Title

Low Complexity Sequences of Rbfox Form Higher-order Complexes with LASR to Regulate Alternative Splicing

Author

Ying, Yi

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Low Complexity Sequences of Rbfox Form Higher-order Complexes with LASR to Regulate Alternative Splicing

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Molecular Biology

by

Yi Ying

2016

ABSTRACT OF THE DISSERTATION

Low Complexity Sequences of Rbfox

Form Higher-order Complexes with LASR

to Regulate Alternative Splicing

by

Yi Ying

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2016

Professor Douglas L. Black, Chair

Alternative splicing is controlled by diverse RNA binding proteins that recognize elements in the pre-mRNA to alter spliceosome assembly. Separate from their RNA binding domains, these proteins often contain intrinsically disordered domains with regions of low-complexity (LC) sequences, but how LC sequences contribute to splicing regulation is not known. In earlier work, we found that splicing regulators of the Rbfox family are bound with a large complex of proteins called the Large Assembly of Splicing Regulators, LASR. Rbfox proteins were shown to regulate splicing in association with

LASR and to alter the activity of LASR components in splicing, but the nature of the Rbfox and LASR interaction was not clear. Here, we show that C-terminal domain of the Rbfox protein interacts with LASR and this interaction is essential for Rbfox activity in splicing. We find that an LC region within the C-terminal domain mediates assembly of Rbfox proteins with LASR into higher-order structures. Repetitive tyrosine residues in this domain are essential to the formation of the higher-order assemblies. The Rbfox LC domain both spontaneously aggregates in solution and forms fibrous structures and hydrogels over time, suggesting a mechanism for higher-order assembly similar to the fibril formation with FUS and other RNA-binding proteins. Exon repression and activation by Rbfox proteins are lost with mutations that disrupt the interaction of Rbfox and LASR. However, blocking higher-order assembly while retaining the Rbfox interaction with LASR, results in selective loss of Rbfox-dependent exon activation. These findings demonstrate that the LC domains of RNA-binding proteins and their self-assembly play a crucial role in splicing regulation. In addition to simple RNA recognition, higher-order assembly and its associated aggregation properties of phase separation and/or fiber formation offer additional mechanisms for tuning regulatory activities.

The dissertation of Yi Ying is approved.

Arnold J. Berk

Reid C. Johnson

Yi Xing

Douglas L. Black, Committee Chair

University of California, Los Angeles

2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am very grateful to all the people I met that walked me through either good or bad days in my PhD adventure, and those I met earlier that prepared me ready for this adventure in my life.

Firstly, I would like to express my sincere gratitude to my PhD advisor Doug Black for tremendous support of my PhD study and research, for his patience, motivation and immense knowledge. I am very grateful for every conversation we had on research projects as well as other non-scientific questions. So much I learned from him to become a mature scientist and so much I still need to learn. I could not imagine I could accomplish all these without him. I would also like to thank my committee for meeting with me every year and all the comments and questions that broaden my perspectives of my project.

I cannot thank enough to all the Black lab members both past and present who helped and accompanied me for the past six years. Thanks to Amy Pandya-Jones and Shalini Sharma for their encouragement even before I started graduate school. Working with them as an international undergraduate, I was so excited by the beauty of science and convinced me to pursue a career as a scientist. I would like to thank Andrey Damianov for teaching me a lot in doing experiments when I started in the lab and helpful discussions throughout my study. Thanks to Xiao-Jun Wang for teaching and helping me purify recombinant proteins to speed up the project. Thanks to Chia-Ho Lin for help in data analysis and coding questions. I would also like to thank my peers in lab: Areum Han, Pop Wongpalee, Anthony Linares, Celine Vuong for all the time together, all the

suggestions and discussions. Thank again to all the Black lab members and to our lab manager Julia Nikolic for making it such a great lab.

Many thanks to Andrey Daminanov, Pop Wongpalee, Celine Vuong, Xiao-Jun Wang, Xin Liu, Xiao Xiao for proofreading my dissertation and helpful comments.

I would like to thank my parents and my loving husband Xiao Xiao for supporting me all the time. Without them, I would not be strong enough to go through all those tough days in my life. Thanks to my friends here in US and back in China for keeping the loneliness away from me.

Lastly, I would thank China Scholarship Council for financial support, as well as UCLA dissertation year fellowship.

Chapter 2 is a reprint of a manuscript with permission from Elsevier and Copyright Clearance Center.

Damianov, A., **Ying, Y.**, Lin, C.H., Lee, J.A., Tran, D., Vashisht, A.A., Bahrami-Samani, E., Xing, Y., Martin, K.C., Wohlschlegel, J.A.*, et al.* (2016). Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. Cell *165*, 606-619.

VITA

EDUCATION

PhD Student	**University of California, Los Angeles** September 2010 ~ Present
B.S.	**Zhejiang University**, China Biological Science, June 2010

PUBLICATIONS

Damianov, A., **Ying, Y.**, Lin, C.H., Lee, J.A., Tran, D., Vashisht, A.A., Bahrami-Samani, E., Xing, Y., Martin, K.C., Wohlschlegel, J.A.*, et al.* (2016). Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. Cell *165*, 606-619.

Zhang, X.Q., **Ying, Y.**, Ye, Y., Xu, X.W., Zhu, X.F. and Wu, M. (2010). Thermus arciformis sp. nov., a thermophilic species from a geothermal area. International Journal of Systematic and Evolutionary Microbiology.

Zhang, W.J., Zhang, X.Q., **Ying**, **Y.,** Xu, X.W., Wu, M. (2009). Analysis of Bacterial diversity in deep-sea sediments from Pacific polymetallic nodule province. Journal of Zhejiang University (Science Edition) Chinese

PRESENTATIONS

2015	Eukaryotic mRNA processing, CSHL Talk: Rbfox-mediated alternative splicing via its interaction with a large assembly of splicing regulators, LASR
2014	Gordon Conference: Post-transcriptional Gene Regulation Poster: Characterization of a Large Rbfox Splicing Regulatory Complex
2013	Eukaryotic mRNA processing, CSHL Poster: Characterization of a Large Splicing Regulatory Complex
2012	The 17th Annual Meeting of the RNA Society Poster: Characterization of a Large Rbfox Protein Complex

HONORS AND AWARDS

2015-2016	UCLA Dissertation Year Fellowship
2010-2014	Four-year scholarship from China Scholarship Council
2009	National Science Scholar Base Scholarship, China
2009	First Prize, Outstanding Undergraduate Award, Zhejiang University
2009	Original Research Prize, Zhejiang University
2008	Excellent Student Award
2007	National Scholarship, China

CHAPTER 1

INTRODUCTION TO ALTERNATIVE PRE-MRNA SPLICING REGULATION AND RBFOX PROTEIN FAMILY

Alternative pre-mRNA splicing regulation

Alternative pre-mRNA splicing is the major source of proteomic complexity from a limited repertoire of genes in metazoans. Through this process, mRNA isoforms with different coding potential or stability can be generated from one identical pre-mRNA, thus allowing production of proteins with distinct functions or in different quantities. Near all human genes undergo alternative splicing to give rise to diverse protein products (Barbosa-Morais et al., 2012; Merkin et al., 2012).

Pre-mRNA transcripts can undergo many different types of alternative splicing. The most common type is that a cassette exon can be either included or excluded in mRNA. One derivative type is mutually exclusive splicing where one exon or the other is included, but not both in the mRNA. Other types include alternative usage of 5’ terminal exons, 3’ terminal exons and 5' or 3' splice sites within the exons. Finally, the excision of an intron can be suppressed to retain intronic sequence in the mRNA (Black, 2003). Many genes show multiple types of alternative splicing, leading to a complex combination of exons and a large family of related but distinctly encoded proteins.

The recognition of the exon-intron junctions, the removal of the intron, and the subsequent joining of the exons are carried out by a highly dynamic RNA-protein

complex called the spliceosome. Core spliceosomal components, including five small nuclear ribonucleoproteins (snRNPs) and numerous auxiliary proteins, are assembled stepwise onto each intron through recognizing the splice site sequences of a pre-mRNA (Wahl et al., 2009). The 5' and 3' splice sites sequences are located at the two ends of an intron, and the branch point adenosine is usually located ~15 to 50 nucleotides upstream of the 3' splice site, followed by a polypyrimidine tract (Black, 2003). Besides the primary splice site sequences, the splice site choice is modulated by multiple auxiliary regulatory sequences throughout the pre-mRNA. RNA elements that act positively to stimulate spliceosome assembly are called splicing enhancers. Conversely, RNA sequences act as splicing silencers or repressors to block spliceosome assembly. Splicing enhancers and silencers have both exonic and intronic varieties. Often these splicing enhancers and silencers are bound by a large number of regulatory proteins, many of which can directly bind to the pre-mRNA.

Classic regulatory proteins identified include the SR protein and heterogeneous nuclear ribonucleoproteins (hnRNPs) families (Singh and Valcarcel, 2005). Both these families are usually expressed in multiple cell types. Historically, SR proteins are widely viewed as splicing activators that promote exon inclusion by interacting with core spliceosomal proteins, whereas hnRNP proteins are negative regulators that repress exon inclusion. However, global studies of alternative splicing regulation and RNA-protein mapping revealed that some hnRNP proteins act as splicing activators in a context-dependent and position-dependent manner. Thus, the functional distinction

between these two classes of proteins is blurred by discoveries that members of either one can activate or repress splicing in a context-dependent manner (Fu and Ares, 2014). Some tissue-specific RNA-binding proteins were identified as splicing factors, such as MBNL, CELF and Rbfox protein families. They play important roles in establishing a cell-type specific splicing profile to maintain cell identity (Han et al., 2013).

The pre-mRNA sequences carry the “splicing code” that is recognized and decoded by spliceosomal proteins and regulatory splicing factors. To understand the “splicing code”, many different methods have been developed to detect the binding of splicing factors to regions of pre-mRNA transcripts *in vitro* and *in vivo*. SELEX (systematic evolution of ligands by exponential enrichment) is an *in vitro* approach, which allows the identification of high-affinity RNA motifs (Bouvet, 2001). RNA Bind-n-Seq (RBNS) adapted high-throughput SELEX with deep sequencing for quantitative mapping of RNA binding specificity. CLIP-seq (UV crosslinking and immunoprecipitation) captures RNA-protein recruitment sites *in vivo* (Ule et al., 2003). Developments in technologies advance our understanding of sequence specificities of RNA-binding proteins. Nevertheless, assigning binding sites to splicing factor solely based on primary RNA sequence is still very challenging. The majority of RNA-binding proteins bind to very short and partially degenerate RNA sequences, and one RNA sequence can be assigned to multiple RNA-binding proteins. Combinatorial control by multiple splicing factors on one pre-mRNA transcript within a cell makes it more complicated to crack the “splicing code”.

Much remains to be learned about the mechanisms and the regulatory networks of alternative splicing.

Rbfox protein family

Rbfox proteins belong to a family of tissue-specific splicing regulators. They consist? It consists of three members in mammals: Rbfox1 (also known as A2BP1) (Jin et al., 2003; Underwood et al., 2005), Rbfox2 (also known as RBM9) (Underwood et al., 2005), Rbfox3 (also known as NeuN) (Kim et al., 2009). Rbfox1 is highly enriched in the brain, heart and skeletal muscle (Gao et al., 2016; Gehman et al., 2011; Kuroyanagi, 2009; Underwood et al., 2005), whereas Rbfox2 shows a broader expression pattern in multiple cell types, including all these three tissues as well as embryotic stem cells (Gehman et al., 2012; Singh et al., 2014; Wei et al., 2015; Yeo et al., 2009a). Rbfox3 is exclusively expressed in mature neurons (Kim et al., 2009). The Rbfox proteins and particularly Rbfox1 have been associated with many human neurodevelopmental disorders including genetic generalized epilepsy (GGE), childhood focal epilepsy and autism spectrum disorders (Barnby et al., 2005; Bhalla et al., 2004; Martin et al., 2007; Sebat et al., 2007). Central nervous system-specific deletion of Rbfox1 or Rbfox2 in mice results in neurological defects resembling those in human diseases (Gehman et al., 2012; Gehman et al., 2011). While all the Rbfox proteins are expressed in neurons, they exhibit temporally and spatially distinct patterns of expression during neurodevelopment (Gehman et al., 2012), suggesting that they have distinct physiological activities besides

biochemical redundancy at the molecular level. Similarly, Rbfox1 and Rbfox2 showed different physiological functions during heart and muscle development. Rbfox2 is expressed in heart from embryo to adult whereas Rbfox1 is induced in postnatal heart (Kalsotra et al., 2008). As expected, Rbfox2 but not Rbfox1 is required for myoblast differentiation (Singh et al., 2014). When Rbfox1 and Rbfox2 are both expressed in mature heart, cardiac-specific ablation of Rbfox2 causes dilated cardiomyopathy that leads to heart failure (Wei et al., 2015), while Rbfox1 but not Rbfox2 is markedly diminished in pressure-overloaded hearts (Gao et al., 2016), suggesting that they have both overlapping and unique roles. As expressed in other cell types, Rbfox2 has also been shown to establish a splicing program involved in pluripotent stem cell differentiation in cooperation with another splicing factor MBNL1 (Venables et al., 2013b). In addition, Rbfox2 plays an important role in specifying the mesenchymal tissue-specific splicing profiles both in normal and in cancer tissues (Venables et al., 2013a).

All Rbfox proteins contain a single highly conserved RNA recognition motif (RRM) (Auweter et al., 2006), which is identical in Rbfox1 and Rbfox2 and only slightly altered in Rbfox3. Despite the high conservation of RRM, one alternative exon skipping within RRM results in a dominant-negative isoform with decreased RNA-binding affinity, which inhibits Rbfox-dependent splicing activation (Damianov and Black, 2010). Increased expression of this dominant-negative isoform of Rbfox2 has also been found in diabetic hearts at early stages. By interacting with wild-type Rbfox2, dominant-negative isoform of Rbfox2 inhibits the splicing activity of the wild-type protein (Nutter et al., 2016).

The flanking N and C terminal domains that are diversified through the use of alternative promoters and alternative splicing patterns (Damianov and Black, 2010). Different isoforms of Rbfox proteins show different tissue specificity. For instance, two mutually exclusive alternative exons B40 and M43, in the middle of C-terminal domain of Rbfox1 and Rbfox2, are expressed in brain and muscle respectively (Nakahata and Kawamoto, 2005). Variable N and C terminal domains of Rbfox proteins might affect the structure and potential activity of the protein.

Inclusion of exon 19 at the 3' region of Rbfox1 generates a protein isoform ending with the amino acid sequence TALVP, while skipping of this exon results in a FAPY C-tail which is required for proper localization into the nucleus (Lee et al., 2009). Thus, Rbfox1 protein with TALVP tail is predominantly cytoplasmic. It has been reported that by binding to the 3' UTR of mRNA transcripts, cytoplasmic Rbfox1 promotes the stability and/or translation of target transcripts involved in synaptic function, calcium signaling and autism (Lee et al., 2016). Similarly in Drosophila, the cytoplasmic Rbfox proteins regulate the translation of the germ cell maintenance factor *pumilio* by binding to its 3' UTR, which is essential in maintaining normal germ cell differentiation (Carreira-Rosario et al., 2016). One proposed mechanism of regulating mRNA stability and translation by cytoplasmic Rbfox1 is through its competition with microRNA binding at the 3' UTR (Lee et al., 2016).

Unlike many RNA-binding proteins that have very degenerate binding sites, Rbfox proteins bind to RNA element GCAUG with high specificity (Jin et al., 2003). They usually enhance alternative exon inclusion when binding downstream of this exon, while

repressing exon inclusion when binding upstream or within the exon (Jin et al., 2003). Genome-wide analyses by RNA sequencing have revealed hundreds of targets regulated by Rbfox proteins. CLIP (UV crosslinking and immunoprecipitation) coupled with high-throughput sequencing has been used to discover the Rbfox binding sites on the target transcripts. Besides the exon-proximal binding sites of Rbfox proteins which are better studied so far, distal binding sites (>500 nucleotides from any exon) were shown to be active in Rbfox regulation through a conserved long-range RNA-RNA base-pairing interaction (Jangi et al., 2014; Lovci et al., 2013; Weyn-Vanhentenryck et al., 2014; Yeo et al., 2009a). Computational analyses of genome-wide CLIP and expression data revealed splicing regulatory networks controlled by the Rbfox proteins and provide insights of their roles in brain development and autism (Weyn-Vanhentenryck et al., 2014). Rbfox2 cross-regulates alternative splicing events of other RNA-binding proteins to alter their expression by nonsense mediated decay, which further affects the splicing controlled by these RNA-binding proteins. A multilayer splicing regulatory network controlled by Rbfox proteins on top of many other RNA-binding proteins was revealed, thus offering an explanation for how regulatory splicing networks are tuned by the Rbfox proteins (Jangi et al., 2014).

However, the molecular mechanisms that ensure precise regulation of alternative splicing in a position-dependent manner by Rbfox proteins remain poorly understood. Most studies to date have focused on the splicing repression by Rbfox proteins. It has been shown that Rbfox proteins repress the inclusion of calcitonin specific exon 4 by

blocking U2AF65 binding and suppressing exonic splicing enhancer in exon 4. Thus, Rbfox proteins block both the formation of the pre-spliceosomal E complex and its transition to E' complex (Zhou et al., 2007; Zhou and Lou, 2008). Rbfox proteins also repress a different exon in F1γ gene via blocking the formation of pre-spliceosomal E complex when a weak 5' splice site is more preferable (Fukumura et al., 2007; Fukumura et al., 2009). Protein partners of Rbfox proteins have also been examined to illuminate the molecular mechanism of Rbfox regulation. HnRNP H1 and hnRNP F were found to interact with Rbfox2 to repress the exon in FGFR2 minigene by antagonizing the binding of SRSF1 (Mauger et al., 2008). A different study reported that hnRNP H1 works together with RALY and TFG to specifically interact with the C-terminal domain of Rbfox proteins (Sun et al., 2012). However, how these protein partners are involved in Rbfox regulation is not well defined. In *C. elegans,* Rbfox proteins/ASD-1 and SUP-12 proteins cooperatively interact with *egl-15* RNA by sandwiching a G base to form a stable complex where SUP-12 binds to the GUGUG sequence juxtaposed of Rbfox binding site (Kuroyanagi et al., 2007; Kuwasako et al., 2014). Different domains of Rbfox proteins were analyzed for their requirements in Rbfox activity as well. MS2-tethering assays showed that both RRM and C-terminus of Rbfox1 are required for exon repression when tethered to the upstream intron, whereas C-terminus tethered to the downstream intron is sufficient for exon activation (Sun et al., 2012). But the underlying molecular mechanism of the differential requirements of Rbfox domains in splicing activation and repression is not well understood.

The primary goal of this dissertation is to understand the molecular mechanisms by which Rbfox proteins regulate alternative splicing. We started with identification of Rbfox protein partners in the nucleus where Rbfox splicing regulation happens. Strikingly, nearly the entire pool of nuclear Rbfox is associated with the high molecular weight (HMW) material containing the chromatin. Furthermore, in this nuclear fraction the Rbfox proteins participate in an unexpected large protein complex which sediments in the 55S region in glycerol gradients. We call this large multimeric protein complex as a Large Assembly of Splicing Regulators, LASR. By characterizing the subunits of LASR complex and how LASR is involved in Rbfox-mediated splicing regulation, described by chapter 2, we have identified a novel mechanism that Rbfox protein affect splicing of a broader set of exons by interacting with RNA-binding proteins within LASR (Damianov et al., 2016).

References

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

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R.*, et al.* (2012). The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science *338*, 1587-1593.

