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

UTD Author(s): Michael Qiwei Zhang

©2014 The Authors. Published by Elsevier Inc.

Creative Commons 3.0 BY-NC-ND License.
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,1,9 Adrian R. Krainer,3 Robert B. Darnell,2,* and Chaolin Zhang1,9

1Department of Systems Biology, Department of Biochemistry and Molecular Biophysics, Center for Motor Neuron Biology and Disease, Columbia University, New York, NY 10032, USA
2Howard Hughes Medical Institute, Laboratory of Molecular Neuro-Oncology, Rockefeller University, New York, NY 10065, USA
3Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
4Ludwig Institute for Cancer Research, University of California, San Diego, La Jolla, CA 92039, USA
5Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
6Merck Research Laboratories, Merck & Co., Inc., Rahway, NJ 07065, USA
7Department of Molecular and Cell Biology, Center for Systems Biology, The University of Texas at Dallas, Richardson, TX 75080, USA
8Bioinformatics Division, Center for Synthetic and Systems Biology, TNLIST, Tsinghua University, Beijing 100084, China
9These authors contributed equally to this work

*Correspondence: darnellr@rockefeller.edu (R.B.D.), cz2294@columbia.edu (C.Z.)

http://dx.doi.org/10.1016/j.celrep.2014.02.005

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

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, Hnrnpb3, 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. UV crosslinking results in a protein-RNA complex, which is followed by UV crosslink site and RNA partial RNase digestion, immunoprecipitation, 3’ linker ligation, RNA-labeling, and SDS-page. Then, digestion with proteinase K, RNA purification, 5’ linker ligation, cross-linked residue, RT, cDNA read-through, and cDNA purification & circularization result in a cleavage site for relinearization & elution of PCR & sequencing.

(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.

(C) Pooled Rbfox CLIP tags and Nova-binding sites.

(D) Pooled tags

(E) Pooled peaks

(F) Rbfox3 (log₁₀ tag count)

(G) Standard CLIP (log₁₀ tag count)

1140 Cell Reports 6, 1139–1152, March 27, 2014 ©2014 The Authors

(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; ≥94% amino acid identity), they are expected to bind and regulate largely overlapping sets of transcripts (Gallaher 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 occupied by their very similar RRMs. 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 × 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 × 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).
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 < 0.1 and proportional change of exon inclusion \(|\Delta D| \geq 0.1\) (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 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.
either the CLIP data derived from mouse brain or the bio-informatically 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 weight 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
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, 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 × 10⁻³) and “neuron projection” (Benjamini FDR < 3.8 × 10⁻³), 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 × 10⁻³) (Ranganathan and Ross, 1997).
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 x 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 changes in autistic brains (Δ| > 0.1, and FDR ≤ 0.05), a very significant overlap compared to random chance (odds ratio = 3.4, p < 1.4 x 10^-16; Fisher’s exact test; Figure 5A). The autistic brains compared here were selected to have low expression level of Rbfox1 (Voineagu et al., 2011) (4.1-fold downregulation compared to E17 cortex, respectively, whereas 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 x 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
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.

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.