Barnby, G., Abbott, A., Sykes, N., Morris, A., Weeks, D.E., Mott, R., Lamb, J., Bailey, A.J., Monaco, A.P., and International Molecular Genetics Study of Autism, C. (2005). Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT. American journal of human genetics *76*, 950-966.

Berchowitz, L.E., Kabachinski, G., Walker, M.R., Carlile, T.M., Gilbert, W.V., Schwartz, T.U., and Amon, A. (2015). Regulated Formation of an Amyloid-like Translational Repressor Governs Gametogenesis. Cell.

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

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annual review of biochemistry *72*, 291-336.

Bouvet, P. (2001). Determination of nucleic acid recognition sequences by SELEX. Methods in molecular biology *148*, 603-610.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. Science *324*, 1729-1732.

Calabretta, S., and Richard, S. (2015). Emerging Roles of Disordered Sequences in RNA-Binding Proteins. Trends Biochem Sci *40*, 662-672.

Carreira-Rosario, A., Bhargava, V., Hillebrand, J., Kollipara, R.K., Ramaswami, M., and Buszczak, M. (2016). Repression of Pumilio Protein Expression by Rbfox1 Promotes Germ Cell Differentiation. Developmental cell *36*, 562-571.

Conchillo-Sole, O., de Groot, N.S., Aviles, F.X., Vendrell, J., Daura, X., and Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC bioinformatics *8*, 65.

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

Damianov, A., Ying, Y., Lin, C.H., Lee, J.A., Tran, D., Vashisht, A.A., Bahrami-Samani, E., Xing, Y., Martin, K.C., Wohlschlegel, J.A.*, et al.* (2016). Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. Cell *165*, 606-619.

Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C., Eckmann, C.R., Myong, S., and Brangwynne, C.P. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. Proceedings of the National Academy of Sciences of the United States of America *112*, 7189-7194.

Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet *15*, 689-701.

Fukumura, K., Kato, A., Jin, Y., Ideue, T., Hirose, T., Kataoka, N., Fujiwara, T., Sakamoto, H., and Inoue, K. (2007). Tissue-specific splicing regulator Fox-1 induces exon skipping by interfering E complex formation on the downstream intron of human F1gamma gene. Nucleic acids research *35*, 5303-5311.

Fukumura, K., Taniguchi, I., Sakamoto, H., Ohno, M., and Inoue, K. (2009). U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. Nucleic acids research *37*, 1907-1914.

Gao, C., Ren, S., Lee, J.H., Qiu, J., Chapski, D.J., Rau, C.D., Zhou, Y., Abdellatif, M., Nakano, A., Vondriska, T.M.*, et al.* (2016). RBFox1-mediated RNA splicing regulates cardiac hypertrophy and heart failure. The Journal of clinical investigation *126*, 195-206.

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

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

Han, H., Irimia, M., Ross, P.J., Sung, H.K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N.*, et al.* (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature *498*, 241-245.

Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G.*, et al.* (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell *149*, 768-779.

Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., Uversky, V.N., Vidal, M., and Iakoucheva, L.M. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. Plos Comput Biol *2*, 890-901.

Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., and Parker, R. (2016). ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. Cell *164*, 487-498.

Jangi, M., Boutz, P.L., Paul, P., and Sharp, P.A. (2014). Rbfox2 controls autoregulation in RNA-binding protein networks. Genes & development *28*, 637-651.

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

Kalsotra, A., Xiao, X., Ward, A.J., Castle, J.C., Johnson, J.M., Burge, C.B., and Cooper, T.A. (2008). A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. Proceedings of the National Academy of Sciences of the United States of America *105*, 20333-20338.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J.*, et al.* (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753-767.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome research *12*, 656-664.

Kim, K.K., Adelstein, R.S., and Kawamoto, S. (2009). Identification of neuronal nuclei (NeuN) as Fox-3, a new member of the Fox-1 gene family of splicing factors. The Journal of biological chemistry *284*, 31052-31061.

Kuroyanagi, H. (2009). Fox-1 family of RNA-binding proteins. Cellular and molecular life sciences : CMLS *66*, 3895-3907.

Kuroyanagi, H., Ohno, G., Mitani, S., and Hagiwara, M. (2007). The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo. Molecular and cellular biology *27*, 8612-8621.

Kuwasako, K., Takahashi, M., Unzai, S., Tsuda, K., Yoshikawa, S., He, F., Kobayashi, N., Guntert, P., Shirouzu, M., Ito, T.*, et al.* (2014). RBFOX and SUP-12 sandwich a G base to cooperatively regulate tissue-specific splicing. Nature structural & molecular biology *21*, 778-786.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell *155*, 1049-1060.

Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S.L. (2014). Poly-dipeptides encoded by the C9ORF72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science.

Lee, J.A., Damianov, A., Lin, C.H., Fontes, M., Parikshak, N.N., Anderson, E.S., Geschwind, D.H., Black, D.L., and Martin, K.C. (2016). Cytoplasmic Rbfox1 Regulates the Expression of Synaptic and Autism-Related Genes. Neuron *89*, 113-128.

Lee, J.A., Tang, Z.Z., and Black, D.L. (2009). An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. Genes & development *23*, 2284-2293.

Li, H., Qiu, J., and Fu, X.D. (2012a). RASL-seq for massively parallel and quantitative analysis of gene expression. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Chapter 4*, Unit 4 13 11-19.

Li, P.L., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F.*, et al.* (2012b). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336-U129.

Lin, Y., Protter, D.S., Rosen, M.K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Molecular cell *60*, 208-219.

Lovci, M.T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T.Y., Stark, T.J., Gehman, L.T., Hoon, S.*, et al.* (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. Nature structural & molecular biology *20*, 1434-1442.

Maekawa, S., Leigh, P.N., King, A., Jones, E., Steele, J.C., Bodi, I., Shaw, C.E., Hortobagyi, T., and Al-Sarraj, S. (2009). TDP-43 is consistently co-localized with ubiquitinated inclusions in sporadic and Guam amyotrophic lateral sclerosis but not in familial amyotrophic lateral sclerosis with and without SOD1 mutations. Neuropathology *29*, 672-683.

Martin, C.L., Duvall, J.A., Ilkin, Y., Simon, J.S., Arreaza, M.G., Wilkes, K., Alvarez-Retuerto, A., Whichello, A., Powell, C.M., Rao, K.*, et al.* (2007). Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics *144B*, 869-876.

Mauger, D.M., Lin, C., and Garcia-Blanco, M.A. (2008). hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. Molecular and cellular biology *28*, 5403-5419.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. Science *338*, 1593-1599.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. Cell *163*, 123-133.

Murakami, T., Qamar, S., Lin, J.Q., Schierle, G.S., Rees, E., Miyashita, A., Costa, A.R., Dodd, R.B., Chan, F.T., Michel, C.H.*, et al.* (2015). ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. Neuron *88*, 678-690.

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

Neelamraju, Y., Hashemikhabir, S., and Janga, S.C. (2015). The human RBPome: From genes and proteins to human disease. J Proteomics *127*, 61-70.

Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D.*, et al.* (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Molecular cell *57*, 936-947.

Nutter, C.A., Jaworski, E.A., Verma, S.K., Deshmukh, V., Wang, Q., Botvinnik, O.B., Lozano, M.J., Abass, I.J., Ijaz, T., Brasier, A.R.*, et al.* (2016). Dysregulation of RBFOX2 Is an Early Event in Cardiac Pathogenesis of Diabetes. Cell reports.

Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M.*, et al.* (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell *162*, 1066-1077.

Patel, S.S., Belmont, B.J., Sante, J.M., and Rexach, M.F. (2007). Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. Cell *129*, 83-96.

Schwartz, J.C., Wang, X., Podell, E.R., and Cech, T.R. (2013). RNA seeds higher-order assembly of FUS protein. Cell reports *5*, 918-925.

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

Singh, R., and Valcarcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. Nature structural & molecular biology *12*, 645-653.

Singh, R.K., Xia, Z., Bland, C.S., Kalsotra, A., Scavuzzo, M.A., Curk, T., Ule, J., Li, W., and Cooper, T.A. (2014). Rbfox2-coordinated alternative splicing of Mef2d and Rock2 controls myoblast fusion during myogenesis. Molecular cell *55*, 592-603.

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

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

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212-1215.

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

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P.*, et al.* (2009). Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. Science *323*, 1208-1211.

Venables, J.P., Brosseau, J.P., Gadea, G., Klinck, R., Prinos, P., Beaulieu, J.F., Lapointe, E., Durand, M., Thibault, P., Tremblay, K.*, et al.* (2013a). RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. Molecular and cellular biology *33*, 396-405.

Venables, J.P., Lapasset, L., Gadea, G., Fort, P., Klinck, R., Irimia, M., Vignal, E., Thibault, P., Prinos, P., Chabot, B.*, et al.* (2013b). MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. Nature communications *4*, 2480.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wang, J.T., Smith, J., Chen, B.C., Schmidt, H., Rasoloson, D., Paix, A., Lambrus, B.G., Calidas, D., Betzig, E., and Seydoux, G. (2014). Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically-disordered proteins in C-elegans. eLife *3*.

Wei, C., Qiu, J., Zhou, Y., Xue, Y., Hu, J., Ouyang, K., Banerjee, I., Zhang, C., Chen, B., Li, H.*, et al.* (2015). Repression of the Central Splicing Regulator RBFox2 Is Functionally Linked to Pressure Overload-Induced Heart Failure. Cell reports.

Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q.*, et al.* (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell reports *6*, 1139-1152.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009a). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nature structural & molecular biology *16*, 130-137.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009b). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nature structural & molecular biology *16*, 130-137.

Zhou, H.L., Baraniak, A.P., and Lou, H. (2007). Role for Fox-1/Fox-2 in mediating the neuronal pathway of calcitonin/calcitonin gene-related peptide alternative RNA processing. Molecular and cellular biology *27*, 830-841.

Zhou, H.L., and Lou, H. (2008). Repression of prespliceosome complex formation at two distinct steps by Fox-1/Fox-2 proteins. Molecular and cellular biology *28*, 5507-5516.

CHAPTER 2

RBFOX PROTEINS REGULATE SPLICING AS PART OF A LARGE MULTIPROTEIN COMPLEX LASR

Chapter 2 is a reprint of a manuscript with permission from Elsevier and Copyright Clearance Center.

Damianov, A., Ying, Y., Lin, C.H., Lee, J.A., Tran, D., Vashisht, A.A., Bahrami-Samani, E., Xing, Y., Martin, K.C., Wohlschlegel, J.A.*, et al.* (2016). Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. Cell *165*, 606-619.

Copyright permission:

This Agreement between Yi Ying ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Licensee: Yi Ying

License Date: Aug 15, 2016

License Number: 3930061480656

Publication: Cell

Title: Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR

Type Of Use: reuse in a thesis/dissertation

SUMMARY

Rbfox proteins control alternative splicing and post-transcriptional regulation in mammalian brain and are implicated in neurological disease. These proteins recognize the RNA sequence (U)GCAUG, but their structures and diverse roles imply a variety of protein-protein interactions. We find that nuclear Rbfox proteins are bound within a large assembly of splicing regulators (LASR), a multimeric complex containing the proteins hnRNP M, hnRNP H, hnRNP C, Matrin3, NF110/NFAR-2, NF45, and DDX5, all approximately equimolar to Rbfox. We show that splicing repression mediated by hnRNP M is stimulated by Rbfox. Virtually all the intron-bound Rbfox is associated with LASR, and hnRNP M motifs are enriched adjacent to Rbfox crosslinking sites in vivo. These findings demonstrate that Rbfox proteins bind RNA with a defined set of cofactors and affect a broader set of exons than previously recognized. The function of this multimeric LASR complex has implications for deciphering the regulatory codes controlling splicing networks.

INTRODUCTION

Patterns of alternative pre-mRNA splicing are controlled by specialized proteins that assemble onto the pre-mRNA and control splice site choices (Fu and Ares, 2014; Lee and Rio, 2015). The Rbfox family of splicing regulators includes three mammalian paralogs Rbfox1 (A2BP1), Rbfox2 (RBM9), and Rbfox3 (NeuN) that are all expressed in neurons and can show specific expression in other cell types (Kuroyanagi, 2009).

Rbfox proteins affect a wide range of synaptic and neurodevelopmental functions. Conditional Rbfox1 deletion in the mouse brain leads to a seizure phenotype, whereas mice without brain Rbfox2 exhibit cerebellar defects and ataxia (Gehman et al., 2011, 2012). Mutations in Rbfox1 and Rbfox3 are found in human epilepsy patients (Bhalla et al., 2004; Lal et al., 2013a, 2013b), whereas other findings connect Rbfox1 with autism spectrum disorders and spinocerebellar ataxias (Bill et al., 2013). Rbfox2 is also upregulated in certain cancers and controls exons during the epithelial-mesenchymal transition (Braeutigam et al., 2014; Venables et al., 2013). These findings have led to substantial clinical interest in Rbfox protein function.

The Rbfox proteins all contain a single highly conserved RNA recognition motif (RRM) that specifically binds the sequences UGCAUG and GCAUG (Auweter et al., 2006; Jin et al., 2003; Lambert et al., 2014). Alternative promoters and alternative splicing produce multiple protein isoforms from each Rbfox locus, with varying subcellular localization and splicing activity (Damianov and Black, 2010; Lee et al., 2009; Nakahata and Kawamoto, 2005). In addition to controlling splicing patterns, cytoplasmic Rbfox isoforms bind to 3′ UTR targets to affect downstream gene expression (Lee et al., 2016). Target transcripts for both nuclear and cytoplasmic proteins encode many proteins essential to neuronal development and synaptic activity (Gehman et al., 2011, 2012; Lee et al., 2016; Lovci et al., 2013; Weyn-Vanhentenryck et al., 2014).

Typically, binding of Rbfox to a (U)GCAUG element downstream of the alternative exon promotes its splicing, whereas binding to an upstream element, or an element within the exon, represses exon inclusion (Jangi et al., 2014; Lovci et al., 2013; Tang et al., 2009; Weyn-Vanhentenryck et al., 2014; Yeo et al., 2009; Zhang et al., 2008). However, many (U)GCAUG elements proximal to alternative exons do not affect splicing or exhibit Rbfox binding. Conversely, (U)GCAUG elements located more than 500 nucleotides away from a target exon can function through base-pairing interactions that bring the bound protein closer to the exon (Lovci et al., 2013). Moreover, some Rbfox binding sites identified in genome-wide assays do not encompass a (U)GCAUG element. Thus, the features determining whether and how Rbfox will affect splicing are not understood.

We find that nuclear Rbfox proteins function within a large macromolecular complex containing a distinct set of other splicing factors that affect the recruitment of Rbfox to its targets.

RESULTS

Rbfox Proteins Engage in Distinct Protein and RNA Interactions in Different Nuclear Compartments

To examine the portion of Rbfox protein engaged with nascent RNA, we isolated nuclei from brain tissue followed by lysis in Triton X-100. The majority of Rbfox proteins pelleted with

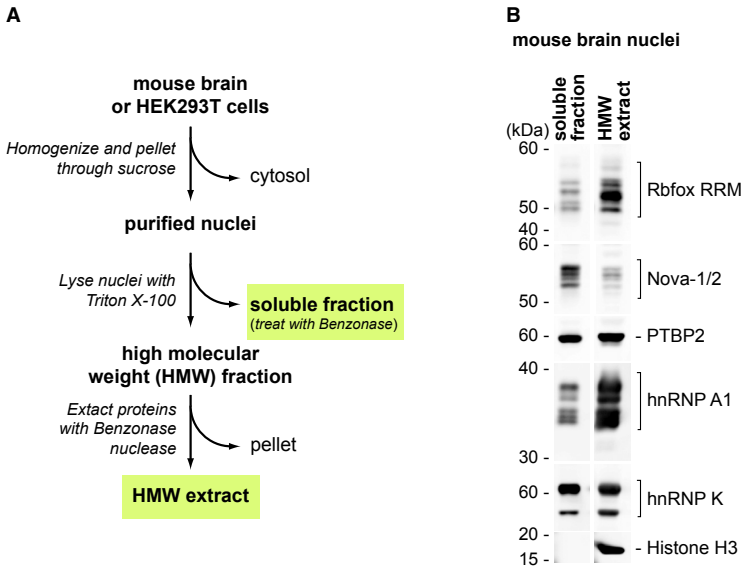


Figure 1. Rbfox Proteins Are Found in the HMW Nuclear Fraction

(A) Preparation of the soluble and HMW nuclear fractions.

(B) Immunoblot analysis of soluble and HMW fractions from mouse brain. Rbfox1, Rbfox2, and Rbfox3 were detected with an antibody recognizing their nearly identical RRMs.

See also Figure S1.

chromatin and other high-molecular-weight (HMW) material, with less than 10% of the protein present in the soluble nuclear fraction (Figures 1A and 1B). Rbfox proteins could be extracted from the pellet fraction using RNase and more efficiently with Benzonase nuclease that cleaves both RNA and DNA (Figures 1A and 1B) and which removed more than 99% of the bulk RNA from the fraction. We used this method to prepare subnuclear extracts from both mouse brain and cultured cells. Note that this extraction differs from methods to isolate nascent RNA, in which the chromatin is pelleted in high salt and urea to eliminate many intermolecular interactions while maintaining RNA polymerase association with DNA and nascent RNA (Khodor et al., 2011; Pandya-Jones and Black, 2009; Wuarin and Schibler, 1994). Our intent was to preserve protein-protein interactions while eliminating interactions mediated by RNA. The pellet fraction contains HMW material that is spun down with chromatin but is not necessarily in direct interaction with it. The release upon nuclease treatment indicates that a large portion of Rbfox protein is associated with RNA (see below).

Different types of splicing regulators exhibited different partitioning between the HMW fraction and the soluble nucleoplasm (Figures 1B and S1). Similar to Rbfox, the majority of hnRNPs A1, A2/B1, and Q/R, as well as the SR proteins, were found in the HMW pellet and released with nuclease (Figures 1B and S1; in Figure S1, compare the input lanes). These proteins are enriched in nuclear speckles, indicating that nuclear speckle material also pelleted with the HMW fraction (Misteli et al., 1998; Tripathi et al., 2012). Unlike Rbfox, splicing factors such as PTBP2 and hnRNP K showed approximately equal distribution between the HMW and soluble fractions. Other proteins such as Nova were enriched in the soluble nuclear fraction (Figures 1B and S1).

To examine Rbfox/RNA interactions in these two nuclear compartments, we performed individual nucleotide resolution cross-linking immunoprecipitation (iCLIP) on fractions of adult mouse brain (König et al., 2010). UV irradiation of triturated brain tissue did not alter the location of the Rbfox proteins (Figure S2A). In vivo crosslinking data were generated for Rbfox1, Rbfox2, and Rbfox3 in the HMW fractions of adult mouse forebrain and hindbrain and for Rbfox1 in the soluble nuclear fraction of these regions. Crosslinked Rbfox1, 2, and 3 were also isolated from the soluble nuclear fraction of whole mouse brain (Table S1A). Crosslinked sequences were aligned to the mouse genome, and significant clusters of iCLIP tags were defined (König et al., 2010). Crosslinking commonly generates an iCLIP tag terminating one nucleotide downstream. This position was frequently, but not always, within a (U)GCAUG element (see below). The identified clusters overlapped with previous maps of Rbfox binding from unfractionated brain tissue (Lovci et al., 2013; Weyn-Vanhentenryck et al., 2014; Figures S2B and S2C). iCLIP clusters for Rbfox1, 2, and 3 were found adjacent to the majority of alternative exons whose splicing was altered by brain-specific deletion of the Rbfox1 or Rbfox2 genes (Gehman et al., 2011, 2012). The overlap between Rbfox1, 2, and 3 binding sites supports the hypothesis that functional redundancy reduces the magnitude of splicing changes in Rbfox knockout mice.