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>
<thead>
<tr>
<th>Entrez Gene ID</th>
<th>Gene Symbol</th>
<th>Tissue Expression</th>
<th>Autism versus Control Splicing Change</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>11538</td>
<td>Adnp</td>
<td>non-brain-specific</td>
<td></td>
<td>transcription factor</td>
</tr>
<tr>
<td>11784</td>
<td>Apba2</td>
<td>brain-specific</td>
<td></td>
<td>neurotransmitter release</td>
</tr>
<tr>
<td>11789</td>
<td>Apc</td>
<td>non-brain-specific</td>
<td></td>
<td>Wnt signaling</td>
</tr>
<tr>
<td>11941</td>
<td>Atp2b2</td>
<td>brain-specific</td>
<td></td>
<td>calmodulin binding</td>
</tr>
<tr>
<td>319974</td>
<td>Autos2</td>
<td>non-brain-specific</td>
<td></td>
<td>unknown</td>
</tr>
<tr>
<td>30948</td>
<td>Bin1</td>
<td>non-brain-specific</td>
<td>Y</td>
<td>synaptic vesicle endocytosis</td>
</tr>
<tr>
<td>12288</td>
<td>Cacna1c</td>
<td>non-brain-specific</td>
<td></td>
<td>ion channel</td>
</tr>
<tr>
<td>12289</td>
<td>Cacna1d</td>
<td>non-brain-specific</td>
<td></td>
<td>ion channel</td>
</tr>
<tr>
<td>12291</td>
<td>Cacna1g</td>
<td>non-brain-specific</td>
<td></td>
<td>ion channel</td>
</tr>
<tr>
<td>54725</td>
<td>Cadm1</td>
<td>non-brain-specific</td>
<td>Y</td>
<td>cell adhesion</td>
</tr>
<tr>
<td>320405</td>
<td>Cadps2</td>
<td>non-brain-specific</td>
<td>Y</td>
<td>synaptic vesicle exocytosis</td>
</tr>
<tr>
<td>100072</td>
<td>Camta1</td>
<td>non-brain-specific</td>
<td></td>
<td>calmodulin binding</td>
</tr>
<tr>
<td>67300</td>
<td>Chtc</td>
<td>non-brain-specific</td>
<td></td>
<td>receptor localization</td>
</tr>
<tr>
<td>104318</td>
<td>Csnk1d</td>
<td>non-brain-specific</td>
<td></td>
<td>cell signaling</td>
</tr>
<tr>
<td>74006</td>
<td>Dnm1I</td>
<td>non-brain-specific</td>
<td></td>
<td>mitochondrial and peroxisomal division</td>
</tr>
<tr>
<td>75560</td>
<td>Ep400</td>
<td>non-brain-specific</td>
<td></td>
<td>chromatin binding</td>
</tr>
<tr>
<td>13876</td>
<td>Erg</td>
<td>non-brain-specific</td>
<td></td>
<td>transcription factor</td>
</tr>
<tr>
<td>14107</td>
<td>Fat1</td>
<td>non-brain-specific</td>
<td></td>
<td>cell adhesion</td>
</tr>
<tr>
<td>268566</td>
<td>Gphn</td>
<td>non-brain-specific</td>
<td></td>
<td>receptor localization</td>
</tr>
<tr>
<td>74053</td>
<td>Grip1</td>
<td>non-brain-specific</td>
<td></td>
<td>glutamate receptor binding</td>
</tr>
<tr>
<td>108071</td>
<td>Grm5</td>
<td>brain-specific</td>
<td></td>
<td>G-protein coupled receptor</td>
</tr>
<tr>
<td>227753</td>
<td>Gsn</td>
<td>non-brain-specific</td>
<td>Y</td>
<td>cytoskeleton</td>
</tr>
<tr>
<td>14886</td>
<td>Gtf2i</td>
<td>non-brain-specific</td>
<td>Y</td>
<td>transcription factor</td>
</tr>
<tr>
<td>16531</td>
<td>Kcnma1</td>
<td>brain-specific</td>
<td>Y</td>
<td>ion channel</td>
</tr>
<tr>
<td>242274</td>
<td>Lrcc7</td>
<td>brain-specific</td>
<td></td>
<td>synapese assembly</td>
</tr>
<tr>
<td>18027</td>
<td>Nfia</td>
<td>non-brain-specific</td>
<td></td>
<td>transcription factor</td>
</tr>
<tr>
<td>319504</td>
<td>Nrcam</td>
<td>brain-specific</td>
<td>Y</td>
<td>cell adhesion</td>
</tr>
<tr>
<td>18189</td>
<td>Nrxn1</td>
<td>brain-specific</td>
<td></td>
<td>cell adhesion</td>
</tr>
<tr>
<td>18191</td>
<td>Nrxn3</td>
<td>brain-specific</td>
<td></td>
<td>cell adhesion</td>
</tr>
<tr>
<td>80883</td>
<td>Ntng1</td>
<td>brain-specific</td>
<td></td>
<td>cell adhesion</td>
</tr>
<tr>
<td>233977</td>
<td>Ppfia1</td>
<td>non-brain-specific</td>
<td>Y</td>
<td>cell signaling</td>
</tr>
<tr>
<td>353211</td>
<td>Prune2</td>
<td>non-brain-specific</td>
<td>Y</td>
<td>unknown</td>
</tr>
<tr>
<td>268859</td>
<td>Rbfox1</td>
<td>non-brain-specific</td>
<td></td>
<td>RNA metabolism</td>
</tr>
<tr>
<td>268902</td>
<td>Robo2</td>
<td>non-brain-specific</td>
<td></td>
<td>axon guidance</td>
</tr>
<tr>
<td>110876</td>
<td>Scr2a1</td>
<td>brain-specific</td>
<td></td>
<td>ion channel</td>
</tr>
<tr>
<td>20273</td>
<td>Scr8a</td>
<td>brain-specific</td>
<td></td>
<td>ion channel</td>
</tr>
<tr>
<td>58234</td>
<td>Shank3</td>
<td>non-brain-specific</td>
<td></td>
<td>synapese assembly</td>
</tr>
<tr>
<td>76376</td>
<td>Slc24a2</td>
<td>brain-specific</td>
<td></td>
<td>ion transport</td>
</tr>
<tr>
<td>72055</td>
<td>Slc38a10</td>
<td>non-brain-specific</td>
<td></td>
<td>ion/amino acid transport</td>
</tr>
<tr>
<td>94229</td>
<td>Slc4a10</td>
<td>brain-specific</td>
<td></td>
<td>ion transport</td>
</tr>
<tr>
<td>64213</td>
<td>St7</td>
<td>non-brain-specific</td>
<td></td>
<td>cell signaling</td>
</tr>
<tr>
<td>53416</td>
<td>Stk39</td>
<td>non-brain-specific</td>
<td></td>
<td>serine/threonine kinase</td>
</tr>
<tr>
<td>20910</td>
<td>Stxbp1</td>
<td>non-brain-specific</td>
<td>Y</td>
<td>neurotransmitter release</td>
</tr>
<tr>
<td>64009</td>
<td>Syne1</td>
<td>non-brain-specific</td>
<td>Y</td>
<td>cytoskeleton</td>
</tr>
<tr>
<td>22084</td>
<td>Tsc2</td>
<td>non-brain-specific</td>
<td></td>
<td>chaperone</td>
</tr>
<tr>
<td>22138</td>
<td>Ttn</td>
<td>low-brain-expression</td>
<td></td>
<td>sarcomere structure</td>
</tr>
</tbody>
</table>

(Continued on next page)
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 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 cross-linking. 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, differentially expressed 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 etiology 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 ChIPs analysis as described previously (Moore et al., 2014; Zhang and Darnell, 2011). ChIPs 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

Semi-quantitative 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 Read 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, Jinhiao Ma and Judith Kribelbauer for assistance with the preparation of living cells. Wiley Interdiscip Rev. RNA

REFERENCES


Cell Reports 6, 1139–1152, March 27, 2014 ©2014 The Authors 1151


SUPPLEMENTAL EXPERIMENTAL PROCEDURES

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

Transient Transfection of Rbfox1, 2, and 3 in HEK293T Cells
To test cross-reactivity of Rbfox antibodies, 1.5×10⁶ HEK293T cells were plated onto 10 cm dishes the day before transfection. Cells were then transfected with 2 µg of pRbfox1, pRbfox2, or pRbfox3 individually, or the empty pcDNA vector as a control, using FuGENE HD (Roche) as described by the manufacturer. After 48 hrs, cells were rinsed with ice-cold PBS and collected in lysis buffer (1X PBS, 0.1% SDS, 0.5% NaDOC, and 0.5% (v/v) NP-40) for protein extraction.

Immunoblot Analysis
A total of 20 µg protein per lane was loaded into 10% or 4-12% Novex NuPage SDS-PAGE gels (Invitrogen). After protein transfer onto nitrocellulose (Millipore) 0.45 µm membranes, the following antibodies were used for immunoblot: α-Rbfox1 (Millipore), α-Rbfox2 (Bethyl Laboratories), α-Rbfox3 polyclonal (in-house), α-Rbfox3 monoclonal (Millipore), and α-GAPDH (Millipore).

HITS-CLIP Experiments using Mouse Brain
Rbfox1, 2, and 3 CLIP experiments were performed with whole brain tissue lysate of about two-week old (P15) CD1 mice using two different protocols. Each CLIP experiment used approximately 1/4-1/3 of a whole brain (100-150 mg tissue lysate). The following antibodies were used: α-Rbfox1 (Millipore), α-Rbfox2 (Bethyl Laboratories), α-Rbfox3 polyclonal (in-house) and monoclonal (Millipore). The standard CLIP libraries were prepared as described previously (Darnell, 2010; Licatalosi et al., 2008; Moore et al., 2014). The BrdU-CLIP protocol is composed of the following steps. Protein-RNA complexes are immunoprecipitated using each specific antibody followed by stringent washes and ligation of the 3’ linker. After digestion of the protein, the RNA is purified and reverse transcribed using a dNTP mix where dTTP is replaced with BrdUTP. The RT primer has 5’ and 3’ anchor sequences used later for PCR amplification separated by an APE1 cleavage site. The BrdUTP incorporated into the resulting cDNA allows on-bead cDNA purification using an anti-BrdU antibody (clone IIB5; Santa Cruz Biotech or other vendors). The purified cDNA is then circularized and relinearized on-bead using CircLigase II and APE1, respectively, which places the two PCR anchors in the correct orientation. PCR is performed using a real time amplification system; SYBR green is added to the PCR reaction mix to monitor the concentration of DNA, and the PCR reaction is stopped when the fluorescence intensity (RFU) reaches an empirically-determined threshold. More details of the protocol are described in Supplemental Protocols. The resulting libraries prepared with both standard CLIP and BrdU-CLIP protocols were sequenced by Illumina Hi-Seq 2000 to obtain 51-nt reads.