The HMW and soluble fractions differed dramatically in the positions of Rbfox binding. In other cells, most unspliced RNA was found in the HMW pellet even after stringent wash conditions (Bhatt et al., 2012; Khodor et al., 2011; Pandya-Jones and Black, 2009). In agreement with this, the majority of Rbfox binding in the soluble fraction from brain was in 3′ UTR regions (Figure 2A), whereas 90% of crosslinking events from the HMW fraction were in introns, indicating predominant Rbfox association with unspliced RNA in this compartment. The patterns of Rbfox binding are illustrated on the *Snap25* transcript (Figure 2B), which contains a pair of Rbfox-regulated mutually exclusive exons (Gehman et al., 2011, 2012; Johansson et al., 2008). In the HMW fraction, the majority of crosslinking events were in the intron downstream of the two exons, with the 3′ UTR showing only one prominent Rbfox-binding cluster. In the soluble fraction, intron clusters were largely absent (in keeping with the spliced structure of the transcript), whereas the 3′ UTR showed a broadly distributed set of clusters (Figure 2B). Similar patterns of binding were seen for multiple other transcripts (see Supplemental Information). The 3′ UTR binding in the soluble nuclear fraction

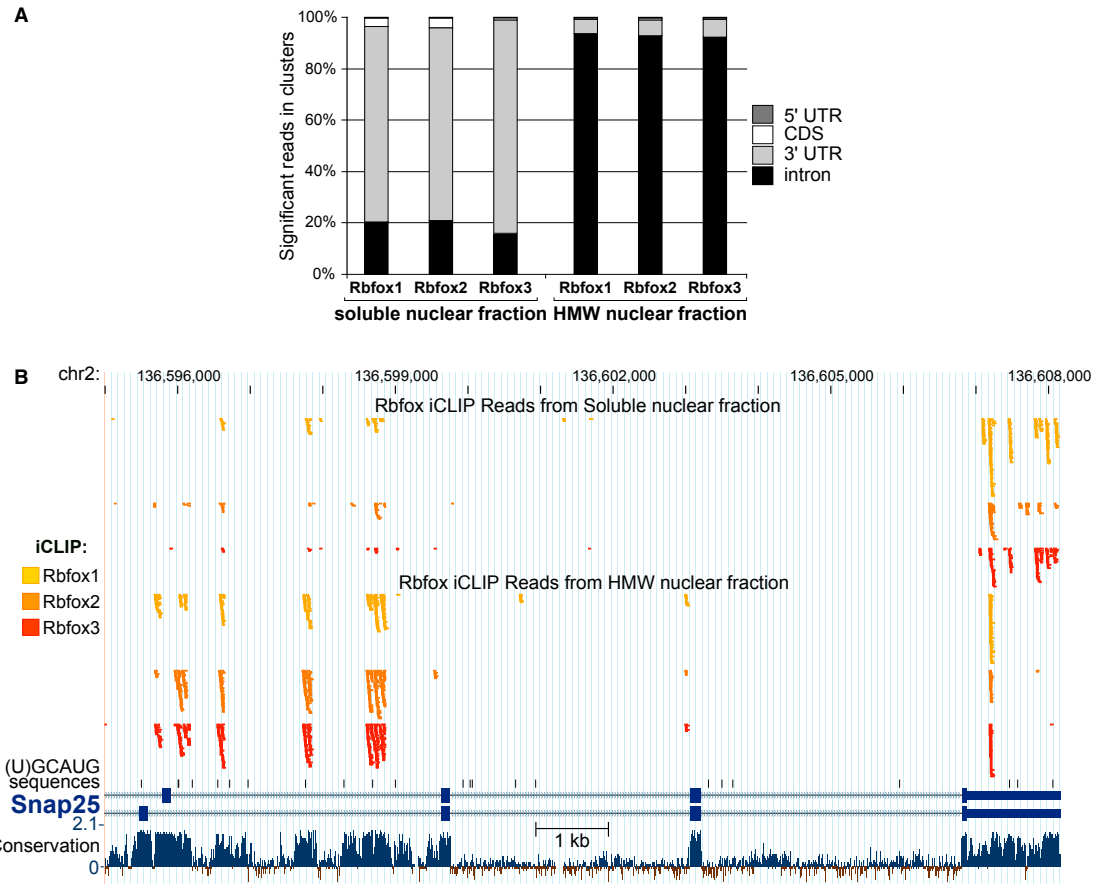


Figure 2. Rbfox Recruitment to Introns in the HMW Fraction and to 3′ UTRs in the Soluble Nuclear Fraction

(A) Distribution of Rbfox iCLIP tags in 5′ UTRs, CDS, 3′ UTRs, and introns. Data for mouse brain iCLIP clusters of Rbfox1, 2, and 3 with width ≥2 nt are shown. See Table S1 and Figure S2 for additional detail.

(B) Genome browser view of iCLIP reads mapped to the 3′ portion of the *Snap25* gene. iCLIP tracks from the soluble nuclear fraction of whole brain and the cerebellar HMW fraction are aligned as indicated. Significant iCLIP reads from Rbfox1, 2, and 3 are colored as indicated. GCAUG motifs are shown below. See Data S1 for additional examples.

was similar to that seen in the cytoplasm (Lee et al., 2016) and may represent processed mRNA that is not yet exported from the nucleus. RT-PCR measurements indicated that, unlike the limited intron crosslinking in the soluble fraction, the lack of binding was not due to reduced amounts of 3′ UTR RNA in the HMW fraction (data not shown). The 3′ UTR of *Snap25* mRNA was present in both fractions, but not bound by Rbfox until release to the soluble pool. The patterns of crosslinking indicate that more than simple (U)GCAUG recognition determines Rbfox binding to RNA.

The Nuclear Rbfox Isoforms Are Subunits of a Much Larger Complex of Proteins

Immunoprecipitation (IP) and mass spectrometry indicated that the Rbfox proteins specifically associated with other proteins in the brain HMW fraction (data not shown), but characterization of these complexes required additional purification. To examine Rbfox interactions in detail, we generated HEK293 cell lines stably expressing N-terminally Flag-tagged Rbfox1, 2, or 3 (Figure 3A). Each cell line contained a single flipped-in Rbfox transgene expressing physiological or lower levels of protein compared to brain. The Rbfox proteins exhibited the same enrichment in the HMW nuclear fraction of these cells as in neurons (see below). We immunopurified the Rbfox proteins from this fraction. Elution with Flag peptide yielded a defined set of copurifying proteins seen on Coomassie-stained gels at nearly equal stoichiometry to the tagged Rbfox protein (Figure 3A). The pattern of copurifying bands was nearly identical for Rbfox1, 2, and 3. Multidimensional protein identification technology

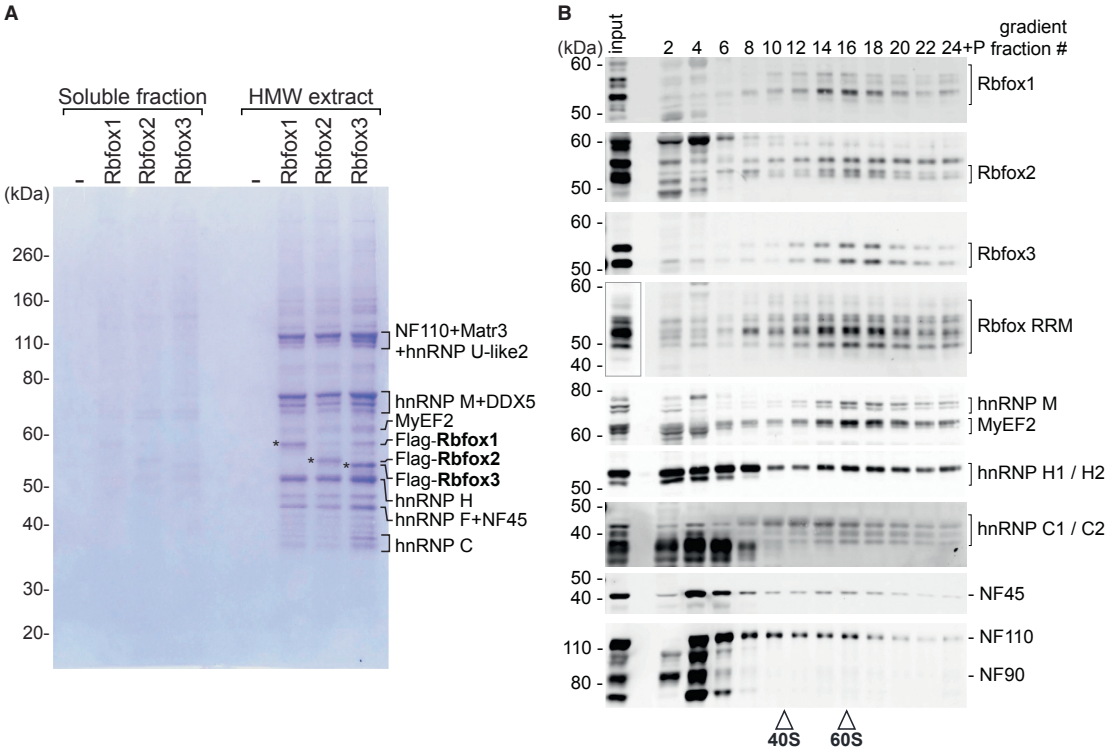


Figure 3. Rbfox Proteins Coprecipitate from the HMW Fraction with a Distinct Set of Other RNA-Binding Proteins

(A) Immunopurification of Rbfox proteins from HEK293 nuclear fractions. Soluble and HMW nuclear extracts were prepared from cells stably expressing Flag-tagged Rbfox1, 2, or 3. Proteins were immunoprecipitated with anti-Flag, eluted with Flag peptide, resolved on SDS-PAGE, and stained with SimplyBlue. The major interacting proteins are indicated on the right. Flag-Rbfox bands are indicated by asterisks. (-) denotes nuclear fractions from the parental HEK293 cell line that does not express Flag-tagged protein.

(B) Sedimentation of protein complexes from mouse brain HMW extract through 10%–50% glycerol gradients. Gradient fractions from top to bottom run from left to right. 40S and 60S markers from a parallel gradient are indicated below. Rbfox1, Rbfox2, Rbfox3 and their binding partners are indicated on the right. MyEF2 and hnRNP M are detected with common antibody, as are the *ILF3* gene products NF110 and NF90. The boxed area indicates a lower exposure of the same gel to reduce the band intensities in those lanes.

See Figures S1 for additional data.

(MuDPIT) and immunoblots identified the major interacting partners (Figures 3A and S3A). Two splicing factors, hnRNP M and hnRNP H (which was previously shown to bind to Rbfox2; Mauger et al., 2008), were present at slightly higher stoichiometry to Rbfox. Proteins at near equal amounts to Rbfox included matrin3, NF110/NFAR-2, NF45, and the DEAD-box helicase DDX5/p68. Somewhat lower levels of hnRNP U-like-2 and hnRNP C were also present, as were relatives of the above proteins including MyEF2 (an hnRNP M homolog), hnRNP F, and DDX17/p72 (Figures 3A and S3A). The same set of proteins was found to associate with Rbfox proteins in mouse brain (Figure 3B; data not shown). Numerous other RNA-binding proteins and splicing factors were absent from the Rbfox purifications, including the major SR proteins, hnRNP A1/B2, A2/B1, R, Q, K, PTBP1, and others (Figure S3A; data not shown).

Some interactions with Rbfox were specific to particular isoforms. NF110/NFAR-2, the largest product of the *ILF3* gene, has an extended C terminus absent from the more-abundant NF90/NFAR-1 isoform (Saunders et al., 2001). Notably, NF90/NFAR-1 did not copurify with Rbfox proteins (Figure S3A), implicating the C-terminal domain of NF110 in this interaction (Reichman and Mathews, 2003). Similarly, hnRNP H1 and H2 copurify with Rbfox more efficiently than the homologous hnRNP F (Figures 3A and S3A), whereas, for hnRNP M and hnRNP C, family members bound with equal efficiency (Figures 3A and 3B). The Rbfox-interacting proteins were not seen in other immunoprecipitates, such as PTBP1, from brain or HEK293 cells. The strikingly similar stoichiometry of the copurifying proteins, their isoform specificity, and the absence of other RNA-binding proteins in the isolate indicated that Rbfox bound with a specific

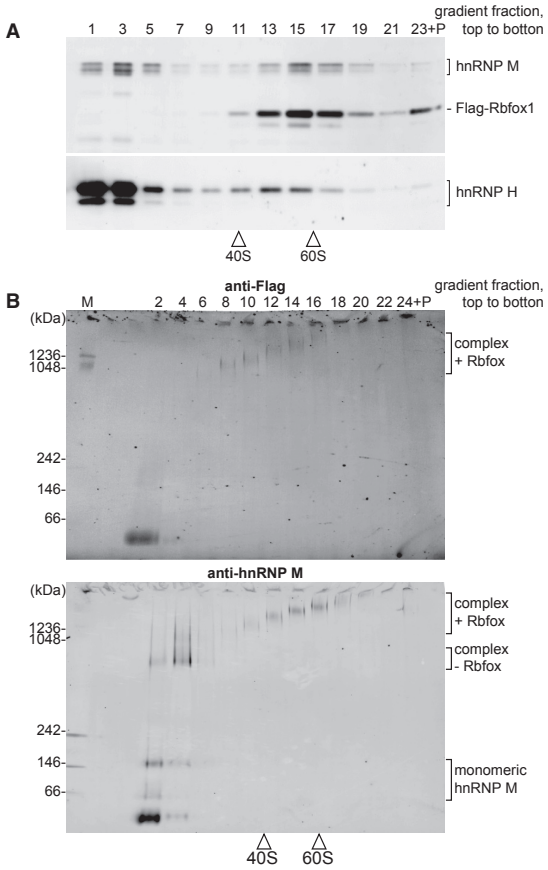


Figure 4. The 55S Rbfox Complex Is Present in HEK293 Cells and Is Heterogeneous in Size

(A) Gradient sedimentation of HMW extract from HEK293T cells transiently expressing Flag-tagged Rbfox1. Proteins detected by immunoblot are indicated as in Figure 3B.

(B) Native gel analysis of complexes separated on glycerol gradients. HMW extracts from HEK293 cells stably expressing Flag-tagged Rbfox3 were fractionated, and protein complexes were resolved by native PAGE and probed by immunoblot. Flag-Rbfox3 (top) and hnRNP M (bottom) are indicated on the right.

See Figure S3 for further analyses.

set of interactors and was not copurifying with these proteins via a common interaction with RNA.

Because multiple proteins were isolated in amounts equal to the tagged Rbfox, it appeared that the Rbfox proteins associated with a single complex, rather than making multiple independent interactions. To examine this, we immunoprecipitated Flag-tagged hnRNP M and hnRNP H proteins transiently expressed in HEK293T cells (Figures S3B and S3C). These overexpressed proteins distributed between the soluble nucleoplasm and the HMW fraction. In the HMW, but not the soluble nuclear fraction, both hnRNP M and H copurified with the same set of proteins found with Flag-Rbfox. These proteins were again isolated at similar stoichiometry to each other, although in this case less abundantly than the overexpressed hnRNP M or H (Figures S3B and S3C). HnRNP M interacted with two additional proteins SRSF14 and hnRNP Q/R. The proteins copurified with hnRNP H or M in the absence of coexpressed Rbfox protein, but when HA-tagged Rbfox3 was transfected, this protein was isolated with Flag-hnRNP M (Figure S3B). These results indicate that the Rbfox proteins in the HMW fraction associate with a defined complex of other proteins that can assemble without Rbfox.

To examine the size of the Rbfox complexes, we separated the soluble and HMW nuclear fractions from mouse brain by sedimentation through glycerol density gradients (Figure 3B). Interestingly, the majority of all three Rbfox proteins in the HMW fraction sedimented as a very large protein complex with an average size of 55S (Figure 3B; see the anti-Rbfox RRM panel). The abundant proteins copurifying with Flag-Rbfox all showed cosedimenting peaks, including hnRNP M/MyEF2, hnRNP H, hnRNP C, NF110, and NF45 (Figure 3B). Unlike Rbfox, where the majority of the protein was 55S in size, portions of the other proteins sedimented as likely monomers (fractions 1 and 2) and smaller complexes (fractions 4–6). In the soluble nuclear fraction, these proteins also sedimented as free proteins and small complexes at the top of the gradient (data not shown). Numerous other splicing factors, including hnRNP A1, hnRNP K, Nova, PTBP2, the SR proteins, and others that did not copurify with Rbfox proteins, did not sediment in the 55S fraction but were found as free proteins or complexes up to about 20S in size (Figure S1). Notably NF90, containing the same N-terminal double-stranded RNA-binding domains as NF110 but lacking its C-terminal domain, failed to cosediment with Rbfox just as it did not copurify with the Flag-Rbfox proteins (Figure 3B).

We also examined gradients of the HMW fraction from HEK293 cells, probing for Rbfox1 and its interacting proteins hnRNP H and M. Similar to brain, all the Flag-Rbfox1 was found at ~55S, accompanied by peaks of hnRNP H and M (Figure 4A). Some hnRNP H and M were also present as smaller complexes and apparent monomers (Figure 4A), and only these smaller forms were seen in the soluble nuclear fraction (Figure S3D). The 55S peak of hnRNP H and M was also observed in HEK293 cells not expressing an Rbfox transgene (data not shown). Because these cells express low levels of Rbfox2, it is not clear whether Rbfox is responsible for the high S value of the H and M proteins or whether additional complexes assemble from these proteins. The data indicate that nearly the entire nuclear pool of Rbfox proteins is associated with a large nuclease-resistant complex of proteins in both the brain and HEK293 cells.

The total mass of one copy of each protein in the flag eluate is not sufficient to yield a 55S complex, which is expected to be in the megadalton range. To further assess the size and heterogeneity of these protein complexes, we analyzed each fraction across the gradient by native protein gel and immunoblot to identify Flag-Rbfox- and hnRNP M-containing species. Probing these gels with anti-Flag antibody identified slowly migrating complexes in the fractions containing Rbfox protein (Figure 4B, top). Interestingly, these complexes increased in size in fractions of successively higher S value. This indicates that the Rbfox complexes are heterogeneous, with an average size of 55S.

Probing the gel with anti-hnRNP M antibody labeled the same complexes containing Rbfox, a small species in fraction 2 that is likely monomeric hnRNP M, and a complex lacking Rbfox in gradient fractions 2–6 (Figure 4B, bottom). This complex may be a precursor to the larger 55S population of complexes containing Rbfox.

Taken together, these data indicate that the Rbfox proteins assemble with a large complex of proteins containing hnRNP M, hnRNP H, Matrin3, hnRNP U-like-2, hnRNP C, NF110, NF45, and DDX5/17. We call this complex a large assembly of splicing regulators (LASR). We next examined how the interaction of Rbfox with LASR might affect splicing.