RNA-Seq in HeLa Cells
HeLa cells with or without Rbfox2 knockdown were generated previously (Zhang et al., 2008). Briefly, an shRNA against human Rbfox2 (shRbfox2) was cloned into the MSCV retroviral vector (Dickins et al., 2005). To create stable cell pools, HeLa cells were infected with MSCV expressing shRbfox2 or the empty retroviral vector (control).
replaced the medium 24 hrs after infection. 24 hrs later, infected cells were selected with puromycin (2 µg mL⁻¹) for 72 hrs.

To reduce splicing precursors and intermediates, cells were fractionated to enrich cytoplasmic RNA. Specifically, cells were lysed in gentle lysis buffer (10 mM HEPES pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.5 % (v/v) NP-40), and the nuclei were pelleted at 2,300 g for 5 min. Cytoplasmic RNA was extracted from the supernatant by Trizol and treated with DNase I (Promega). We used 4 µg of cytoplasmic RNA from each sample to prepare RNA-Seq libraries following the manufacturer’s instructions (Illumina). cDNA was size-selected using an agarose gel, and a band around 250 nt (corresponding to a fragment size of 185 nt with adaptors excluded) was excised for PCR amplification. The library was sequenced on Illumina GA IIx sequencer to produce paired-end 32-nt reads. Each of the two samples was sequenced in three lanes to increase the coverage.

Rbfox1 KO Mice
Rbfox1<sup>loxp/loxp</sup> mice (described previously by Gehman et al. 2011) and Nestin-Cre<sup>+/+</sup> mice were purchased from Jackson lab and bred to make heterozygous (Rbfox1<sup>loxp/+<br/>Nestin-Cre<sup>+/+</sup></sup>) mice. The heterozygous mice were then crossed with Rbfox1<sup>loxp/loxp</sup> to obtain homozygous (Rbfox1<sup>loxp/loxp/Nestin-Cre<sup>+/+</sup></sup>) offspring. Adult Rbfox1 KO mice at around 9 months and age-matched WT controls were used for validation assays (duplicates for each genotype).

RT-PCR Validation
Semi-quantitative RT-PCR in HeLa cells was performed as previously described (Zhang et al., 2008). For RT-PCR validation, total RNA was extracted from brains of WT and homozygous Rbfox1 KO (Rbfox1<sup>loxp/loxp/Nestin-Cre<sup>+/+</sup></sup>) mice using Trizol reagent (Invitrogen) and treated with DNase I. Candidate exons for validation were selected to cover the whole range of prediction confidence, but exons with complex splicing patterns or very high (≥0.9) or low (≤0.1) inclusion levels in WT adult mouse brains (Dillman et al., 2013) were avoided. Semi-quantitative RT-PCR analysis of these exons was performed as previously described (Zhang et al., 2010). A total of eight exons with both alternative isoforms detected in WT or Rbfox1 brain samples were included in our quantitative analysis. Primers used in this study are listed in Table S5.

Alternative Splicing Database
The alternative splicing database used in this study was built by aligning RefSeq, mRNA, and EST sequences to the mouse genome (mm10), as described previously (Zhang et al., 2010). Alternative splicing events used in our Bayesian network analysis consist of 16,034 cassette exons, 4,074 events of tandem cassette exons, and 960 events of mutually exclusive exons. A similar database was generated for human (hg19) and rat (rn5) to determine conservation of alternative splicing patterns.

Analysis of HITS-CLIP Data
HITS-CLIP data analysis was performed essentially as described previously (Moore et al., 2014; Zhang and Darnell, 2011). In brief, 51-nt raw reads were first filtered by their quality scores. The standard CLIP reads have a 5-nt random barcode, and the BrdU-CLIP reads have a 5-nt sample multiplexing index in addition to a 9-nt random barcode at the 5’ end. For reads kept for downstream analysis, we required a minimum score of 20 at these index/barcode positions and an average score of 20 for the succeeding 25 positions corresponding to the first 25-nt of the actual CLIP tags. Barcode and index sequences were then removed from the read sequences and the information was recorded separately. The remaining read sequences corresponding to actual RNA tags were mapped to the reference genome (mm9 at the time of the analysis) using the novoalign program (http://www.novocraft.com) and allowing ≤2 mismatches (substitutions, insertions, or deletions) per read. We used the iterative trimming mode of novoalign during alignment, which iteratively removes adaptor sequences and low quality bases at the 3’ end so that CLIP tags of smaller sizes can be mapped properly. A minimum of 25-nt matches was required and only those reads mapped unambiguously to the genome (single hits) were kept for downstream analysis. Using an iterative expectation maximization-like statistical model (Darnell et al., 2011; Moore et al., 2014), we collapsed CLIP tags mapping to the same genomic positions into a single tag; tags with sufficiently distinct barcode sequences were kept, however. This step removed PCR duplicates while retaining genuinely unique CLIP tags (Darnell et al., 2011; Moore et al., 2014) and resulted in unique CLIP tags that represent independent captures of protein-RNA interactions. Deletions, insertions, and substitutions—collectively denoted as mutations—in unique CLIP tags were recorded for crosslinking-induced mutation site (CIMS) analysis. Coordinates of unique CLIP tags and mutations therein were then liftOver to mm10 for further analysis.

CLIP tag clusters were identified by grouping overlapping CLIP tags (Moore et al., 2014), and the statistical significance of CLIP tag cluster peak height was evaluated using scan statistics as described previously (Charizanis et al., 2012). This approach compares the observed peak height with the height expected by chance when CLIP tags are randomly shuffled on a gene-by-gene basis. CIMS analysis was performed as described previously (Moore et al., 2014; Zhang and Darnell, 2011).