Rbfox Can Repress Splicing through the Binding Element for Another LASR Component hnRNP M

The iCLIP and biochemical analyses together indicate that the intron-bound Rbfox protein is all associated with LASR and thus likely regulates splicing as a part of this assembly. LASR components are also found as free proteins and presumably have functions in addition to their role in LASR. The LASR subunit hnRNP M regulates splicing during the epithelial mesenchymal transition (EMT) (Hovhannisyan and Carstens, 2007; Xu et al., 2014). Interestingly, the activity, but not the expression, of hnRNP M was found to change during EMT, whereas Rbfox2 is upregulated during this transition (Venables et al., 2013; Xu et al., 2014). To test the effect of Rbfox on hnRNP M splicing activity, we created a three exon minigene based on DUP51EK (Amir-Ahmady et al., 2005), where the second exon contains an hnRNP M consensus binding motif, UGGUGGUG, as defined by CLIP analysis (Huelga et al., 2012). We mutated other potential hnRNP M binding sites and all GCAUG motifs to create DUP-51M1. An equivalent minigene carried a mutation in the hnRNP M site (DUP-51ΔMsite; Figure 5A). In gel shift assays, purified hnRNP M bound this exon, but not the mutant exon, whereas purified Rbfox2 did not bind either the wild-type or mutant exon (Figures S4A–S4C). Depletion of hnRNP M by RNAi stimulated splicing of DUP-51M1 exon 2, as did mutation of the M site, confirming that hnRNP M acts as a silencer of the exon (Figure S4D). We also found that hnRNP M crosslinked to DUP-51M1 pre-mRNA in vivo and that this crosslinking is reduced by the M site mutation (Figure S4E).

DUP51M1 and DUP-51ΔMsite were transfected into HEK293T cells, with and without an Rbfox3 expression plasmid. Coexpression of Rbfox3 strongly inhibited target exon splicing of DUP-51M1 but had minimal effect on DUP-51ΔMsite (Figure 5B). Thus, Rbfox can alter the splicing of an exon containing an hnRNP M site. The in vivo association of Rbfox3 with this transcript was confirmed by crosslinking, anti-Flag IP, and RT-PCR of the pre-mRNA (Figure 5C). The DUP-51M1 transcript was readily detected in the RNA crosslinked to Rbfox3, and this was strongly reduced by mutation of the M site (Figure 5C, bottom panel). GAPDH RNA, which crosslinked at moderate levels to the overexpressed Rbfox3, was used to normalize band intensities. The M site mutation did not reduce the level of precursor RNA to cause the reduced Rbfox binding (data not shown). The crosslinking of Rbfox3 indicated close proximity to the reporter pre-mRNA that was dependent on the M site.

To generally assess the effect of Rbfox proteins on hnRNP M activity, we profiled exon inclusion across the transcriptome in a modified HEK cell line where the low level of endogenous Rbfox2 was eliminated by CRISPR/Cas9 deletion of an early exon in the Rbfox2 gene (data not shown). Rbfox1 was introduced at the flip-in locus to yield two cell lines plus and minus Rbfox1 expression (Figure 6). HnRNP M could be efficiently depleted from both lines by RNAi without affecting Rbfox1 expression (Figure 6A). Profiling alternative exon use in the four conditions in triplicate by RNA-seq and rMATS (Tables S1B and S4; Shen et al., 2014), we identified hnRNP M-dependent exons in the presence of Rbfox1 and Rbfox1-dependent exons in the presence of hnRNP M (Figures 6B and 6C). We found that the effect of hnRNP M on both exon activation and exon repression was strikingly reduced when Rbfox1 was absent (Figure 6B). Of 225 exons whose inclusion was increased or decreased by M depletion in Rbfox1-expressing cells, 144 showed a 15% or greater reduction in this effect in the absence of Rbfox1. To control for bias arising from exon selection, we carried out the reverse analysis of hnRNP M-dependent exons defined in the absence of Rbfox1. A significantly smaller fraction of these exons was sensitive to Rbfox1 (two-sided p value $< 2.0 \times 10^{-5}$ by Fisher's exact test). Thus, the shift in splicing observed in Figure 6B was not due to biased exon selection. These results confirmed that hnRNP M activity was affected by Rbfox.

Reciprocal analyses examined the dependence of exons on Rbfox1 in the presence and absence of hnRNP M (Figure 6C). RNAi depletion of hnRNP M was not as stringent as Rbfox knockout, leaving about 10% of the endogenous hnRNP M and all of the possibly redundant paralog MyEF2 (data not shown). Nevertheless, of 726 Rbfox1-dependent exons identified in the hnRNP M-expressing cells, 329 were less affected by Rbfox1 when hnRNP M was absent (defining the same parameters as Figure 6B). We found that depletion of other LASR components often inhibited cell growth and/or altered expression of Rbfox proteins. Testing the effects of these proteins on Rbfox activity will likely require different approaches.

The Binding of Rbfox to Its Target Element Is Affected by Adjacent Sequence Motifs

To assess whether LASR components affect where Rbfox binds, we analyzed sites of Rbfox crosslinking for enriched binding motifs. Among the LASR subunits, hnRNP M binds to GU-rich motifs (Huelga et al., 2012), hnRNP C to polyU sequences (Görlach et al., 1994; König et al., 2010; Swanson and Dreyfuss, 1988), and hnRNP H to polyG and GGGA sequences (Caputi and Zahler, 2001; Chou et al., 1999; Huelga et al., 2012). We defined high-confidence binding sites as intronic sequences containing overlapping iCLIP clusters for all three Rbfox proteins in either the forebrain or hindbrain HMW fractions. These regions were analyzed for pentamer frequencies (see Supplemental Experimental Procedures; Table S2). Based on the previous CLIP analyses (Huelga et al., 2012), we defined possible hnRNP M-binding motifs as all pentamers containing three Gs and two Us or three Us and two Gs but without more than two Gs or Us in a row. This definition includes all the described binding pentamers, but it is likely that some sequences within this group bind hnRNP M better than others. HnRNP C pentamers included U_5 and all pentamers containing four continuous U nucleotides. Similarly, hnRNP H pentamers included G_5 and all pentamers

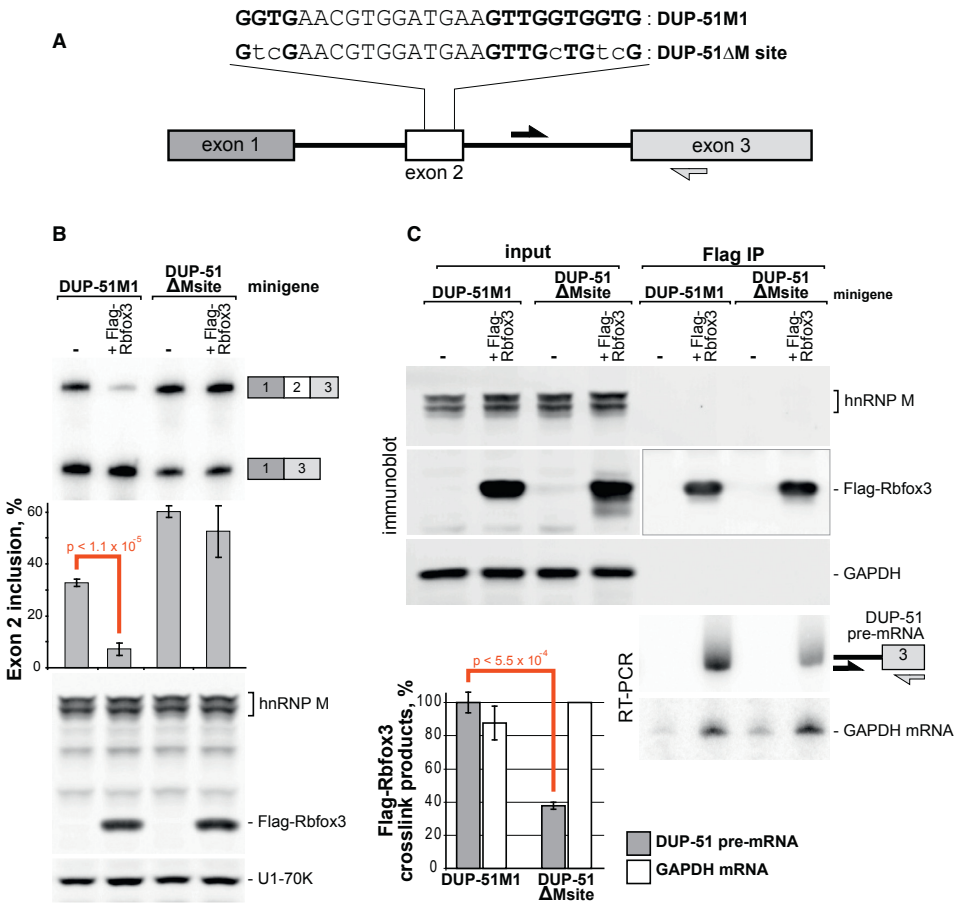


Figure 5. Rbfox3 Can Regulate Alternative Splicing through an hnRNP M Binding Site

(A) Diagram of the minigene DUP-51M1 and its mutant DUP-51ΔMsite. The M binding site is in bold. Arrows indicate primers used to detect DUP-51 pre-mRNA.

(B) DUP-51M1 or DUP-51ΔMsite were transfected into HEK293T cells with control vector (-) or Flag-Rbfox3 expression vector. Exon 2 splicing was measured by RT-PCR with primers in the flanking exons. The spliced products are indicated (top). Average exon inclusion with SD from four experiments is quantified below. Rbfox3 expression caused a 4.6-fold decrease in DUP-51M1 exon 2 splicing. Statistical significance (red) was measured by unpaired, two-tailed, unequal variance Student's t test. HnRNP M, Flag-tagged Rbfox3, and U1-70K, as a loading control, are indicated (bottom).

(C) As in (B), cells expressing DUP-51 minigenes and Flag-Rbfox3 were UV irradiated in vivo and lysed under denaturing conditions to prevent copurification of hnRNP M. RNA:protein crosslinks were immunoprecipitated with anti-Flag. HnRNP M, Flag-Rbfox3, and GAPDH in the lysates (lanes: input) and immunoprecipitates (lanes: Flag IP) were measured by immunoblot (top). The boxed area indicates a lower exposure of the same gel to reduce the band intensities in those lanes. DUP-51 pre-mRNA and GAPDH mRNA were detected by RT-PCR (bottom). Amounts of coprecipitated RNA normalized to the Rbfox3 protein over three experiments are graphed, with means, SD, and p value as in (B).

See Figures S4 and S6 for additional analyses.

containing four continuous Gs. Within the large set of Rbfox-binding regions, both hnRNP M and hnRNP C motifs were highly enriched, whereas hnRNP H motifs were depleted relative to a control distribution (Figure S5).

GCAUG was the most-enriched pentamer in the high-confidence binding regions, and this motif was often aligned precisely at the crosslink site (Figures 2B and S5). However, as seen previously, many iCLIP clusters did not contain the GCAUG motif (Jangi et al., 2014; Lovci et al., 2013; Weyn-Vanhentenryck et al., 2014; Yeo et al., 2009). We subdivided the intronic binding regions into two sets: (A) those containing a GCAUG or a UGCAU within ±40 nucleotides of the crosslink site and (B) those without such a proximal Rbfox-binding motif. These two sets of sequences were analyzed for pentamer frequencies (Figures 7A and 7B; with all motif scores in Table S3). Binding regions in set A showed the expected enrichment of its component

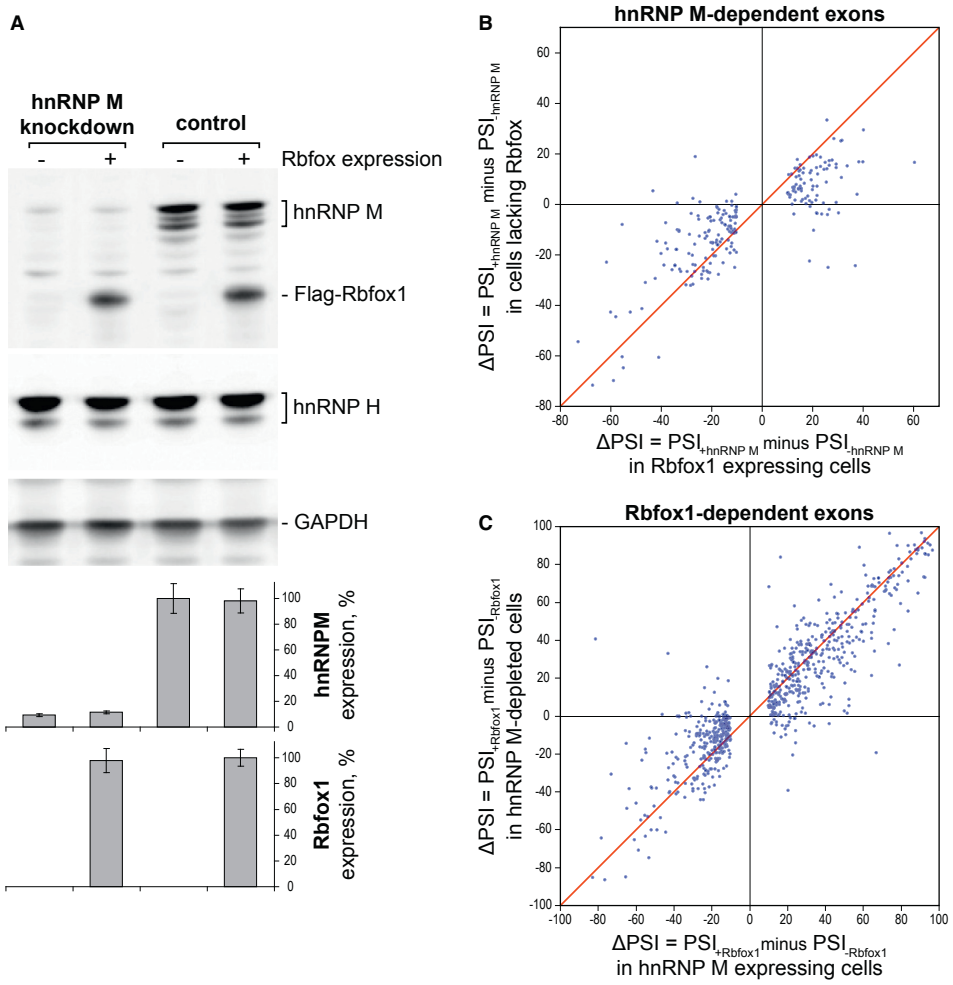


Figure 6. Rbfox1 Stimulates hnRNP M Splicing Activity across Many Exons

(A) Immunoblot of hnRNP M and Rbfox1 in Rbfox2-null HEK293 cells. Rbfox2-knockout cells and derivative cells with Flag-Rbfox1 at the Flp-in locus were grown in doxycycline, transfected with control or hnRNP M-targeted shRNA plasmids and harvested 84 hr post-transfection. Relative protein expression over three experiments is graphed below with SD.

(B) Comparison of hnRNP M splicing activity in cells expressing Flag-Rbfox1 to that in cells not expressing Rbfox proteins. hnRNP M-regulated exons on this chart were defined in Rbfox1-expressing cells as showing a $|\Delta PSI| > 10$ (PSI in hnRNP M-expressing cells minus PSI in hnRNP M-knockdown cells) and FDR < 0.5. x axis, ΔPSI values for these exons in Rbfox1 expressing cells. y axis, corresponding ΔPSI values in Rbfox1-lacking cells.

(C) Rbfox1 splicing activity in hnRNP M-expressing cells and hnRNP M-depleted cells is compared as in (B). See Table S4 for the rMATS analysis.

GCAUG and UGCAU pentamers (Figure S5). In the set B binding regions, U pentamers were the most-enriched motifs and were more common than in set A (compare Figures 7A and 7B). Similar to the entire high-confidence set of iCLIP clusters (Figure S5), GU pentamers were highly enriched in both subsets of binding regions. These data indicate that binding sites for other LASR components are commonly found adjacent to sites of Rbfox binding.

To examine the positions of particular motifs relative to the crosslink sites, we further refined two smaller groups of intronic Rbfox iCLIP clusters. Group 1 contained a GCAUG within five nucleotides upstream or downstream of the crosslinking site defined as position 0 (orange lines in Figure 7C). As seen previously (Jangi et al., 2014; Weyn-Vanhentenryck et al., 2014), the GCAUG pentamer most frequently began at either position −4 or 0, consistent with crosslinking to the second or the first guanine of a GCAUG motif (Figure 7C, top). The second group of clusters (red lines in Figure 7C) contained crosslink sites more than 100 nt away from the nearest GCAUG. In group 2 clusters,

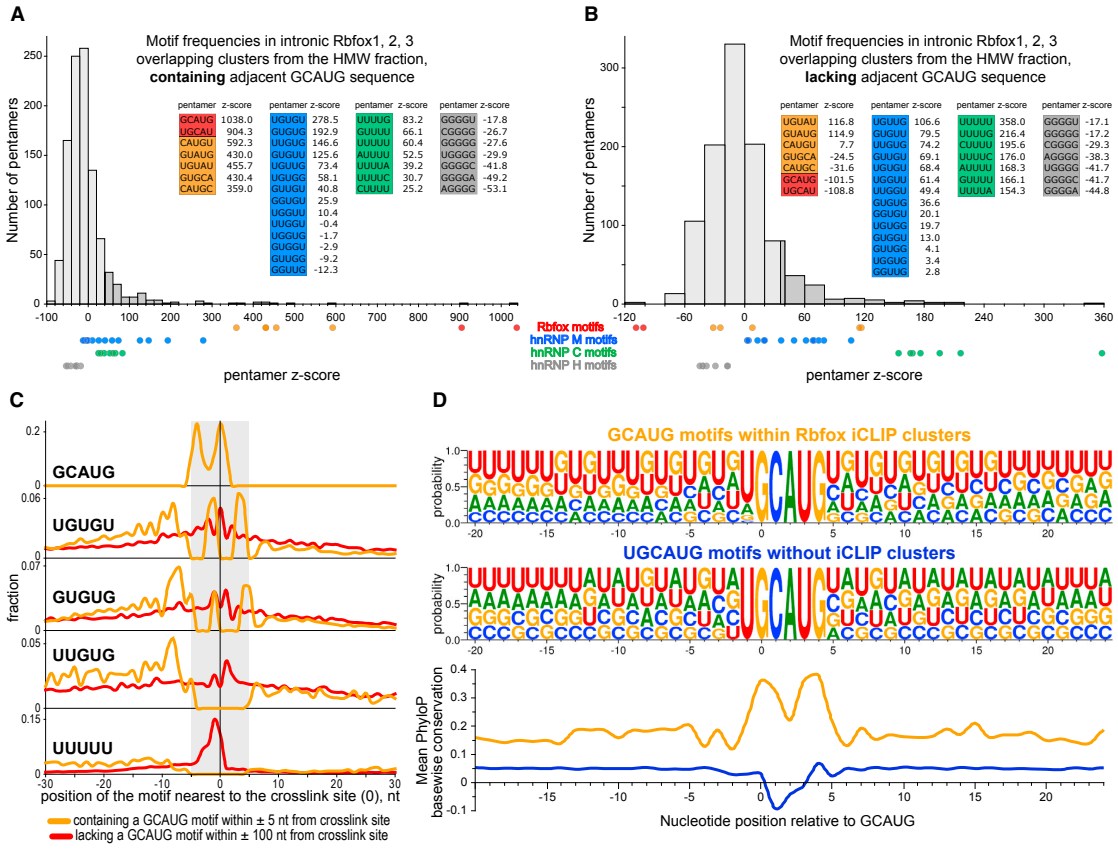


Figure 7. Enrichment of Sequence Motifs near Sites of Rbfox Binding

Histogram of pentamer enrichment *Z* scores within 40 nucleotides of the crosslink sites. iCLIP clusters showing overlap for all three Rbfox paralogs in forebrain or hindbrain HMW fractions were analyzed.

(A and B) Motif enrichments were calculated for crosslink sites less than 40 nucleotides from the nearest GCAUG motif (A) or for sites more than 40 nt from this motif (B). The top 10% of *Z* scores is shaded darker gray. The Rbfox-binding GCAUG, UGCAU (red), and similar pentamers (orange) are indicated as dots below and sorted by *Z* score above. Motifs recognized by hnRNP M (blue), hnRNP C (green), and hnRNP H (gray) are similarly shown.

(C) Motif distribution near the crosslink sites. The fraction of sequences with an individual motif aligning at each nucleotide relative to the crosslink site is plotted. Smaller groups of the cluster subsets A and B were analyzed: (1) those containing a GCAUG sequence within five nucleotides of the crosslink site (orange lines) and (2) those with a crosslink >100 nt away from the nearest GCAUG (red lines).

(D) WebLOGO plots of the sequence adjacent to (U)GCAUG motifs (Crooks et al., 2004). (Top) Intronic sequences containing a GCAUG motif within 5 nt of an Rbfox crosslink site are shown. (Middle) Sequences containing UGCAUG from introns with Rbfox iCLIP clusters but >100 nt away from the nearest crosslink site are shown. (Bottom) Mean PhyloP placental conservation scores of the WebLOGO sequences are shown (blue line, no Rbfox crosslinking; orange line, crosslinked to Rbfox).

See Figures S5 and S6 for additional information.

Rbfox crosslinking is presumably determined by other interactions than GCAUG recognition, although it is possible that secondary structure brings a distal GCAUG motif close to the crosslinked region. The U_5 motif was strongly enriched precisely at the crosslink sites of the group 2 clusters (red line, Figure 7C, bottom), with the most-frequent crosslink at U_2. This motif showed no enrichment at particular positions adjacent to the GCAUG crosslink sites (orange line).

Individual GU pentamers were enriched at particular positions in both groups of crosslinking sites (Figure 7C shows the GU pentamers with the top *Z* scores). In group 1 clusters, these GU pentamers were enriched at upstream and downstream positions directly abutting the crosslinked GCAUG, as well as further upstream of the GCAUG motif. Because all the Rbfox is assembled with LASR, the common position of these pentamers presumably reflects the binding of another LASR protein—most likely hnRNP

M. The group 2 clusters showed different patterns of GU pentamer placement (Figure 7C, red lines), with peaks very close to the crosslink site, and dispersed enrichment in the adjacent sequence.

The most-enriched GU-rich pentamer in these group 2 clusters was UGUUG. To examine the activity of this element in splicing, we made a new splicing reporter with the duplicated element UGUUGUGUUG in the exon (DUP-50M1; diagramed in Figure S6A). This fused element also contains a UGUGU pentamer that is enriched in group 2 clusters. Purified hnRNP M, but not Rbfox2, bound this exon in vitro, and the M binding was eliminated by mutation of the element (Figures S4A–S4C). This element, but not its mutant, rendered the exon repressible by Rbfox3 and stimulated crosslinking of Rbfox3 to the transcript in vivo (Figures S6B and S6C), similar to the hnRNP M element defined by CLIP (Figures 5B and 5C).

For every GCAUG sequence within an iCLIP cluster, there can be dozens of exact UGCAUG sequences within the same transcript that do not exhibit crosslinking and presumably do not bind Rbfox. To examine the contribution of adjacent nucleotides to Rbfox binding, we aligned the central GCAUG motifs from the group 1 clusters and plotted the nucleotide frequency at adjacent positions. The highest probability sequence flanking the GCAUG from position −15 to position +15 consisted of alternating G and U nucleotides (Figure 7D, top). This pattern was compared to UGCAUG hexamers from introns containing iCLIP clusters but which did not generate a cluster themselves. Nucleotides adjacent to these non-Rbfox-binding hexamers exhibited a different pattern of A and U enrichment (Figure 7D, middle). GCAUG motifs bound by Rbfox were also more conserved than the unbound motifs, with PhyloP conservation scores peaking within the GCAUG and extending into adjacent nucleotides (Figure 7D, bottom). These data indicate that the surrounding nucleotide context of a (U)GCAUG element contributes to its recruitment of an Rbfox protein.

DISCUSSION

A Large Protein Assembly for Rbfox Proteins

Much of splicing occurs in conjunction with transcription, and splicing factors are concentrated adjacent to active loci in nuclear speckles thought to consist of dense networks of intermolecular interactions. After nuclear lysis, unspliced RNA remains associated with chromatin and other HMW assemblies. However, biochemical analyses of pre-mRNA splicing usually employ proteins and RNPs eluted from intact nuclei at moderate salt that may not engage in all interactions defining their function. To assess the contacts of splicing regulators in compartments more immobile than the soluble nucleoplasm, we lysed nuclei under mild conditions and examined how proteins partition between the pellet and the supernatant upon centrifugation. The Rbfox proteins were largely found in the HMW pellet, containing chromatin, nuclear speckle components, and unspliced RNA. Splicing regulators were associated with RNA in this fraction, and extraction of the pellet with Benzonase released them in soluble form. In this preparation, the Rbfox proteins were associated with a multimeric complex, the LASR. It will be interesting to examine other regulators in these fractions and perhaps find other new interactions.

The Rbfox protein in the HMW extract was entirely associated with LASR. Isolation of LASR via a tagged Rbfox protein or tagged hnRNP M or H yielded approximately equal quantities of hnRNP M, hnRNP H, and tagged Rbfox and only slightly lower amounts of Matrin3, hnRNP U-like-2, hnRNP C, NF110, NF45, and DDX5. The near equimolar stoichiometry of the components when isolated with different tagged subunits indicated that LASR assembled via specific protein-protein interactions and was not a random aggregation of proteins released by the nuclease. Similarly, the absence of other RNA-binding proteins in the complex, and its resistance to extensive nuclease treatment, indicated that the LASR subunits were not held together by binding to a common RNA. This nuclease resistance distinguishes LASR from previously characterized RNP assemblies such as the 40S hnRNP particle, the DBIRD complex, and the higher-order exon junction complex (Close et al., 2012; Singh et al., 2012; Walker et al., 1980).

LASR subunits were present as free proteins, small complexes that contain components other than Rbfox, and large 55S assemblies that contain Rbfox. The 55S complexes were heterogeneous in size yet still yielded equivalent stoichiometries of all the components, indicating that they may be higher-order assemblies of a unit complex. The unit LASR complex is present in cells without Rbfox (Figures 4B and S3C). This may multimerize to create the 55S forms or interact with a much larger but substoichiometric structure to yield its high S value. Given that all of the Rbfox is in this larger form, it is possible that Rbfox itself mediates the multimerization. In the brain, nuclear Rbfox1 and Rbfox3 were almost entirely associated with the 55S complex. Rbfox2 bound this complex but was also in light fractions, indicating a possible functional difference of Rbfox2. It will be interesting to identify interactions that hold the LASR subunits together, that allow recruitment of Rbfox, and that mediate its higher-order assembly.

The components of LASR engage in a variety of other interactions. LASR shares several components with one of two described microprocessor complexes, including NF45, DDX5, DDX17, hnRNP M, and hnRNP H (Gregory et al., 2004). Instead of Rbfox, the microprocessor contains Drosha and DGCR8 that carry out miRNA processing. Other described Rbfox interactions include with U1C, hnRNP K, Sam68, RALY, PSF, TFG, and Ataxin2, as well as the aforementioned hnRNP H contact (Huang et al., 2012; Kim et al., 2011; Mauger et al., 2008; Shibata et al., 2000; Sun et al., 2012). Recombinant Rbfox added to an in vitro splicing extract inhibited assembly of a pre-spliceosomal E-complex on an Rbfox-repressed exon (Fukumura et al., 2007; Zhou and Lou, 2008). This activity will be interesting to assess for involvement of the LASR complex.

Rbfox Regulation of Splicing in the Context of LASR

We find that the expression of Rbfox altered the activity of hnRNP M in controlling splicing and that M binding sites increased the crosslinking of Rbfox to a reporter RNA in vivo. Thus, the hnRNP M component of LASR can apparently allow Rbfox to alter splicing through an indirect interaction with the RNA. In examining what constitutes a functional binding site, we found that sequences adjacent to Rbfox-crosslinked GCAUG pentamers were enriched for motifs that potentially bind hnRNP M. We further found that atypical non-(U)GCAUG sites of Rbfox

crosslinking were enriched for M and C motifs, possibly indicative of LASR-mediated recruitment of Rbfox. Although additional work is needed to confirm this mechanism, such an hnRNP M/Rbfox interaction can explain how the activity, but not the expression of hnRNP M, increases during the EMT (Venables et al., 2013; Xu et al., 2014).

The multiple RNA-binding domains within LASR raise questions regarding the optimal arrangement of regulatory motifs and how the architecture of the complex might enforce co-recognition of certain motifs. The enrichment of GU-rich motifs adjacent to GCAUG-binding elements may derive from the hnRNP M in the LASR contacting the RNA simultaneously with Rbfox. However, other GU-binding proteins cannot be ruled out. In *C. elegans*, an Rbfox family member can cooperatively assemble with the Sup12 protein binding to an adjacent GU-rich motif (Kuwasako et al., 2014). It is not clear whether mammalian Sup12-like proteins also cooperate with Rbfox or bind LASR. Substantial efforts are being directed at understanding a "splicing code" that would predict the splicing pattern of a pre-mRNA from its sequence (Barash et al., 2010; Zhang et al., 2010). A group of splicing regulators acting within a common complex provides a mechanistic explanation for the co-occurrence of certain binding motifs.

The non-RRM domain proteins of LASR also have interesting features. The double-stranded RNA-binding protein NF110/NFAR2 and the helicase DDX5 were shown to interact and to affect transcriptional regulation (Fuller-Pace, 2013; Ogilvie et al., 2003; Reichman and Mathews, 2003; Saunders et al., 2001). DEAD-box proteins like DDX5 often have lower RNA helicase activity than the DEAH-box family and may instead act as switches to control assembly steps, with ATP hydrolysis toggling the protein between conformational or binding states (Singh et al., 2015). Studies demonstrate DDX5 involvement in splicing of particular exons, where it may alter assembly of pre-mRNP complexes (Guil et al., 2003; Kar et al., 2011; Liu, 2002).

Most attention has focused on the RNA recognition properties of splicing regulators, but these proteins also engage in complex protein/protein interactions. An Rbfox splice variant lacking most of the RNA-binding domain, but retaining the N- and C-terminal domains can block splicing activation, but not splicing repression, by full-length Rbfox (Damianov and Black, 2010). The C-terminal domain is required for both splicing repression and activation by an MS2-tethered Rbfox protein (Sun et al., 2012). The significance of the Rbfox1 and Rbfox3 C-terminal domains is underscored by their mutation in familial epileptic syndromes (Lal et al., 2013a, 2013b). Elucidating the consequences of these mutations, as well as understanding Rbfox1 roles in autism spectrum disorders and spinal cerebellar ataxia, will require a clearer description of Rbfox interactions in nuclear and cytoplasmic mRNA metabolism.

EXPERIMENTAL PROCEDURES

Cell Culture and Tissue Isolation

Stable HEK293 lines expressing HA-Flag-tagged Rbfox1, 2, or 3 proteins were prepared using the Flp-In T-REx System (Life Technologies). An Rbfox2-deficient clone derived from this cell line was obtained by CRISPR/Cas9-guided deletions in the first constitutive Rbfox2 exon. Brain tissue was obtained from 6-week-old C57BL/6J male mice (Charles River Laboratories). Transfection of HEK293 cells was as described (Damianov and Black, 2010). For transient expression of recombinant proteins, cells were harvested 48 hr post-transfection or post-induction with 0.5 μg/ml doxycycline. For RNAi, hnRNP M was targeted with short hairpin RNAs (shRNAs) as described in the Supplemental Experimental Procedures. All experiments were approved by the UCLA Institutional Animal Care and Use Committee (ARC# 1998-155-53).

RT-PCR

Total RNA was extracted with TRIzol (Life Technologies) from cells or tissues. DNA was removed with RQ1 DNase. Reverse transcription was carried out with SuperScript III (Life Technologies) and gene-specific reverse primers. Minigene and GAPDH products were amplified for 15–18 PCR cycles and detected as described (Damianov and Black, 2010). Primer and minigene sequences are listed in Supplemental Experimental Procedures.

RNA-Protein Crosslinking In Vivo

Monolayer HEK293T cultures were irradiated with UV (254 nm) at 75 mJ/cm^2 on ice in a Stratalinker 1800 (Stratagene). Mouse brain samples were triturated in ice-cold HBSS solution and UV irradiated at 600 mJ/cm^2.

Preparation of Whole-Cell Lysates for Reporter Experiments

UV-irradiated HEK293T cells were lysed 5 min on ice with ten packed cell volumes of buffer (20 mM HEPES-KOH [pH 7.5], 150 mM NaCl, 0.5 mM DTT, 1 mM EDTA, 0.6% Triton X-100, 0.1% SDS, and 50 μg/μl yeast tRNA) and centrifuged at 20,000 × *g* for 5 min at 4°C. The supernatants were 5× diluted with buffer (20 mM HEPES-KOH [pH 7.5], 150 mM NaCl, 0.5 mM DTT, 1 mM EDTA, 1.25× cOmplete protease inhibitors [Roche], and 50 μg/μl yeast tRNA). Lysates were spun for 10 min at 20,000 × *g*, 4°C prior to IP.

Subcellular Fractionation

Nuclei from cell cultures or tissues were purified as described (Grabowski, 2005), resuspended in ten volumes of buffer (10 mM HEPES-KOH [pH 7.6], 15 mM KCl, 1 mM EDTA, 0.15 mM spermine, 0.5 mM spermidine), and pelleted at 1,000 × *g* for 5 min at 4°C. Nuclei were lysed for 5 min in ten volumes of ice-cold lysis buffer (20 mM HEPES-KOH [pH 7.5], 150 mM NaCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 1.25× protease inhibitors, and 0.6% Triton X-100). Soluble and HMW fractions were separated by centrifugation at 20,000 × *g* for 5 min at 4°C. Samples for iCLIP are described in the Supplemental Information. To extract nuclease-resistant protein complexes, the soluble fraction was removed and an equal volume of lysis buffer added to the HMW pellet. Soluble and HMW fractions were incubated at 25°C on a rotator with 5 U/μl of Benzonase (Sigma) until the HMW pellet was resuspended and then cleared by centrifugation for 10 min at 20,000 × *g*, 4°C.

Immunoprecipitation

Nuclear fractions or whole-cell lysates were incubated overnight at 4°C with 5- to 7.5-μl-packed M2 FLAG agarose beads (Sigma). For nuclear fractions, beads were washed four times with wash buffer (20 mM HEPES-KOH [pH 7.5], 150 mM NaCl, and 0.05% Triton X-100). For whole-cell lysates, beads were washed five times with wash buffer containing 1M NaCl and twice with standard wash buffer. Flag-tagged proteins were eluted from beads over 2 hr at 4°C in 50–100 μl of elution buffer (20 mM HEPES-KOH [pH 7.5], 100 mM NaCl, and 150 ng/μl of 3×FLAG peptide [Sigma]). RNA-protein crosslinks were eluted with elution buffer plus 5 μg of yeast tRNA.

Protein Analysis

Immunoprecipitated proteins were subjected to MuDPIT as described (Sharma et al., 2014) and analyzed by SDS-PAGE with immunoblotting or protein staining as described (Damianov and Black, 2010). Protein complexes were resolved on 10%–50% glycerol gradients (20 mM HEPES-KOH [pH 7.5], 150 mM NaCl, 0.5 mM DTT, and 1× protease inhibitor) in 14 × 89 mm tubes (Beckman Coulter Genomics). Extracts (250 μl) were loaded and spun in a SW41Ti rotor (Beckman Coulter) at 32,000× RPM for 12 hr at 4°C. Gradients were fractionated top to bottom into 24 × 500 μl fractions. Fractions were analyzed by immunoblot or resolved on 3%–12% NativePAGE Novex Bis-Tris Gels (Life Technologies). Primary antibodies are listed in Supplemental Experimental Procedures.

iCLIP

iCLIP libraries were prepared following König et al. (2010), with changes to allow for the differing RNA content of the cellular fractions. iCLIP libraries were sequenced on a HiSeq2000 (Illumina). Data analyses were performed as in König et al. (2010), with few modifications. In brief, PCR duplicate iCLIP reads were removed using random barcodes. Unique reads were mapped to mouse genome mm9/NCBI37 using Bowtie, allowing two mismatches (Langmead et al., 2009). Mapped reads were assigned to the longest transcripts in the Known Gene table (Hsu et al., 2006) and divided into 5′ UTR, CDS, intron, and 3′ UTR regions. Motif enrichment and the modified iCLIP protocol are described in the Supplemental Experimental Procedures.

Genome-wide Splicing Analysis

Total TRIzol-extracted RNA was treated with TURBO DNase (Ambion) and polyA plus RNA isolated on Oligo-dT. cDNA libraries were prepared using TruSeq Kits (Illumina). Read properties are shown in Table S1B. Alternative splicing was analyzed by rMATS (Shen et al., 2014) and expressed as changes in percent-spliced-in values (ΔPSI). Exons showing splicing change (|ΔPSI| > 10 with false discovery rate [FDR] less than 0.5) between control and hnRNP M-depleted samples from Flag-Rbfox1-expressing cells were considered hnRNP M regulatory targets. Similarly, Rbfox1-dependent exons were defined in cells expressing hnRNP M.

ACCESSION NUMBERS

The accession number for the iCLIP and RNA-seq data reported in this paper is GEO: GSE71468.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, four tables, and one data file and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.03.040.

AUTHOR CONTRIBUTIONS

Conceptualization, A.D. and D.L.B.; Methodology, A.D., Y.Y., and D.L.B.; Investigation, A.D., Y.Y., J.-A.L., D.T., and A.A.V.; Software, E.B.-S. and Y.X.; Formal Analysis, C.-H.L. and E.B.-S.; Data Curation, C.-H.L.; Writing – Original Draft, A.D. and D.L.B.; Writing – Review and Editing, A.D., Y.Y., J.-A.L., Y.X., K.C.M., and D.L.B.; Supervision, A.D., Y.X., J.A.W., K.C.M., and D.L.B.; Project Administration, K.C.M. and D.L.B.; Funding Acquisition, Y.X., K.C.M., J.A.W., and D.L.B.

ACKNOWLEDGMENTS

We thank Julian König and Jernej Ule for communicating the iCLIP protocol ahead of publication and Celine Vuong and Anthony Linares for critical reading of the manuscript. This work was supported by the Howard Hughes Medical Institute (to D.L.B.); NIH grants R21 MH101684 (to K.C.M.), R01 GM088432 (to Y.X.), R01 GM105431 (to Y.X.), and R01 GM089778 (to J.A.W.); a NARSAD Young Investigator Award (to J.-A.L.); the China Scholarship Council (to Y.Y.); an Alfred P. Sloan Fellowship (to Y.X.); and the Broad Center for Stem Cell Research at UCLA.

Received: July 1, 2015
Revised: December 18, 2015
Accepted: March 22, 2016
Published: April 21, 2016

REFERENCES

Amir-Ahmady, B., Boutz, P.L., Markovtsov, V., Phillips, M.L., and Black, D.L. (2005). Exon repression by polypyrimidine tract binding protein. RNA *11*, 699–716.

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

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

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

Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279–290.

Bill, B.R., Lowe, J.K., Dybuncio, C.T., and Fogel, B.L. (2013). Orchestration of neurodevelopmental programs by RBFOX1: implications for autism spectrum disorder. Int. Rev. Neurobiol. *113*, 251–267.

Braeutigam, C., Rago, L., Rolke, A., Waldmeier, L., Christofori, G., and Winter, J. (2014). The RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion. Oncogene *33*, 1082–1092.

Caputi, M., and Zahler, A.M. (2001). Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H′/F/2H9 family. J. Biol. Chem. *276*, 43850–43859.