Crosslinking-induced truncation site (CITS) analysis was performed using a computational method similar to a previous method used to analyze iCLIP data (Konig et al., 2010; Wang et al., 2010) with several modifications. We removed all CLIP tags with deletions since a majority of these likely represent tags where the crosslink sites were read through during reverse transcription. We clustered the remaining CLIP tags based on their potential sites of truncation, defined as the nucleotide immediately 5’ of each CLIP tag. The key to identifying CITS is to distinguish tags that were read through from those with premature truncation. We used a stringent statistical test to evaluate whether the observed frequency of potential truncations in each site is significantly higher than that expected by chance. This was done using a random redistribution of potential truncations in each CLIP tag cluster, which is more stringent than the gene-by-gene permutation used in the previous analyses (Konig et al., 2010; Wang et al., 2010). The stringent permutation scheme is important because read-through is frequent for Rbfox CLIP data. Similar to identification of significant CLIP tag cluster peaks, scan statistics were used for statistical assessment.
To estimate the rate of truncation at the crosslink site, we used the same method as described previously (Sugimoto et al., 2012), using the formula \( p(\text{BrdU-CLIP}) = f \times p(\text{RT}) + (1-f) \times p(\text{BG}) \), where \( p(\text{BrdU-CLIP}) = 0.062 \), \( p(\text{RT}) = 0.14 \), and \( p(\text{BG}) = 0.004 \) are deletion rates in BrdU-CLIP, standard CLIP (i.e., error of reverse transcription), and in RNA-Seq data (background), respectively, and \( f \) is the rate of reverse transcriptase read-through. The background deletion rate in RNA-Seq data was estimated based on our previous data (Zhang and Darnell, 2011). The estimated rate of read-through for Rbfox CLIP is \( f = 43\% \).

**Motif Discovery and Enrichment Analysis**

*De novo* motif analysis was performed using 21-nt sequences [-10,10] around CIMS or CITS. For CIMS, all 1,158 sequences were used for analysis. For CITS, a randomly sampled subset of 1,000 sequences was used for analysis. Repetitive sequences were masked before de novo motif discovery using the MEME program with the following parameters: -mod zoops -nmotifs 10 -minw 6 -maxw 8, with the remaining parameters left at their default values (Bailey and Elkan, 1994). To examine motif enrichment relative to protein-RNA crosslink sites, the starting positions of UGCAUG, VGCAUG (V=nonU), or additional variants of the Rbfox consensus binding motif were recorded in sequences flanking the crosslink sites. Background motif frequency was estimated from control sequences of positions -500~401 nt and 398~497 nt relative to the crosslink sites, to normalize the observed motif frequency.

**Motif Enrichment and Conservation Score**

We previously used branch length score (BLS) (Stark et al., 2007) to evaluate the strength of Rbfox motif sites (Zhang et al., 2008). This scoring method considers only the conservation, but not the overall enrichment, of motif sites. For example, a site present only in the reference species without orthologs in any other species receives a BLS score of zero, even if the motif is specific and enriched in CLIP tag clusters. To address this caveat, we defined a new scoring metric of motif-site strength reflecting both enrichment and conservation of motif sites in CLIP tag clusters compared to regions without any CLIP tags. Specifically, we obtained 6,101 CLIP tag clusters satisfying the following criteria as foreground sequences: i) peak height \( \geq 10 \); ii) located in exons or flanking intronic regions (exon+1kb extension on each side, or exon+ext1k); and iii) no overlap with repeat-masked regions. We also obtained a list of sequences in 90,637 exon+ext1k regions without any CLIP tags as background. We then searched for UGCAUG elements in the foreground and background sequences and calculated the branch length score (BLS) for each site. A distribution of BLS was estimated in CLIP+ and CLIP- sequences, denoted as \( P(\text{BLS}|\text{CLIP+}, \text{region}) \) and \( P(\text{BLS}|\text{CLIP-}, \text{region}) \), respectively, where the type of region can be introns, coding exons, or noncoding exons. The overall frequency of motif sites in foreground and background sequences and calculated the branch length score (BLS) for each site. A distribution of BLS was estimated in CLIP+ and CLIP- sequences, denoted as \( f(\text{CLIP+}) \) and \( f(\text{CLIP-}) \), respectively, were calculated and normalized by the length of sequences in each set. The preliminary motif enrichment and conservation score (MECS) for each UGCAUG site was calculated as \( \log_2 [f(\text{CLIP+})/f(\text{CLIP-})] + \log_2 [P(\text{BLS}|\text{CLIP+}, \text{region})/P(\text{BLS}|\text{CLIP-}, \text{region})] \). We then performed a 3rd-order polynomial regression between the preliminary MECS score and BLS score. The fitted value for the given BLS score was then used as the final MECS score for each motif site (*Figure S2 J,K*). The same procedure was used to estimate the MECS scores for VGCAUG elements.
RNA-Seq Data Analysis

For the HeLa cell RNA-Seq data, raw 5´ and 3´ reads were mapped independently by Eland (Illumina) to the human genome (hg18 at the time of analysis) with iterative 3´ trimming. We required a minimal matched region of 22 nt and ≤ 2 mismatches for alignment. Reads were also mapped to an exon-junction database with the same criteria, in addition to requiring ≥ 4 nt overlap on each side of exon junctions. Only reads unambiguously mapped to unique loci in the genome or exon junctions (single hits) were kept for further analysis. The downstream analysis, including inference of transcript structure and quantification of gene expression and splicing, was performed as described previously (Charizanis et al., 2012) except that the proportional change of exon inclusion between the two conditions was estimated using the ASPIRE algorithm (Ule et al., 2005).

In addition to the HeLa cell RNA-Seq data, we also analyzed three published RNA-Seq datasets: one dataset profiled six different mouse tissues including cortex, cerebellum, heart, kidney, liver and testis (76-nt reads; each tissue has 2-6 biological replicates; GEO accession: GSE30352) (Brawand et al., 2011); one dataset profiled E17 and adult mouse cortex (80-nt single end reads, 4 biological replicates for E17 and 3 biological replicates for adult; GEO accession: GSE39866) (Dillman et al., 2013); and one dataset profiled autistic and control postmortem human brains (74-nt reads; 3 biological replicates for autistic and control brains, respectively; GEO accession: GSE30573) (Voineagu et al., 2011). These datasets were analyzed with the same pipeline (http://zhanglab.c2b2.columbia.edu/index.php/Quantas). Reads in each dataset were mapped to the mouse (mm10) or human (hg19) genome by OLego (http://zhanglab.c2b2.columbia.edu/index.php/OLego) (Wu et al., 2013) with 14-nt non-overlapping seeds, requiring ≤ 3 mismatches per read. An exon-junction database was provided for alignment. Only reads unambiguously mapped to the genome or exon junctions (single hits) were used for downstream analysis. Since these are single-end reads, the resulting mapped reads were counted in each exon or exon junction to quantify splicing changes, as described previously (Charizanis et al., 2012). Fisher’s exact test was used to evaluate the statistical significance of splicing changes, and the false discovery rate was estimated by the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). An alternative splicing event was called in the two compared conditions if the event had sufficient read coverage (coverage≥20), |ΔI|≥0.1, and Benjamini FDR≤0.05 (Charizanis et al., 2012), unless otherwise indicated.

Integrative Modeling using Bayesian Network Analysis

We initially focused on cassette exons using our established Bayesian network framework (Zhang et al., 2010). For evidence of Rbfox-dependent splicing in the Bayesian network analysis, we considered the following datasets: i) differential splicing between WT and Rbfox1 KO mouse brain, as measured by exon-junction microarrays (GEO accession: GSE28421) (Gehman et al., 2011) where we estimated the splicing changes, denoted ΔI_rank, using the ASPIRE3 algorithm (Wang et al., 2010); ii) differential splicing (ΔI) between HeLa cells with or without Rbfox2 expression, as described above; iii) tissue-specific splicing (ΔI) of exons in cortex and cerebellum compared to several other tissues (liver, kidney and testis) using a published RNA-Seq dataset (Brawand et al., 2011); iv) tissue-specific splicing of exons (ΔI) in heart compared to several other tissues (liver, kidney and testis) (Brawand et al., 2011). For RNA-Seq data, only those exons with sufficient coverage (≥ 20 reads) were used; missing values were assigned to the other exons. Tissue-
specific splicing of exons are indirect indicators of Rbfox-dependent splicing, so the information was used for model training but not for the prediction of individual targets (Figure 4A in the main text and below).

For motif sites, we considered both UGCAUG and VGCAUG elements, and measured their strength by their MECS score. Since multiple motif sites are typically present in the alternatively splicing region, we derived a summarized motif site score by weighting each site according to their strength and the distance to the splice sites, as described previously (Zhang et al., 2010). Each cassette exon was measured by six regional motif site cluster scores denoted as $s_{UI5ss}$, $s_{UI3ss}$, $s_{E3ss}$, $s_{E5ss}$, $s_{DI5ss}$, and $s_{DI3ss}$ (UI: upstream intron; DI: downstream intron; E: exon). In each region (UI, E, and DI), the maximum of the two was used to represent motif site strength in the region for Bayesian network analysis. In contrast to our previous analysis (Zhang et al., 2010), the actual distance of motif sites to the splice sites was used for weighting without adjustment. The CLIP tag cluster score in each region was obtained similarly, using the same weighting function based on the peak height of the clusters (Zhang et al., 2010).

To train the model, we compiled 132 exons that have been previously validated as Rbfox targets in humans or mice, mostly using cell culture systems. A subset of these exons was analyzed by detailed mutation analysis to verify direct regulation by Rbfox, while the others, a majority, were validated only by RT-PCR. Therefore, we excluded a few exons that had inconsistent directions of splicing changes in different studies or that did not have CLIP tags or the (U)GCAUG motif sites in the alternatively spliced region, because these are most likely indirect targets. The final set of validated Rbfox targets was composed of 121 exons (87 and 34 exons experiencing Rbfox-dependent inclusion and exclusion, respectively). A class label (activation or repression by Rbfox) was assigned to each of these exons. Besides these validated target exons, we also included additional exons satisfying any of the following criteria for model training: i) exons with $|\Delta I_{\text{rank}}| \geq 0.6$ in the comparison of WT vs. Rbfox1 KO mice using exon junction microarrays; ii) exons with $|\Delta I| \geq 0.08$ in the comparison of control vs. shRbfox2 in HeLa cell RNA-Seq data; iii) exons with $|\Delta I| \leq 0.08$ in the comparison of brain or cerebellum vs. other tissues and in comparison of heart vs. other tissues. These exons were not assigned a class label during model training. Overall, we obtained 551 non-redundant exons to estimate model parameters, with a relatively balanced representation of exons likely to be activated or repressed by Rbfox and of non-target exons.

The trained Bayesian network model was then applied to 16,034 cassette exons to predict direct Rbfox targets. The same model was also applied to 4,074 events of tandem cassette exons and 960 events of mutually exclusive exons, assuming these alternative splicing events are regulated by Rbfox through similar mechanisms.

**Gene Ontology (GO) and Protein Domain Analysis**

GO and protein domain analysis was performed using the online tool DAVID using all protein-coding genes expressed in the brain as the background (Dennis et al., 2003).

**Candidate Autism-Susceptibility Genes**
The list of candidate autism-susceptibility genes was obtained from the Simons Foundation Autism Research Initiative (SFARI) autism gene database (http://gene.sfari.org) (Basu et al., 2009). The mouse ortholog of each human gene was determined using the HomoloGene database (http://www.ncbi.nlm.nih.gov/homologene) complemented by manual searches.

**Statistical Analysis**

All statistical analyses were performed using the statistical software R (R Development Core Team, 2008).
SUPPLEMENTAL RESULTS

Optimization of Rbfox CLIP Conditions

Several Rbfox antibodies were tested to evaluate their specificity in recognizing the targeted proteins. To assess potential cross-reactivity of antibodies with different Rbfox family members, we transfected pRbfox1, pRbfox2, and pRbfox3 individually into HEK293T cells, which have little to no endogenous expression of the Rbfox proteins. Immunoblot was then used to detect each protein in these cells using different antibodies. We identified one antibody for each that recognizes the targeted Rbfox family member without discernible cross-reactivity (Figure S1A). For Rbfox3, in addition to a rabbit polyclonal antibody, a commercial monoclonal antibody is also largely specific despite minor cross-reactivity with Rbfox2.

We then optimized the conditions for the immunoprecipitation (IP) of RNA-protein complexes using mouse brain tissue. Each antibody was able to precipitate its specific Rbfox paralog. When an increasing amount of antibody was used for IP, the IP eluate gave increased yield of the targeted protein as revealed by immunoblot and an autoradiogram labeling RNA crosslinked to the protein; reciprocal depletion of the specific protein, but not the other paralogs, was observed in the supernatant (Figure S1 B,C and data not shown). As a control, we performed IP of crosslinked tissue using nonspecific IgG antibodies or non-crosslinked tissue using Rbfox antibodies, which resulted in minimal background (Figure S1C). Under high RNase concentrations, IP of Rbfox crosslinked with radiolabeled RNA using specific Rbfox antibodies resulted in a doublet band at around 50 kD (Figure S1 C, D), corresponding to the molecular weight of the proteins observed in several previous studies (Dredge and Jensen, 2011; Tang et al., 2009). At a lower RNase concentration, a smear of labeled RNA of varying sizes above the doublet band was observed (Figure S1D), as expected in CLIP experiments (Moore et al., 2014). It is also important note that the size of the doublet band at around 50 kD is similar, but not identical, for different Rbfox family members (Figure S1D). This observation provides support for IP specificity since four different antibodies would not be expected to give nonspecific bands of such similar patterns. Instead, the result is consistent with the notion that multiple homologous variants are present in all Rbfox family members. We isolated crosslinked protein-RNA complexes roughly 20 kD above the molecular weight of each protein for further experimentation.