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

Close, P., East, P., Dirac-Svejstrup, A.B., Hartmann, H., Heron, M., Maslen, S., Chariot, A., Söding, J., Skehel, M., and Svejstrup, J.Q. (2012). DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. Nature *484*, 386–389.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. *14*, 1188–1190.

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

Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat. Rev. Genet. *15*, 689–701.

Fukumura, K., Kato, A., Jin, Y., Ideue, T., Hirose, T., Kataoka, N., Fujiwara, T., Sakamoto, H., and Inoue, K. (2007). Tissue-specific splicing regulator Fox-1 induces exon skipping by interfering E complex formation on the downstream intron of human F1gamma gene. Nucleic Acids Res. *35*, 5303–5311.

Fuller-Pace, F.V. (2013). The DEAD box proteins DDX5 (p68) and DDX17 (p72): multi-tasking transcriptional regulators. Biochim. Biophys. Acta *1829*, 756–763.

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

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

Görlach, M., Burd, C.G., and Dreyfuss, G. (1994). The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. J. Biol. Chem. *269*, 23074–23078.

Grabowski, P.J. (2005). Splicing-active nuclear extracts from rat brain. Methods *37*, 323–330.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235–240.

Guil, S., Gattoni, R., Carrascal, M., Abián, J., Stévenin, J., and Bach-Elias, M. (2003). Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. Mol. Cell. Biol. *23*, 2927–2941.

Hovhannisyan, R.H., and Carstens, R.P. (2007). Heterogeneous ribonucleoprotein m is a splicing regulatory protein that can enhance or silence splicing of alternatively spliced exons. J. Biol. Chem. *282*, 36265–36274.

Hsu, F., Kent, W.J., Clawson, H., Kuhn, R.M., Diekhans, M., and Haussler, D. (2006). The UCSC Known Genes. Bioinformatics *22*, 1036–1046.

Huang, S.C., Ou, A.C., Park, J., Yu, F., Yu, B., Lee, A., Yang, G., Zhou, A., and Benz, E.J., Jr. (2012). RBFOX2 promotes protein 4.1R exon 16 selection via U1 snRNP recruitment. Mol. Cell. Biol. *32*, 513–526.

Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., et al. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep. *1*, 167–178.

Jangi, M., Boutz, P.L., Paul, P., and Sharp, P.A. (2014). Rbfox2 controls autoregulation in RNA-binding protein networks. Genes Dev. *28*, 637–651.

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

Johansson, J.U., Ericsson, J., Janson, J., Beraki, S., Stanić, D., Mandic, S.A., Wikström, M.A., Hökfelt, T., Ogren, S.O., Rozell, B., et al. (2008). An ancient duplication of exon 5 in the Snap25 gene is required for complex neuronal development/function. PLoS Genet. *4*, e1000278.

Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S., and Wu, J.Y. (2011). RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5′ splice site. Mol. Cell. Biol. *31*, 1812–1821.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H., Marr, M.T., 2nd, and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. *25*, 2502–2512.

Kim, K.K., Kim, Y.C., Adelstein, R.S., and Kawamoto, S. (2011). Fox-3 and PSF interact to activate neural cell-specific alternative splicing. Nucleic Acids Res. *39*, 3064–3078.

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

Kuroyanagi, H. (2009). Fox-1 family of RNA-binding proteins. Cell. Mol. Life Sci. *66*, 3895–3907.

Kuwasako, K., Takahashi, M., Unzai, S., Tsuda, K., Yoshikawa, S., He, F., Kobayashi, N., Güntert, P., Shirouzu, M., Ito, T., et al. (2014). RBFOX and SUP-12 sandwich a G base to cooperatively regulate tissue-specific splicing. Nat. Struct. Mol. Biol. *21*, 778–786.

Lal, D., Reinthaler, E.M., Altmüller, J., Toliat, M.R., Thiele, H., Nürnberg, P., Lerche, H., Hahn, A., Møller, R.S., Muhle, H., et al. (2013a). RBFOX1 and RBFOX3 mutations in rolandic epilepsy. PLoS ONE *8*, e73323.

Lal, D., Trucks, H., Møller, R.S., Hjalgrim, H., Koeleman, B.P., de Kovel, C.G., Visscher, F., Weber, Y.G., Lerche, H., Becker, F., et al.; EMINet Consortium; EPICURE Consortium (2013b). Rare exonic deletions of the RBFOX1 gene increase risk of idiopathic generalized epilepsy. Epilepsia *54*, 265–271.

Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P.A., and Burge, C.B. (2014). RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. Mol. Cell *54*, 887–900.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lee, Y., and Rio, D.C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. Annu. Rev. Biochem. *84*, 291–323.

Lee, J.A., Tang, Z.Z., and Black, D.L. (2009). An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. Genes Dev. *23*, 2284–2293.

Lee, J.A., Damianov, A., Lin, C.H., Fontes, M., Parikshak, N.N., Anderson, E.S., Geschwind, D.H., Black, D.L., and Martin, K.C. (2016). Cytoplasmic Rbfox1 regulates the expression of synaptic and autism-related genes. Neuron *89*, 113–128.

Liu, Z.R. (2002). p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5′ splice site duplex. Mol. Cell. Biol. *22*, 5443–5450.

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

Mauger, D.M., Lin, C., and Garcia-Blanco, M.A. (2008). hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. Mol. Cell. Biol. *28*, 5403–5419.

Misteli, T., Cáceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F., and Spector, D.L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. J. Cell Biol. *143*, 297–307.

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

Ogilvie, V.C., Wilson, B.J., Nicol, S.M., Morrice, N.A., Saunders, L.R., Barber, G.N., and Fuller-Pace, F.V. (2003). The highly related DEAD box RNA helicases p68 and p72 exist as heterodimers in cells. Nucleic Acids Res. *31*, 1470–1480.

Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. RNA *15*, 1896–1908.

Reichman, T.W., and Mathews, M.B. (2003). RNA binding and intramolecular interactions modulate the regulation of gene expression by nuclear factor 110. RNA *9*, 543–554.

Saunders, L.R., Perkins, D.J., Balachandran, S., Michaels, R., Ford, R., Mayeda, A., and Barber, G.N. (2001). Characterization of two evolutionarily conserved, alternatively spliced nuclear phosphoproteins, NFAR-1 and -2, that function in mRNA processing and interact with the double-stranded RNA-dependent protein kinase, PKR. J. Biol. Chem. *276*, 32300–32312.

Sharma, S., Wongpalee, S.P., Vashisht, A., Wohlschlegel, J.A., and Black, D.L. (2014). Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. Genes Dev. *28*, 2518–2531.

Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-seq data. Proc. Natl. Acad. Sci. USA *111*, E5593–E5601.

Shibata, H., Huynh, D.P., and Pulst, S.M. (2000). A novel protein with RNA-binding motifs interacts with ataxin-2. Hum. Mol. Genet. *9*, 1303–1313.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell *151*, 750–764.

Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The clothes make the mRNA: past and present trends in mRNP fashion. Annu. Rev. Biochem. *84*, 325–354.

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

Swanson, M.S., and Dreyfuss, G. (1988). Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. Mol. Cell. Biol. *8*, 2237–2241.

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

Tripathi, V., Song, D.Y., Zong, X., Shevtsov, S.P., Hearn, S., Fu, X.D., Dundr, M., and Prasanth, K.V. (2012). SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. Mol. Biol. Cell *23*, 3694–3706.

Venables, J.P., Brosseau, J.P., Gadea, G., Klinck, R., Prinos, P., Beaulieu, J.F., Lapointe, E., Durand, M., Thibault, P., Tremblay, K., et al. (2013). RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. Mol. Cell. Biol. *33*, 396–405.

Walker, B.W., Lothstein, L., Baker, C.L., and LeStourgeon, W.M. (1980). The release of 40S hnRNP particles by brief digestion of HeLa nuclei with micrococcal nuclease. Nucleic Acids Res. *8*, 3639–3657.

Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q., et al. (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell Rep. *6*, 1139–1152.

Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Mol. Cell. Biol. *14*, 7219–7225.

Xu, Y., Gao, X.D., Lee, J.H., Huang, H., Tan, H., Ahn, J., Reinke, L.M., Peter, M.E., Feng, Y., Gius, D., et al. (2014). Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. Genes Dev. *28*, 1191–1203.

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

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

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

Zhou, H.L., and Lou, H. (2008). Repression of prespliceosome complex forma tion at two distinct steps by Fox-1/Fox-2 proteins. Mol. Cell. Biol. *28*, 5507 5516.

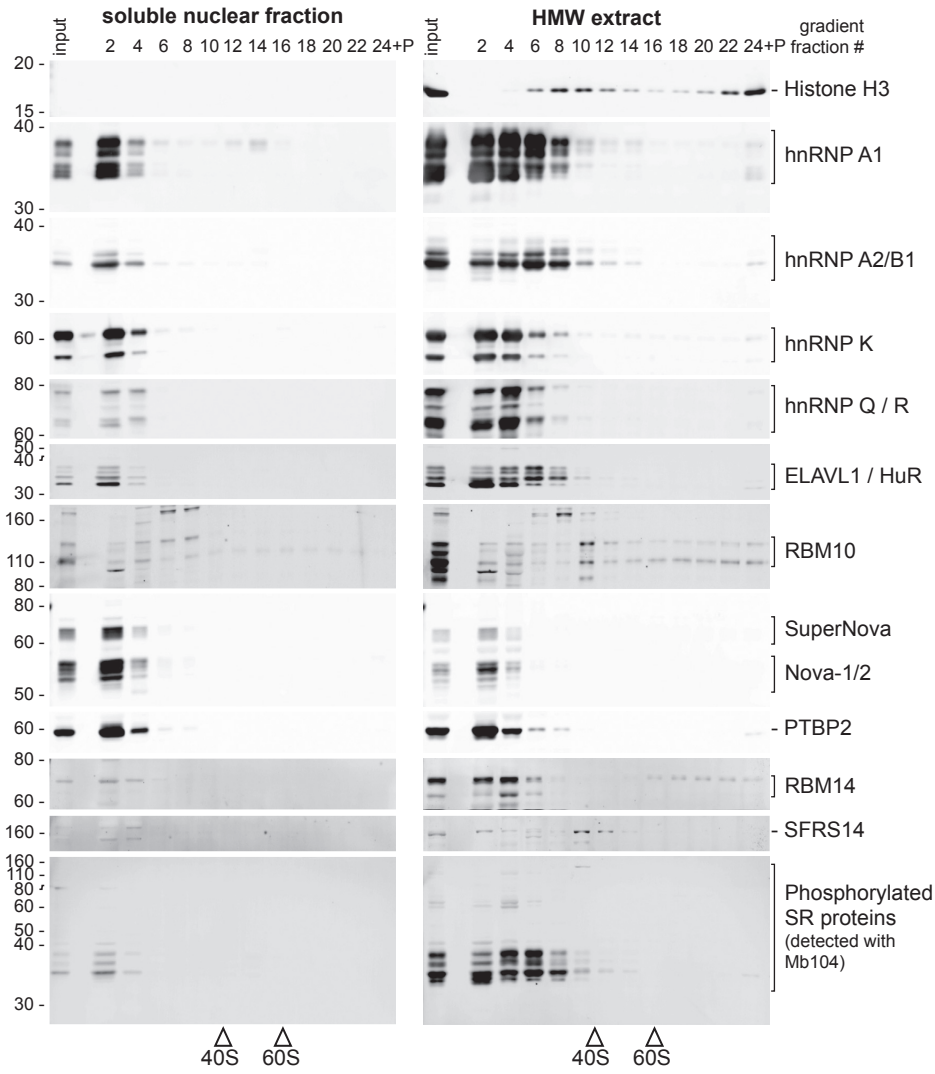


Figure S1. Splicing Factors Not Coprecipitating with Rbfox Proteins Do Not Cosediment in the 55S Glycerol Gradient Fraction, Related to Figures 1 and 3

The soluble nuclear fraction and the HMW extract from mouse brain were subjected to sedimentation through 10%–50% glycerol gradients as in Figure 3B. The proteins detected by immunoblotting are indicated on the right. 40S and 60S markers (arrows) were obtained by sedimenting a HeLa S100 extract in parallel.

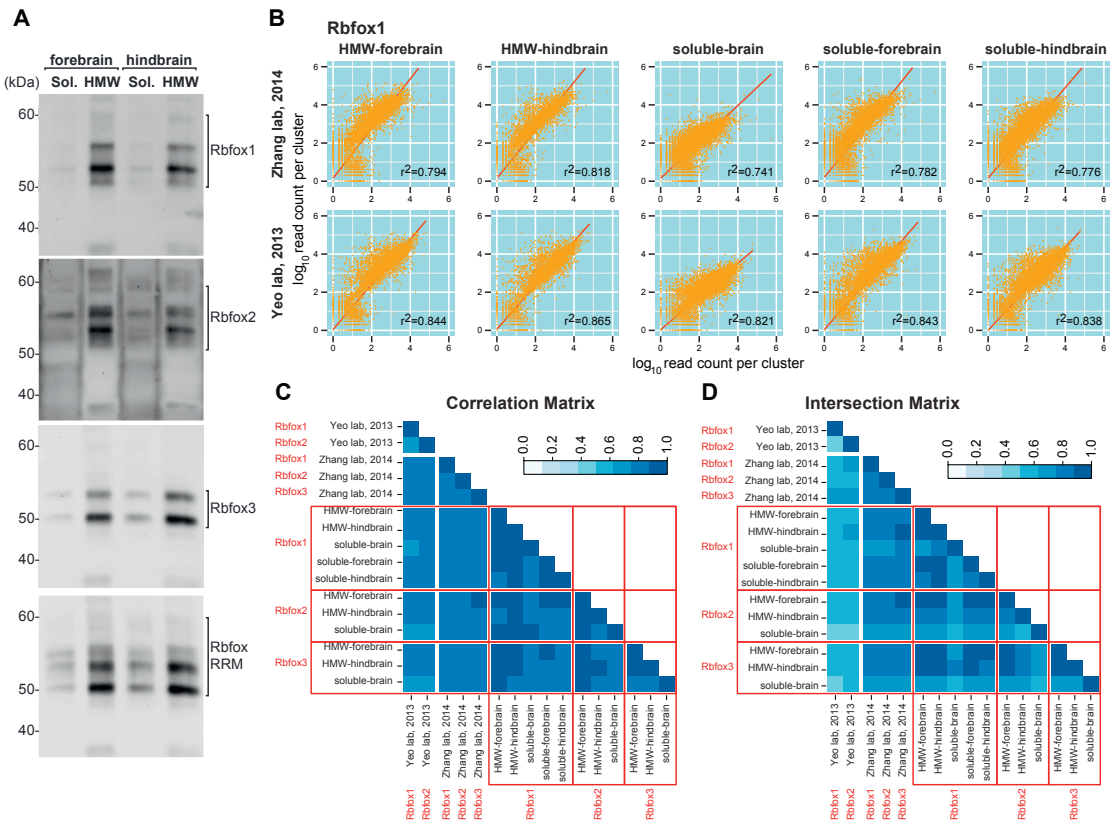


Figure S2. iCLIP Analyses, Related to Figure 2

(A) The Rbfox proteins from mouse brain remain in the nuclear HMW fraction after trituration and UV-irradiation. Immunoblot analysis of the soluble nuclear fraction and the HMW extract from mouse brain, after trituration and irradiation at 600 mJ per cm^2 from a 254 nm ultraviolet light source. Rbfox1, Rbfox2, and Rbfox3 are detected with specific antibodies or simultaneously with an antibody recognizing their common RRM.

(B) Comparison of Rbfox iCLIP clusters with previous CLIP studies (Lovci et al., 2013; Weyn-Vanhentenryck et al., 2014). The iCLIP datasets show strong correlation with the previously published data. The number of reads per cluster in our data were plotted against the same values in two other datasets. Correlation plots are shown with R^2 value for Rbfox1.

(C) Correlation matrix for comparing all the Rbfox pairs of datasets.

(D) Intersection matrix for comparing all the Rbfox pairs of datasets. The number of common clusters divided by the union in each pair of datasets.

See Figure 2.

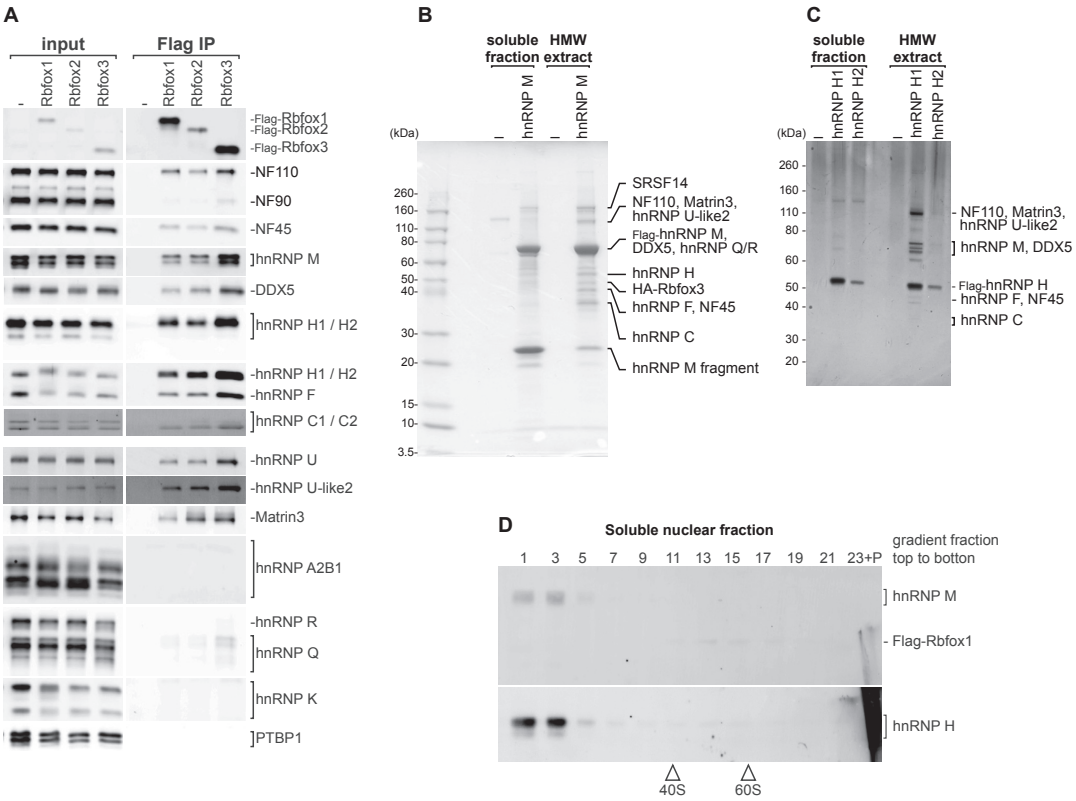


Figure S3. Analyses of Proteins Copurifying with LASR, Related to Figures 3 and 4

(A) Immunoblot validation of Rbfox interacting proteins. Nuclear HMW extracts were prepared from HEK293T cells transfected with control vector (-), or transiently expressing Flag-tagged Rbfox1, Rbfox2, or Rbfox3 proteins. Proteins copurifying with Rbfox proteins were subjected to immunoblot analysis. Aliquots of HMW extracts (input) and anti-Flag precipitates (Flag IP) were analyzed in parallel as indicated on the top. The proteins were probed with specific antibodies as shown on the right.

(B and C) Proteins coprecipitating with hnRNP M and hnRNP H from HEK293T HMW extracts. Anti-Flag immunopurification from nuclear soluble and HMW fractions prepared from human HEK293T cells expressing Flag-tagged hnRNP M, hnRNP H1, or hnRNP H2 as indicated, or transfected with control vector (lanes "-"). The major interacting proteins are indicated on the right. (B) Immunopurification of hnRNP M and interacting proteins. The HEK293T cells in this experiment expressed HA-tagged Rbfox3 in addition to Flag-hnRNP M. The proteins were resolved by SDS-PAGE and stained with SimplyBlue safestain. (C) Immunopurification of hnRNP H and interacting proteins. The HEK293T cells in this experiment did not express ectopic Rbfox proteins. The immunoprecipitated proteins were resolved by SDS-PAGE and silver-stained. Note that the IP of hnRNP H2 was inefficient due to the low expression of this protein, but the same pattern of bands was visible in this lane of the gel.

(D) hnRNP M and hnRNP H present in the soluble nuclear fraction do not participate in large protein complexes. Soluble nuclear fractions were prepared from HEK293T cells transiently expressing Flag-tagged Rbfox1 and were sedimented through 10%–50% glycerol gradients. Sedimentation of HMW extract, prepared in parallel from these cells is shown in Figure 4A. The proteins were detected by immunoblot, indicated on the right as in Figure 4A.

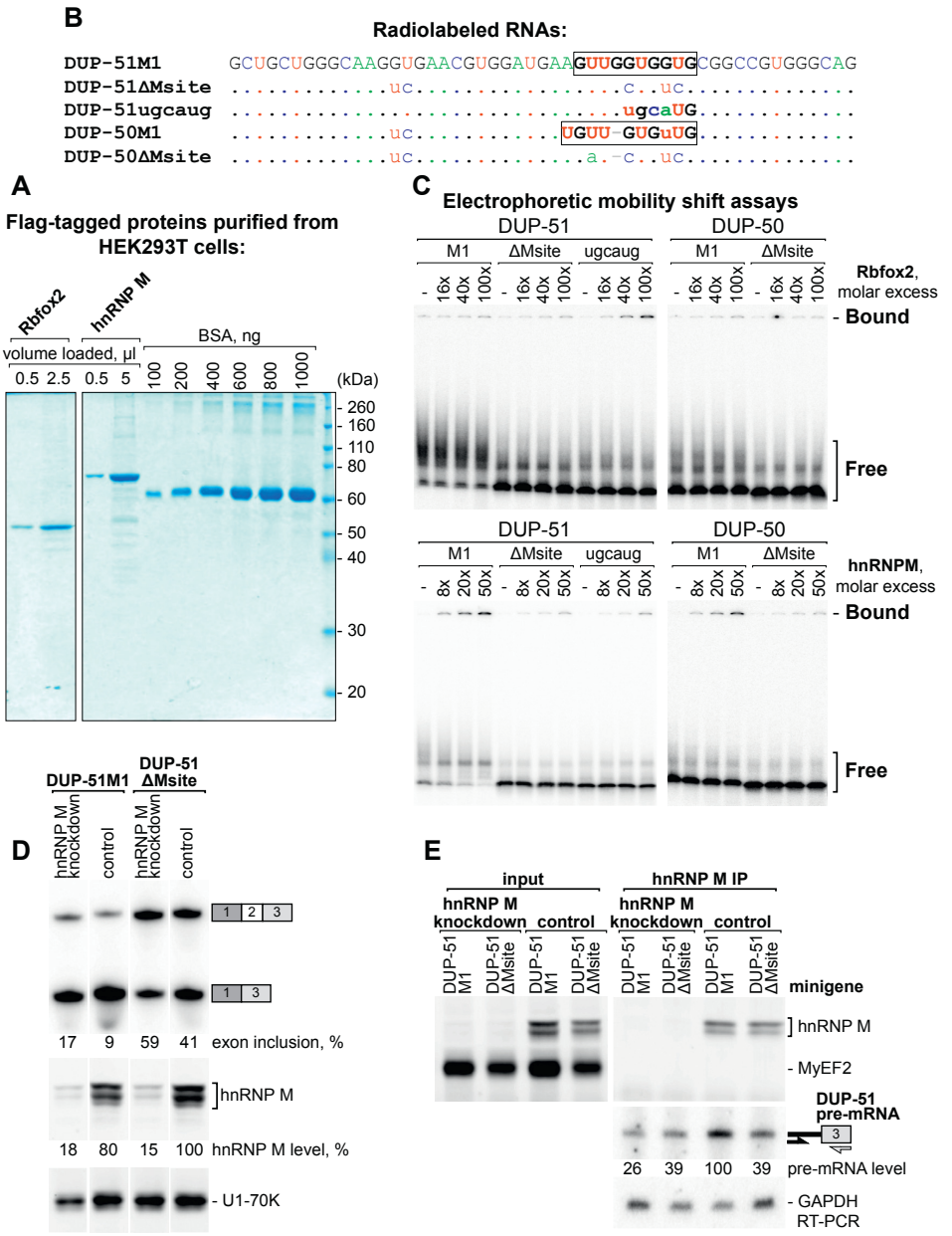


Figure S4. hnRNP M Is Recruited to the DUP-51M1 Pre-mRNA to Regulate Splicing, Related to Figure 5

(A–C) hnRNP M, but not Rbfox2 binds in vitro to the GU-rich sequences in exon 2 of DUP-51 and DUP-50. (A) Coomassie staining of immunopurified Flag-tagged Rbfox2 and hnRNP M. These proteins were transiently expressed in HEK293T cells, immunoprecipitated from whole cell lysates, washed with buffer containing 1M NaCl, and eluted with 3xFlag peptide. Rbfox2 was used because Rbfox3 protein purified this way contained traces of other protein components of the LASR complex (data not shown). (B) Exon 2 of DUP-51 and DUP-50 minigenes. These sequences were transcribed in vitro with internal radiolabel. A version of DUP-51 exon 2, in which the GU-rich site was converted to UGCAUG, was also synthesized. The GU-motifs in DUP-51M1 and DUP-50M1 are boxed in black. (C) Electrophoretic mobility shift assays of the radiolabeled RNAs shown in panel B, after incubation with either Rbfox2 (top) or hnRNP M protein (bottom). 4.5 nM

(legend continued on next page)

RNA was incubated with the indicated molar excess of protein, and in the presence of tRNA and spermidine to reduce non-specific binding. The RNA and RNA:protein complexes were resolved on 7% acrylamide gels (29:1 acrylamide:bis acrylamide ratio) containing 25 mM Tris, 112.5 mM Glycine buffer (pH 8.3). (D and E) HEK293 cells were cotransfected with either the DUP-51M1 or DUP-51ΔMsite minigene along with either a hnRNP M-targeted shRNA vector or a control vector. (D) Exon 2 splicing was analyzed by RT-PCR with primers in the flanking exons (top). The percentage of exon 2 inclusion is indicated below each lane. hnRNP M and U1-70K, which served as a loading control, were detected by immunoblot (bottom). The relative level of hnRNP M in each lane, normalized to U1-70K, is indicated. Note that a lower level of basal exon inclusion is observed in this experiment compared to Figure 5. This is due to the longer cell culture time needed to carry out the double transfection of shRNA and reporter plasmids. This results in a greater level of cell confluency at time of assay, which is seen to reduce exon inclusion in the Dup plasmids. (E) HEK293T cells expressing DUP-51 minigenes and the hnRNP M or control shRNAs, as in D, were UV-irradiated in vivo, lysed under denaturing conditions, and immunoprecipitated with anti-hnRNP M antibody. The levels of hnRNP M and MyEF2 in the lysates (input) and post-IP (hnRNP M IP) were measured by immunoblot (top). Co-precipitated DUP-51 pre-mRNA was detected by RT-PCR (bottom). See also Figure 5.

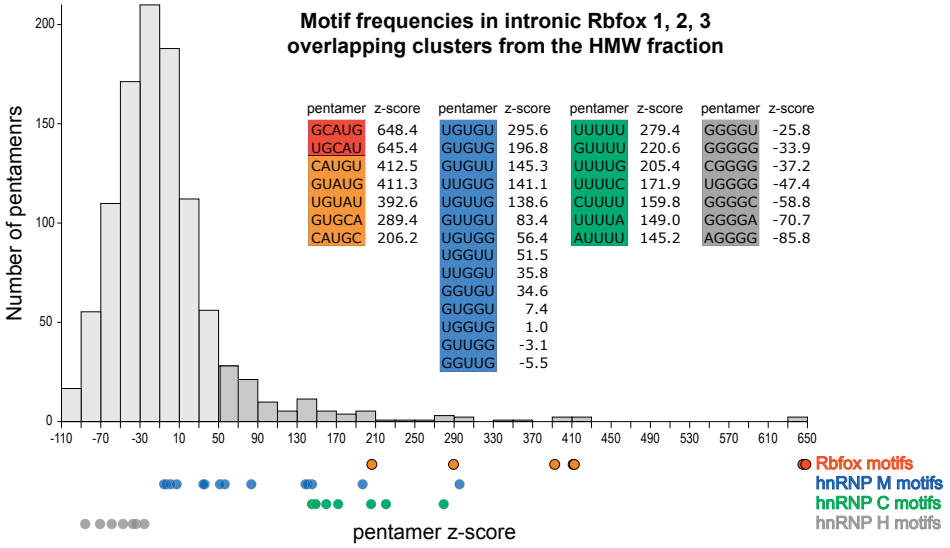


Figure S5. Enrichment of Sequence Motifs near Intronic Crosslink Sites from the Total Set of Overlapping Rbfox1, Rbfox2, and Rbfox3 iCLIP Clusters from Mouse Forebrain or Hindbrain Nuclear HMW Fractions, Related to Figure 7

Histograms of pentamer Z-scores indicating motif enrichment in the Rbfox iCLIP clusters are shown as in Figure 7.

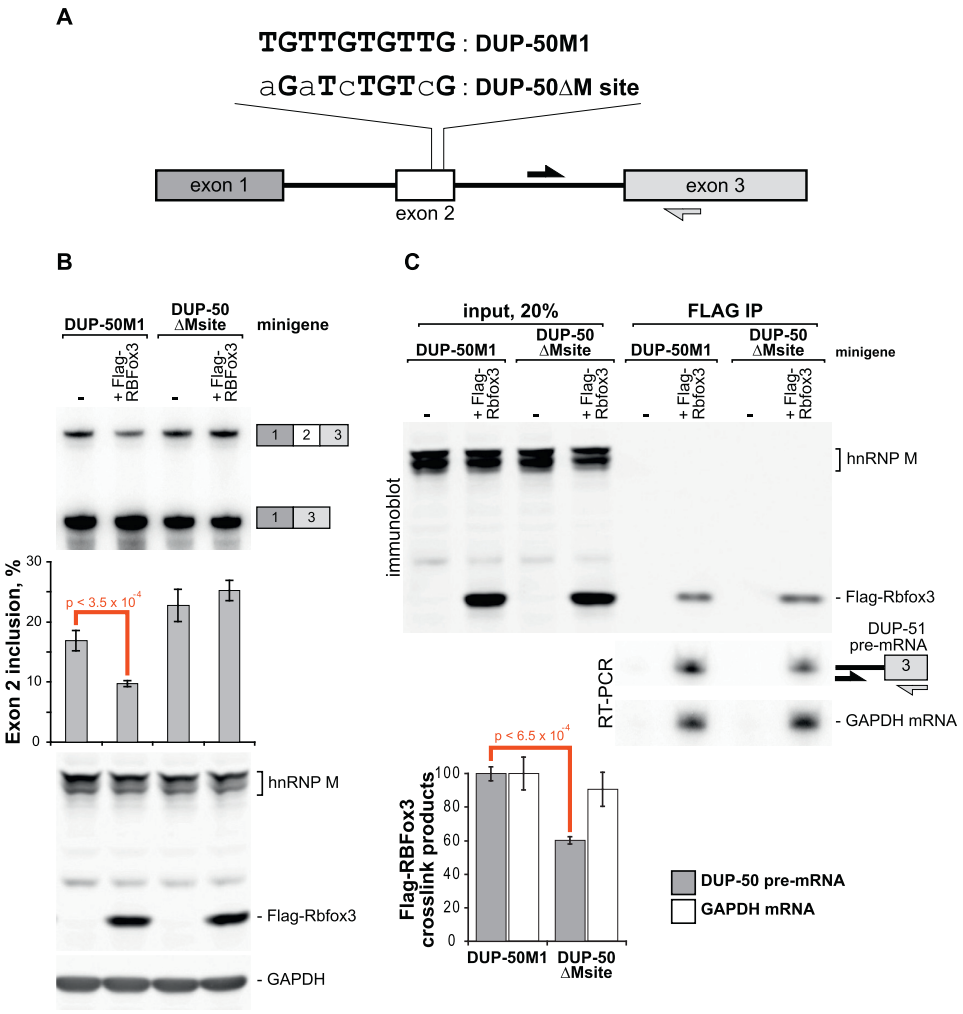


Figure S6. GU-Rich Motifs Enriched in Rbfox Binding Regions Can Confer Rbfox-Dependent Regulation on an Exon, Related to Figure 7

(A) Diagram of the minigenes DUP-50M1 DUP-50ΔMsite.

(B and C) The effect of Rbfox3 on exon 2 splicing (B) and crosslinking to DUP-50 pre-mRNA (C) are analyzed and shown as in Figure 5.

Supplemental material

A

Tissue	Fraction	Protein / antibody	Significant reads	iCLIP clusters	Significant reads in gene regions, %			
					5' UTRs	CDS regions	introns	3' UTRs
Forebrain	HMW	Rbfox1	206,419	27,420	0.3	0.7	92.6	6.4
		Rbfox2	183,165	21,522	0.2	0.9	92.9	5.9
		Rbfox3	153,504	20,456	0.3	0.8	91.8	7.1
	Soluble	Rbfox1	527,355	20,875	0.8	7.1	15.4	76.6
Hindbrain	HMW	Rbfox1	335,818	36,746	0.2	0.7	94.3	4.8
		Rbfox2	246,515	27,583	0.1	1.0	92.7	6.1
		Rbfox3	346,986	38,192	0.1	0.7	92.6	6.6
	Soluble	Rbfox1	303,936	17,378	0.4	3.4	28.7	67.5
Brain	Soluble	Rbfox1	136,483	7,080	0.2	3.5	20.3	76.0
		Rbfox2	77,137	3,345	0.2	3.9	20.8	75.0
		Rbfox3	170,240	3,667	0.1	1.1	15.7	83.0

B

Library ID	-Rbfox, -hnRNP M			+Rbfox, -hnRNP M			-Rbfox, +hnRNP M			+Rbfox, +hnRNP M		
	replicate 1	replicate 2	replicate 3	replicate 1	replicate 2	replicate 3	replicate 1	replicate 2	replicate 3	replicate 1	replicate 2	replicate 3
Number of input reads	29,505,514	27,163,511	23,977,861	20,449,200	20,407,845	21,719,496	31,069,868	32,560,985	30,223,076	39,214,977	23,586,357	32,109,927
Uniquely mapped reads number	26,672,600	24,598,904	16,444,741	16,884,619	18,028,381	19,175,947	27,578,312	29,321,413	26,426,555	34,763,688	15,451,350	29,130,368
Uniquely mapped reads, %	90.40%	90.56%	68.58%	82.57%	88.34%	88.29%	88.76%	90.05%	87.44%	88.65%	65.51%	90.72%
Number of splices, total	7,711,379	7,148,062	4,967,171	5,119,525	5,311,401	5,565,996	8,164,354	8,765,408	7,976,808	9,970,133	4,699,521	9,027,587
Number of splices, sjdb annotated	7,611,822	7,057,173	4,907,062	5,047,202	5,235,023	5,485,171	8,078,355	8,673,291	7,893,654	9,846,061	4,644,933	8,920,595

Table S1: Sequencing and mapping data for iCLIP and TruSeq libraries. Related to Fig. 2 and Fig. 6.

(**A**) Rbfox iCLIP experiments from mouse brain nuclear fractions.

Numbers of iCLIP clusters with width of at least two nucleotides are shown for each experiment. The numbers of significant reads (with FDR less than 0.01) within these clusters are also indicated. The percentages of these reads mapping to 5' untranslated regions (5' UTRs), coding DNA sequence (CDS), introns, and 3' untranslated regions (3' UTRs) of the longest transcripts from known genes are shown for each experiment. See Fig. 2.

(**B**) TruSeq libraries from 293 Flp-In cells. Read numbers and mapping data from the STAR algorithm are indicated for each library. See Fig. 6

HMW extract	intronic Rbfox 1,2,3 iCLIP overlapping clusters	crosslink sites from overlapping clusters			
		all	containing GCAUG or UGCAU ±40nt from crosslink site	containing GCAUG ±5nt from crosslink site	lacking GCAUG and UGCAU ±40nt from crosslink site
Forebrain	3,600	84,861	25,397	3,815	59,464
Hindbrain	4,410	163,702	87,855	15,766	75,847

Table S2: Overlapping Rbfox 1, Rbfox2, and Rbfox3 iCLIP clusters on introns in HMW nuclear fractions. Related to Fig. 2 and Fig. 7.

The numbers of iCLIP clusters from HMW nuclear fraction from forebrain and hindbrain, overlapping by at least one nucleotide, and the crosslinking sites within these clusters are indicated. These crosslink sites are further split into groups based on their proximity to the nearest GCAUG or UGCAU pentamers as described. See Fig. 2 and 7.

Extended Experimental Procedures

Cell culture and tissue isolation

HEK293T and Flp-In™ T-REx™ 293 cells were grown according to ATCC protocols. For transient transfection, cells were grown in 6-well plates and transfected as described (Damianov and Black, 2010). For transient expression of recombinant proteins cells were harvested 48 hours post-transfection or post-induction with 0.5 µg /ml doxycycline. For RNAi, hnRNP M was targeted at the 3' UTR sequences GCATCTTGTTGACATCGAAT and AGATTGCAATGTGCGCAATT with shRNAs expressed from the plasmid pBlsH1 (Gencheva et al., 2010). ShRNA plasmids were first transfected in suspension, followed by transfection in monolayer 36 hours later, and harvested after an additional 48 hours. Stable HEK293 lines expressing HA-Flag-tagged Rbfox1, 2, or 3 proteins were prepared using the Flp-In™ T-REx™ System (Life Technologies). An Rbfox2 deficient clone derived from this cell line was obtained by CRISPR/Cas9-guided deletions in the first constitutive Rbfox2 exon. Forebrain and hindbrain tissues were obtained from 6 week old C57BL/6J male mice (Charles River). All experiments were approved by the UCLA Institutional Animal Care and Use Committee.

Antibodies

Primary antibodies used for immunoblot assays: FLAG (F3165-1MG, Sigma), Rbfox1 (1D10, (Gehman et al., 2011)), Rbfox2 (A300-864A, Bethyl), Rbfox3 (MAB377, Millipore), RbfoxRRM (Damianov and Black, 2010), hnRNP M (NB200-314, Novus Biologicals), MyEF2 (HPA004883-100UL, Sigma), hnRNP H (Chou et al., 1999), hnRNP F (Min et al., 1995), hnRNP C (ab75822, Abcam), Matrin3 (A300-591A, Bethyl), hnRNP U-like2 (ab104042, Abcam), hnRNP U (A300-689A, Bethyl), ILF3 (NF110 and NF90 proteins, NBP1-40682, Novus Biologicals), NF45 (ab28772, Abcam), DDX5 (ab10261-100, Abcam), DDX17 (ab24601-100, Abcam), ELAVL1/HuR (ab14371, Abcam), Nova (Buckanovich and Darnell, 1997), hnRNP A2/B1 (ab6102, Abcam), hnRNP K (ab52600,

Abcam), hnRNP Q/R (R5653-200UL, Sigma), hyperphosphorylated SR proteins (mAb104, (Roth et al., 1990)), PTBP1 (Markovtsov et al., 2000), PTBP2 and U1-70K (Sharma et al., 2005), GAPDH (AM4300, Ambion), Histone H3 (ab1791-100, Abcam).