Conditional Probability Distributions Learned by Bayesian Network

The following observations were made from the estimated Bayesian network model parameters. First, stronger motif sites are more likely to be bound by the protein. For example, MECS scores of 10 and 20 predict Rbfox binding with probabilities of 0.92 and >0.99, respectively (Figure 4B). Second, regions bound by Rbfox are supported by more CLIP tags than regions not bound by Rbfox. About 18% of regions predicted to be bound by Rbfox have a CLIP tag cluster score ≥10, as compared to 0.9% for regions predicted to have no Rbfox binding (Figure 4C). In addition, we also confirmed and quantified the position-dependent RNA map: binding of Rbfox in the downstream intron resulted in Rbfox-dependent inclusion with a probability of 0.99, while binding of Rbfox in the upstream intron or exon results in repression with a probability of 0.75 and 0.61, respectively; binding of Rbfox in both exon and upstream intron increases the probability of repression to 0.84 (Figure 4D). This analysis also showed that altered splicing after perturbation of
Rbfox in HeLa cells or in the Rbfox1 KO mouse model, as well as differential splicing in different tissues, are informative for identifying Rbfox target exons in the brain; exons inferred to be activated or repressed by Rbfox or those with no response had different means of splicing changes. However, the discriminative power is moderate, as reflected in the largely overlapping distributions of splicing changes (Figures 4E and S3 A-C). Finally, alternative splicing of Rbfox target exons is much more likely to preserve the reading frame than that of non-target exons (85% versus 42%) and such alternative splicing events are also more likely observed in other mammalian species (Figure S3 D,E).

**Comparison of Rbfox Target Exons Defined by Bayesian Network with Previously Validated Exons**

In total, 88 of 121 previously validated exons were predicted as Rbfox targets (69/87 exons with Rbfox-dependent inclusion and 19/34 exons with Rbfox-dependent exclusion). This comparison gives a sensitivity of 73% to our Bayesian network prediction, with a conservative assumption that all of these previously validated exons are directly regulated by Rbfox; it is important to note that some of these previously validated Rbfox-dependent exons are likely indirect Rbfox targets, which appears to be particularly true for exons showing Rbfox-dependent exclusion. To address this issue, we focused on 14 cassette exons for which the direct regulation by Rbfox has been previously validated by mutagenesis or RNA IP (Figure S4 and Table S7). Among these, 11 exons were predicted as Rbfox targets by the Bayesian network analysis, indicating a sensitivity of 79%.

**Comparison of Rbfox Target Exons Defined by Bayesian Network and Exons by Previous Studies**

We compared the results of the Bayesian network analysis to our previous motif-based bioinformatic predictions (Zhang et al., 2008). Among the 432 previously predicted Rbfox target cassette exons conserved between human and mouse, 273 (63%) were identified by integrative modeling. While this overlap is substantial (p<1.8×10^{-263}; Fisher’s exact test), we were able to identify important differences in the results. For instance, integrative modeling identified a greatly expanded set of target exons missed by our previous motif-based predictions, presumably using information obtained from the CLIP and other data, while eliminating a number of exons having conserved motif sites but lacking supporting evidence from the other complementary datasets.

Another recent study predicted 933 Rbfox target exons in human using a “leading-edge” analysis based on the presence of RBP motif sites and the correlation of exon splicing with RBP expression in different human tissues and cell lines (Ray et al., 2013). Among the 703 exons annotated as cassette exons in our database, 117 exons have mouse orthologs predicted by our Bayesian network analysis, representing a very significant overlap (p<2.2×10^{-90}; Fisher’s exact test). Again, however, these results suggest that a majority of exons predicted by the two methods are distinct (Figure S5A). Target exons predicted by the Bayesian network analysis more frequently have conserved splicing patterns between human and mouse (78% vs. 26%; Figure S5B), preserved reading frame (89% vs. 57%, as judged from whether the exon size is a multiple of three; Figure S5C), smaller exon size (median: 69 nt vs 96 nt; p<10^{-17}; Figure S5D), and relatively higher expression in the brain (median of expression quantile across all genes: 0.19 vs. 0.25; p<3.3×10^{-7}; Figure S5E). While these differences are not completely unexpected, they suggest that Bayesian network analysis is effective for integrating features that are known to be consistent with tight regulation of alternative splicing events (Xing and Lee, 2006).
SUPPLEMENTAL REFERENCES


Day 1

Note:
The first part of the BrdU-CLIP protocol (up to precipitation of IP’d RNA) is the same as the standard CLIP method, and is available at http://www.rockefeller.edu/labheads/darnellr/CLIP_methods.pdf

Bead Prep: Denhardt’s Blocking

Ab Binding Buffer:

1X PBS, pH 7.4
0.02% Tween-20

50µl Protein-G Dynabeads per sample (25µl per cDNA purification step)

Wash 3 times with Ab binding buffer

Add 225µl Ab binding buffer, 25µl 50X Denhardt’s Solution (Sigma, D2532); total volume is 5X original bead volume

Rotate at RT for at least 45 minutes - 1 hour

RT Reaction:

Add 8µl water to RNA pellet (tap to resuspend, quick spin down). Denature at 65°C for 5 minutes (in microfuge tube), place tube on ice (to avoid loss of RNA, do not overdry pellet and do not pipette until after denaturing step)

Transfer to PCR tube (on ice)

Mix I:

4µl 5X RT Buffer
1µl dATP
1µl dCTP
1µl dGTP
1µl Br-dUTP (8.2mM; Sigma, B0631)
1µl RT Primer (25µM)
9µl total

Add 9µl of Mix I

3 minutes at 75°C, ramp down to 50°C and hold

Mix II:

1µl DTT (0.1M)
1µl RNAsin Plus (Promega, N261)
1µl Superscript III
3µl total

Add 3µl of Mix II (pre-warm to 50°C in PCR block before adding)

45 minutes at 50°C, 15 minutes at 55°C, 5 minutes at 85°C, 4°C hold
**Bead Prep: Ab Binding**

**1X IP Buffer:**

- 0.3X SSPE
- 1mM EDTA
- 0.05% Tween-20

Wash 3 times with Ab binding buffer

Add 20µl Ab binding buffer, 5µl 50X Denhardt’s Solution and 25µl (5µg) αBrdU antibody (Santa Cruz, sc-32323)

Rotate at RT for at least 45 minutes

Wash 3 times with 1X IP Buffer

Following RT reaction, add 1µl (at 2U/µl) RNAse H (Invitrogen 18021-071 or NEB M0297L)

Incubate for 20 minutes at 37°C, hold at 4°C

Add 10µl water (to bring volume above 25µl needed for G-25 column)

Spin through G-25 column to remove free BrdUTP (discard G-25 column as solid radioactive waste)

**cDNA Purification: Immunoprecipitation I**

**2X IP Buffer:**

- 0.6X SSPE
- 2mM EDTA
- 0.1% Tween-20

**4X IP Buffer:**

- 1.2X SSPE
- 4mM EDTA
- 0.2% Tween-20

**Nelson Low Salt Buffer:**

- 15mM Tris pH 7.5
- 5mM EDTA
**Nelson Stringent Buffer:**

15mM Tris-HCl pH7.5  
5mM EDTA  
2.5mM EGTA  
1% Triton X-100  
1% NaDOC  
0.1% SDS  
120mM NaCl  
25mM KCl

Measure volume, add water up to 40µl and add 10µl 50X Denhardt’s Solution and 50µl 2X IP Buffer for a total volume of 100µl (Denhardt’s and 2X IP Buffer can be added to the G-25 column collection tube prior to spinning samples through, volume can then be adjusted up to 100µl)

5 minutes at 70°C, equilibrate to RT

Add to prepared tube of beads (25µl original slurry volume, store remaining beads for second purification at 4°C O/N), rotate at RT for 30-45 minutes

(all washes 3-5 minutes rotating at RT, 1ml)

Wash 1 time with 1X IP Buffer (5X Denhardt’s)  
2 times with Nelson Low Salt Buffer (1X Denhardt’s)  
2 times with Nelson Stringent Buffer (1X Denhardt’s)  
2 times with 1X IP Buffer

**Competitive Elution (with BrdU):**

Add 50µl of 100µM BrdU (Sigma, B5002, diluted in 1X IP Buffer)

Rotate at RT for 30-45 minutes, place on magnet and collect eluate

Spin through G-25 column to remove free BrdU

Measure volume, add water up to 97.5µl, add 37.5µl 4X IP buffer and 15µl 50X Denhardt’s Solution (Denhardt’s and 4X IP Buffer can be added to the G-25 column collection tube prior to spinning samples through)

Store overnight at 4°C
cDNA Purification: Immunoprecipitation II

**CircLigase Wash Buffer:**

33mM Tris-Acetate  
66mM KCl  
(pH 7.8)

5 minutes at 70°C, equilibrate to RT

Add to previously prepared beads, rotate at RT for 30-45 minutes

(all washes 3-5 minutes rotating at RT, 1ml)

Wash 1 time with 1X IP Buffer (5X Denhardt’s)  
2 times with Nelson Low Salt Buffer (1X Denhardt’s)  
2 times with Nelson Stringent Buffer (1X Denhardt’s)  
2 times with CircLigase wash buffer

**Circularization: CircLigase**

**APE1 Wash Buffer:**

50mM Potassium Acetate  
20mM Tris-Acetate  
10mM Magnesium Acetate  
(pH 7.9)

**Reaction Mix:**

2µl CircLigase 10X Reaction Buffer  
4µl Betaine (5M)  
1µl MnCl₂ (50mM)  
0.5µl CircLigase ssDNA Ligase II (50U) (Epicentre, CL9021K)  
12.5 Water  
20µl total

Incubate 1 hour at 60°C in thermomixer (interval: shake at 1300rpm every 30” for 15”)

(all washes 3-5 minutes rotating at RT, 1ml)

Wash 2 times with Nelson Low Salt Buffer  
2 times with Nelson Stringent Buffer  
2 times with APE1 wash buffer

Day 2
Re-linearization: APE1

**Phusion Wash Buffer:**

50mM Tris  
(pH 8.0)

**Reaction Mix:**

2µl  10X Reaction Buffer (NEB 4)  
1.25µl  APE1 (10U/µl, NEB M0282S)  
16.75µl  Water  
20µl total

Incubate for 1 hour at 37°C in thermomixer (interval: shake at 1300rpm for 15” every 30”)

(all washes 3-5 minutes rotating at RT, 1ml)

Wash 2 times with Nelson Low Salt Buffer  
2 times with Nelson Stringent Buffer  
2 times with Phusion wash buffer

**PCR: Phusion Polymerase, SYBR Green**

**Mix I:**

10µl  5X Phusion HF Buffer  
1µl  10mM dNTPs  
37µl  Water  
48µl total

**Mix II:**

0.5µl  P5 (20µM)  
0.5µl  P3 (20µM)  
0.5µl  Phusion DNA Polymerase (NEB, M0530)  
1.5µl total

Add Mix I to beads  
Incubate at 98°C for 1 minute in thermomixer (1200 rpm) to elute cDNA from beads and place on magnet  
Collect supernatant and place in PCR tube with optically clear cap (BioRad, TCS-0803)  
Add 1.5µl Mix II  
Add 0.5µl 50X SYBR Green I (Invitrogen, S7563) diluted in Phusion wash buffer to mix and place in real-time PCR machine
Cycle:

98°C  30”
98°C  10”
60°C  15”
72°C  20”

Remove reaction tubes when RFU signal reaches 500-1000 (BioRad CFX96 Touch™ Real-Time PCR Detection System, or BioRad iQ5)

Purify PCR product using Agencourt AMPure XP beads (Beckman Coulter) according to manufacturers instructions (usually results in 5-10nM in 30μl eluate)

Quantitate using Quant-it kit (Invitrogen) or Tapestation (Agilent), if multiplexing, pool samples according to quantitation results

Libraries are compatible with Illumina single-end HiSeq flowcells (not compatible with MiSeq, or paired-end HiSeq flowcells)

Submit for Illumina sequencing using small RNA sequencing primer listed below
RNA adapter (Dharmacon):

SRA3: 5’-P UCGUAUGCCGUCUUCGCUUG–puromycin-3’

Pre-Adenylated DNA adapter (IDT): Higher ligation efficiency, use with truncated T4 RNA Ligase 2, NEB M0242L)

Pre-A SRA3: 5’-/5rApp/TCGTATGCCTCTCTGCTG/3ddc/-3’

DNA Primers (IDT):

RT Primers:

RT-1
5’pGNNNNNNNGATGCATCGTCGGACTGTAGAACTCT/idSp/CAAGCAGAAGACGCGCATACGA

RT-2
5’pGNNNNNNNGTGCAATCGTCGGACTGTAGAACTCT/idSp/CAAGCAGAAGACGCGCATACGA

RT-3
5’pGNNNNNNNGGCATGATCGTCGGACTGTAGAACTCT/idSp/CAAGCAGAAGACGCGCATACGA

RT-4
5’pGNNNNNNNGAGCTGATCGTCGGACTGTAGAACTCT/idSp/CAAGCAGAAGACGCGCATACGA

RT-5
5’pGNNNNNNNGTCCAATCGTCGGACTGTAGAACTCT/idSp/CAAGCAGAAGACGCGCATACGA

RT-6
5’pGNNNNNNNGTGGATCGTCGGACTGTAGAACTCT/idSp/CAAGCAGAAGACGCGCATACGA

PCR Primers:

Forward (P5):
5’-AATGATACGGCGACCACCAGGATTCAGATTTCTACAGTCCGACG

Reverse (P3):
5’-CAAGCAGAAGACGCGATA

Sequencing Primer: Illumina Small RNA Sequencing Primer

5’ -CGACAGGTTCCAGAGTTCTACAGTCCGACGATC
Fig. S1: Optimization of Rbfox1, 2, and 3 HITS-CLIP conditions, related to Figure 1.
A. HEK293T cells were transfected with one of the Rbfox1, 2, or 3 plasmids or an empty vector control. Total protein extract from these cells was used for immunoblot analysis using each α-Rbfox antibody to test for cross-reactivity. α-GAPDH was used as a loading control.
B. Immunoprecipitation of Rbfox3 in mouse brain tissue using different amounts of α-Rbfox3 rabbit polyclonal antibodies. Immunoblot using another monoclonal antibody detected a doublet band.
C. Immunoprecipitation of Rbfox1 (top left) and Rbfox2 (bottom left) in crosslinked brain tissue using different amounts of antibodies to isolate crosslinked Rbfox-RNA complexes with high RNase concentration. Nonspecific IgG was used as a control. The isolated RNA was radio-labeled and visualized by autoradiogram. Each sample was loaded into two lanes, as labeled below. The post-IP supernatant of the same samples was used for immunoblot with α-Rbfox1 (top right) and α-Rbfox2 antibodies (bottom right).
D. Representative CLIP autoradiogram for Rbfox1, 2, and 3. In each experiment, high-concentration RNAase treatment resulted in a doublet band corresponding to the molecular weight of the protein, and low RNase concentration resulted in a smear above the expected molecular weight. The yellow boxes indicate the regions cut out for CLIP experiments.
Fig. S2: Rbfox binding motif analysis in CIMS and CITS data, related to Figure 2.
A,C. Crosslink sites inferred from CIMS analysis and CITS analysis, respectively. The dotted box indicates the UGCAUG motif directly revealed by anchoring sequences at the crosslink sites. B,D. The strongest de novo motif in positions [-10,10] around CIMS and CITS, respectively. The number of sequences containing the motif and the statistical significance (E-value) are indicated.
E,F,G. Rbfox recognizes some, but not all, variants of the (U)GCAUG consensus motif. Enrichment of NGAAUG (E), NGCCUG (F) and UGCAUG.m1 (G) around CIMS is shown. UGCAUG.m1 represents all hexamers with one mismatch compared to the UGCAUG consensus sequence (with NGCAUG excluded).
H,I. Standard CLIP data were used as a control to perform CITS analysis. Enrichment of UGCAUG (H) and VGCAUG (I) around “mock” CITS derived from standard CLIP tags, presumably without crosslinking-induced truncation, was shown.
J,K. Motif enrichment and conservation score (MECS) for UGCAUG (J) and VGCAUG (K). Motif sites are grouped by their genomic location, and scored separately. The dots represent the preliminary MECS. The final MECS is smoothed by interpolating the preliminary MECS with branch length score (BLS) measuring motif site conservation in mammalian species (the curves).
Fig. S3: Integrative modeling of Rbfox target cassette exons using a Bayesian network, related to Figure 4.

A-C. Conditional probability distributions (CPDs) of splicing changes. For each panel, splicing changes were modeled using a normal distribution. The distributions for three groups of exons (activated by Rbfox, no effect from Rbfox and repressed by Rbfox) are shown in red, gray and blue, respectively. A. WT vs. Rbfox1 KO brain. B. brain vs. other tissues (liver, kidney and testis). C. heart vs. other tissues.

D. CPD of reading frame preservation.

E. CPD of alternative splicing conservation between mouse and human/rat.

F. Evaluation of over-fitting. A total of 121 validated target exons compiled from the literature were used for this evaluation. X-axis shows the FDR of each exon predicted by the full model, which was trained using the complete training dataset (including all validated exons). Y-axis shows the FDR of each validated exon predicted in 10-fold cross validation (CV). In this cross validation procedure, models were trained using 90% of the training data, then used to predict the remaining 10% of exons.
Fig. S4: Direct Rbfox target cassette exons validated in the literature, related to Figure 4.
In each panel, the alternative exon together with the flanking introns and exons are shown, followed by pooled Rbfox CLIP tags and (U)GCAUG elements from top to bottom. Exons activated and repressed by Rbfox are shown in red and blue, respectively.
Fig. S5: Comparison of the Bayesian network (BN) approach with leading-edge analysis by Ray et al., related to Figure 4.

A. The Venn Diagram shows the overlap of Rbfox target cassette exons predicted by the Bayesian network analysis (599 of 772 exons with conserved alternative splicing in human) with those predicted by leading-edge analysis (693 exons). The significance of overlap is evaluated by Fisher’s exact test.
B. The percentage of exons with conserved alternative splicing between human and mouse.
C. The percentage of frame-preserving exons (exon length is a multiple of three).
D. The cumulative distribution of exon size.
E. The cumulative distribution of the quantile of gene expression in the brain that measures the relative abundance of each gene.
Fig. S6: Validation of Rbfox target exons in genes implicated in autism, related to Figure 6. In each panel, the alternative exon together with the flanking introns and exons are shown, followed by pooled Rbfox CLIP tags and (U)GCAUG elements from top to bottom. Exons activated and repressed by Rbfox are shown in red and blue, respectively. The size of the alternative exon and the gene symbol are indicated at the top. The gel image and quantitation of semi quantitative RT-PCR analysis using WT and Rbfox1 KO mouse brains is shown on the right. The bands corresponding to the two alternative isoforms are indicated.