Minigene Sequences

Exon 1:

CTTACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATCCAA
GGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGTCGAACG
TGGATGAAGTTGCTGTCGAGGCCCTGGGCAG

Intron 1:

gttggtatcaaggttacaagacaggtttaaggagaccaatagaaactggccaagtggag
acagagaagactcttgggtttctgatagggcccactgactctctctgcctattggtcta
ttttcccacccttag

Exon 2 of DUP51-M1:

GCTGCTGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGCGGCCGTGGGCAG

Exon 2 of DUP-51ΔMsite:

GCTGCTGGGCAAGtcGAACGTGGATGAAGTTGcTGtcGCGGCCGTGGGCAG

Exon 2 of DUP50-M1:

GCTGCTGGGCAAGTCGAACGTGGATGATGTTGTGTTGCGGCCGTGGGCAG

Exon 2 of DUP-50ΔMsite:

GCTGCTGGGCAAGTCGAACGTGGATGAaGaTcTGTcGCGGCCGTGGGCAG

Intron 2:

gttggtatcaaggtaccaagacaggtttaaggagaccaatagatctggccaagtggaga
cagagaagactcttgggtttctgataggcactgactctctctgcctattggtctatttt
cccacccttag

Exon 3:

GCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCA
CTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGT
GCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAG
TGAGCTGCACTGTGACAAGCTGCACGTGGATCCTTGAGCATCTGGATTCTGCCT

Primers

Gene	RT-PCR product	Forward primer	Reverse primer
DUP-51 minigenes	spliced RNA	GACACCATCCAAGGTGCAC	CTCAAAGAACCTCTGGGTCCAAG
	unspliced RNA	TTGGGTTTCTGATAGGCACTG	CTCAAAGAACCTCTGGGTCCAAG
human GAPDH	mRNA	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG

Motif analyses and comparisons with published Rbfox CLIP studies

Crosslink sites, defined as one nucleotide upstream of each iCLIP read were evaluated for significance using the False Discovery Rate (FDR) method as described (König et al., 2010). Crosslink sites with FDR ≤ 0.01 were used for clustering and downstream analyses. Sites with genomic locations within 20 nucleotides were clustered.

The binding regions, defined as including 40 nucleotides upstream and 40 nucleotides downstream of each crosslinking site, were analyzed for enrichment of nucleotide pentamers. To control for nucleotide frequency biases, pentamer frequencies in the binding regions were compared to the distribution of pentamer frequencies in a large set of 81-nucleotide genomic intervals randomly chosen from the introns containing iCLIP clusters. GCAUG and UGCAU, the two pentamer derivatives of the UGCAUG motif, were the most enriched motifs in the binding regions (red dots, z-score > 645, Fig. S5). Several other near matches to the UGCAUG motif including CAUGU, GUAUG, UGUAU, GUGCA, CAUGC (orange dots, Fig. S5) were also highly enriched. Some of these partial matches were previously shown to bind Rbfox2 (Lambert et al., 2014).

Within the distribution of pentamer z-scores we examined binding motifs for hnRNP's M, C, and H. Based on the previous CLIP analyses (Huelga et al., 2012), we defined possible hnRNP M binding motifs as all pentamers containing 3 G's and 2 U's or 3 U's and 2 G's but without more than 2 Gs or Us in a row. HnRNP C pentamers included U_5 and all pentamers containing 4 continuous U nucleotides. Similarly, hnRNP H pentamers included G_5 and all pentamers containing 4 continuous G's. Most of the potential hnRNP M binding pentamers were enriched in the Rbfox binding regions, with 8 pentamers having Z-scores above 50 (Fig. S5, blue dots), placing them among the 10% most enriched motifs. Since UV-induced RNA-protein crosslinking is biased towards uridines (Sugimoto et al., 2012), motifs with high U content might be overrepresented at iCLIP crosslink sites. To rule out the possibility that the hnRNP M pentamers were enriched simply due to their uridine content, we compared them to equivalent motifs with C or A substituted for G. All of these CU-rich and AU-rich

motifs had lower Z-scores than the corresponding GU-rich motif, with most having negative values (data not shown). Thus, the enrichment of these GU motifs was not simply due to their uridine content.

We also found that the hnRNP C binding pentamers were enriched in the Rbfox binding regions (Fig. S5, green dots). Since hnRNP C elements are highly uridine-rich, we could not distinguish whether their enrichment resulted from direct Rbfox recognition, recruitment of Rbfox by hnRNP C binding, or simply enhanced crosslinking to those elements. The G-rich pentamers that potentially bind hnRNP H were depleted relative to the average pentamer in the intronic Rbfox binding regions (Fig. S5, gray dots). It is possible that the exclusion of hnRNP H binding sites from the sequence immediately adjacent to Rbfox binding is a consequence of the structural relationship of the two proteins within LASR.

To compare the iCLIP data with previous CLIP studies, replicates were combined to create one data set per study. The reads were collapsed into clusters and the read number per cluster was determined. For clusters intersecting between two data sets the reads per cluster were plotted. The correlation matrix shows the r^2 measure of correlation for all pairs of data sets. For the intersection matrix, the number of clusters overlapping between each pair of CLIP data sets was determined and divided by the number of clusters in the union set for that pair.

Modified iCLIP protocol

This follows the original iCLIP protocol developed by the Ule lab (König et al., 2010) with the following differences:

- Sample preparation is modified for compatibility with nuclear fractions.
- Partial digestion of RNA is carried out post immunoprecipitation with Micrococcal nuclease (MNase), similar to the Yeo lab CLIP protocol (Yeo et al., 2009).
- The immunoprecipitated RNA is fragmented in presence of yeast tRNA in predetermined ratio to the MNase, which allows one to skip the titration step.
- The 3' hydroxyl end of the synthetic RNA linker is blocked by a biotin group, and the protein:crosslinked RNA:linker products are then purified on Monomeric Avidin Agarose prior to separation by electrophoresis.

Materials:

Tubes and tips:

Use original Eppendorf 1.5 ml tubes. Tubes from other manufacturers may have lower quality lids. Barrier pipette tips are highly recommended. All plasticware should be RNase-free.

Buffers:

WB_{150} : 20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.1% Triton-X100

WB_{750} : 20 mM HEPES-KOH pH 7.5, 750 mM NaCl, 0.1% Triton-X100

PNK buffer: 20 mM HEPES-KOH pH 7.5, 10 mM $MgCl_2$, 0.2% Tween-20

EGTA buffer : 20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 20 mM EGTA, 0.1% Triton X-100

TE buffer: 10 mM Tris-HCl pH 7.5, 1 mM EDTA.

Prepare these buffers with DEPC-treated MilliQ-H_2O. Keep all of them except the TE buffer on ice.

Oligonucleotides:

L3 linker (RNA): `/Phos/-UGAGAUCGGAAGAGCGGUUCAG-`**`Biotin`**

Primers for reverse transcription:

RTclipGTT: /Phos/-nnnnAACnnnnAGATCGGAAGAGCGTCGTggatccTGAACCGC

RTclipGCC: /Phos/-nnnnGGCnnnn...

RTclipATC: /Phos/-nnnnGATnnnn...

The barcodes are underlined. Order RT primers with different barcodes in such a way that no single nucleotide substitution should convert one barcode to another.

cut_oligo: GTTCAggatccACGACGCTCTTCaaaa

P5solexa: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCGATCT

P3solexa: AAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTCCGATCT

Magnetic beads: Dynabeads Protein G (Life Technologies, Cat# 10004D).

These beads are very tiny and work efficiently with diluted, large volume samples.

Because of their small size, not all beads will migrate to the sidewall of the tube next to the magnet when placed on a magnetic stand - some will remain on the lid of the tube or will be trapped in the foam layer on the top. To minimize the loss of beads with each wash step, gently invert the magnetic rack with the tubes to allow the trapped beads to find their way to the magnet. Let the samples rest for several seconds, then repeat until the foam layer becomes completely clear.

These beads also require rather vigorous shaking to break up the clumps after removal from the magnetic rack. Make sure to completely resuspend the beads at the beginning of each wash step. After resuspending the beads in wash buffer, we typically invert the tubes about 50-60 times.

Do not centrifuge these beads.

Procedure:

1. Preparation of samples:

Sample types:

- **HMW nuclear pellet**

- **Soluble nuclear fraction**, in buffer containing 20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 1x Complete protease inhibitors, 0.6% Triton X-100

- Whole cells

Triturate brains from 6-week old C57BL/6J male mice in ice-cold HBSS solution as described (Ule et al., 2005) and UV-irradiate at 600 mJ/cm^2 (four pulses of 150 mJ/cm^2 each, gently mixing the sample between irradiations), at 4 ˚C in a Stratalinker 1800.

For cells in monolayer culture, irradiate at 450 mJ/cm^2, at 4 ˚C, and harvest with 1x PBS or HBSS, and pellet at 700x g for 2 min at 4 ˚C.

These UV doses were determined experimentally as described (Darnell, 2012).

Prepare HMW nuclear pellet and soluble nuclear fraction as described (see the fractionation protocol), but do not treat samples with Benzonase nuclease or RNases.

Add 10-15 volumes of buffer containing 20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.6% Triton X-100, 0.1% SDS, 1 mM EDTA, and 0.5 mM DTT to the **HMW pellet or the cell pellet**. Mix immediately by pipetting if this is a cell pellet.

Add 20% SDS and 0.5 M EDTA to **soluble nuclear fraction or cytosol**, to final concentrations of 0.1% and 2 mM, respectively.

Sonicate all samples in a Bioruptor (Diagenode) at 4 ˚C with 30 pulses of 30 sec each. Make sure the chromatin is completely resuspended.

Centrifuge at 20,000 x g, 4 ˚C for 5 min. Transfer the supernatants to fresh tubes. Dilute with five volumes of buffer containing 20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.5 mM DTT, 1.25x Complete protease inhibitors EDTA-free (Roche), and 50 μg/ml yeast tRNA. Mix briefly and centrifuge again at 20,000 x g, 4 ˚C for 10 min.

2. Immobilization of antibodies:

Transfer an appropriate amount of Dynabeads to a 1.5 ml tube. We use approximately 10 μl of packed beads per sample. Wash the beads three times in 1000 μl of WB_{150}. Add about 5-50 μg of IgG per 10 μl of packed beads, add 700-1000 μl of WB_{150} and rotate overnight at 4 ˚C or for a few hours at room temperature.

Wash the beads three times with WB_{750} and once with WB_{150}. Transfer the beads in fresh tube(s) using WB_{150}.

3. Immunoprecipitation:

Add each supernatant to Dynabeads loaded with IgG and rotate overnight at 4 ˚C. Transfer the beads to fresh tubes with WB_{150}. Wash five times with 1 ml of WB_{750} and two times with PNK buffer. If using more than 10 μl of beads, at this point all samples should be split into aliquots containing 10 μl of packed beads each.

4. RNA fragmentation:

Collect the Dynabeads at the bottom of the tube using a magnet (not to the sidewall on magnetic rack stand), then completely remove the buffer on top of the beads. Remove the buffer the same way prior to all other enzymatic steps of this protocol.

Add 100 μl of 1x MNase buffer (NEB) containing 5.0 μg of yeast tRNA.

Place on Thermomixer (Eppendorf) and determine the minimum shaking speed at which the beads completely resuspend in the buffer. Set the Thermomixer at 37 ˚C, 15 sec shaking/15 sec rest, and equilibrate the samples at this temperature. Add 50 μl of 1x MNase buffer containing 60 gel units/ml (6 Kunz units/ml) of MNase (NEB, cat# M0247S). Incubate for exactly 5 min and stop the reaction with 500 μl of EGTA buffer.

This ratio of RNA to MNase routinely produces suitable partial digestion at the given incubation conditions.

Wash four times with EGTA buffer and two times with PNK buffer.

5. Dephosphorylation of RNA:

Add 100 µl of 1x FastAP buffer (Fermentas) containing 0.15 U/µl of Fast alkaline phosphatase (Thermo Scientific, cat# EF0651) and 0.2 U/µl of RNaseOUT (Life Technologies, cat# 10777-019). Incubate in a Thermomixer for 90 min at 37 ˚C, 15 sec shaking/20 sec rest. This long incubation is required for removal of the 3' phosphate groups from MNase-cleaved RNA. The reaction is less efficient than dephosphorylation of 5' ends.

Wash four times with WB_{750} (the high salt buffer is required to wash away the phosphatase) and two times with PNK buffer.

6. L3 linker ligation:

Add 40 µl of 0.75x T4 RNA ligase1 buffer (NEB) containing 1 mM ATP, 25% PEG4000 (Sigma, cat# 202398), 0.5 U/µl T4 RNA ligase1 (NEB, cat# M0204S), 0.5 U/ul RNaseOUT, and 6.0 µM L3 linker.

Incubate in Thermomixer overnight at 16 ˚C, 15 sec shaking/4 min rest.

Wash four times with WB_{150} and two times with PNK buffer.

7. ^{32}P labeling by phosphorylation of the 5' ends:

Add 16 µl of 1x PNK buffer (NEB) containing 150 µCi of γ[^{32}P] ATP, 10 units of T4 Polynucleotide Kinase (NEB, cat# M0201S), and 1U/µl of RNaseOUT. Add 24 µl of PNK wash buffer.

Incubate in Thermomixer for 20 min at 37 ˚C, 15 sec shaking/20 sec rest. Wash three times with WB_{150}.

8. Elution from Dynabeads:

Add 50 µl of buffer containing 100 mM Tris-HCl pH 7.5, 0.6% SDS, 5 mM EDTA, 50 mM DTT, and 50 ng/ul yeast tRNA. Incubate in Thermomixer for 10 min at 85 ˚C while shaking continuously. Transfer the eluted material in a fresh tube.

Rinse the Dynabeads with 1200 µl of buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1.25x Complete protease inhibitors, 50 ng/ul yeast tRNA, and

0.1% Triton X-100. Add the supernatant to the eluted material from the previous step and mix briefly.

Centrifuge at maximum speed for five minutes at 4 ˚C. Transfer the supernatant to a fresh tube. This centrifugation helps to completely remove the Dynabeads.

9. Binding on Monomeric Avidin beads:

This is an additional purification designed to remove the IgG heavy chains, which comigrate or run slightly above the Rbfox proteins on SDS-PAGE. We find that this helps to get cleaner results with other proteins as well. The L3 linker must have a 3' biotin residue.

Wash Immobilized Monomeric Avidin Agarose (Thermo Scientific, cat# 20228) three times with WB_{150}. Pellet the beads after each washing step at 500-1,000 x g, 4 ˚C for 1 min, using a swinging bucket rotor not a fixed rotor. This helps minimize the loss of beads.

Remove the supernatant and add one packed-bead volume of WB_{150}. Add 15 µl of resuspended beads to each sample. Incubate on a rotator for 3-4 hours at 4 ˚C.

Wash the beads three times with WB_{150} and once with EGTA buffer. After removing the last wash buffer, centrifuge again and carefully remove the remaining 5-20 µl of supernatant using a P10 pipette.

10. Elution from Monomeric Avidin beads:

Add 30 µl of buffer containing 100 mM Tris-HCl pH 7.5, 10% Glycerol, 2.2% SDS, and 5 mM EDTA. Incubate in Thermomixer for 10 min at 85 ˚C while shaking.

Centrifuge in fixed rotor at 10,000 x g for 1 min. Carefully remove 25 µl of supernatant and transfer to a fresh tube. Add 5 µl of 1x LDS sample buffer (Life Technologies, 4x stock cat# NP0007) containing 300 mM DTT.

Store at -20 ˚C or proceed with the next step.

11. NuPAGE electrophoresis:

Incubate the samples at 90 ˚C for 10 min.

Prerun a 10% NuPAGE Bis-Tris Gel (Life Technologies, cat# NP0307) with 1x MOPS SDS running buffer (Life Technologies, 20x stock cat# NP0001) for 10-15 min at 75 V. Remove the foam from the running buffer on top of the gel, and rinse the wells by pipetting. Load 30 μl of sample per preparative lane. Load 7.5 μl of Novex sharp pre-stained protein standards (Life Technologies, cat#57318) in the analytical lane. Run for 10-20 min at 75 V, then increase the voltage to 110-120 V and continue the electrophoresis as necessary.

12. Electrotransfer:

Equilibrate the gel for 5 min in transfer buffer (25 mM Bis-Tris, 25 mM Bicine,1 mM EDTA, pH 7.2, 10% Methanol, 0.02% SDS). Rinse a piece of Protran BA-85 nitrocellulose membrane (0.45 μM pore size, Whatman GmbH cat#10104594) with milliQ-H_2O and then equilibrate in transfer buffer. Equilibrate a sheet of extra-thick blot paper (Bio-Rad, cat# 1703969) in transfer buffer and place it over the cathode of a Semi-Dry Electrophoretic Transfer Cell (Bio-Rad cat# 170-3940). Place the membrane over the paper, then place the gel over the membrane. Avoid trapping air bubbles. Cover the gel with another sheet of extra-thick blot paper soaked in buffer. Transfer for 60-75 min at 400 mA, not exceeding 15 V.

Briefly rinse with MilliQ-H_2O and wrap the membrane in saran wrap.

13. Size selection of RNA crosslinked to the immunoprecipitated protein:

Expose an X-ray film or a Phosphoimager screen to the membrane. Use appropriate guides to align the image with the membrane. Excise regions from the preparative lanes 20-40 kDa above the protein of interest. Transfer each cut piece of membrane to a separate 1.5 ml tube.

14. Elution of RNA by deproteinization:

Add 300 µl of buffer containing 100 mM Tris-HCl ph 7.5, 50 mM NaCl, 10 mM EDTA, and 2 µg/µl proteinase K and completely submerge the membrane. Incubate for 30 min at 55 ˚C while shaking continuously. Preincubate buffer containing 100 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM EDTA, 7 M Urea, and 0.5 µg/µl proteinase K for 2 min at 55 ˚C. Add 300 µl of this buffer to each tube and continue the incubation for another 30 min.

Transfer the solution to a fresh tube, add equal volume (600 µl) of Phenol: Chloroform 5:1, pH 4.5), and vortex for 30 sec. Centrifuge at maximum speed for 5 min and carefully transfer the aqueous phase to a fresh tube. Add 0.5 µl of 15 µg/µl GlycoBlue (Life Technologies, cat# AM9516), and 60 µl of 3M sodium acetate pH 5.4. Mix briefly, add 660 µl of isopropanol, mix thoroughly, and incubate overnight at -20 ˚C.

Centrifuge at 20,000 x g, 4 ˚C for 30 min. Wash the pellet with 1 ml of 75% Ethanol. Completely remove the ethanol, and air dry the pellet for a maximum of 2 min.

Add 5.70 µl DEPC treated MilliQ-H_2O and let the tube sit on ice for 5-10 minutes.

15. Reverse transcription:

Add 0.5 µl of 10 mM dNTP mix and 0.5 µl of 2 µM RT primer to a PCR tube. Add 5.5 µl of RNA sample and mix 5-10 times by pipetting. Incubate in a PCR machine for 5 min at 70 ˚C with lid heating. Take the tube from 70 ˚C and place it directly on ice for at least 1 min. Set the PCR heat block at 25 ˚C. Equilibrate the tube at this temperature for 30-60 sec and add 3.5 µl of a mix containing 2 µl of 5x First strand buffer, 0.5 µl of 100 mM DTT, 0.5 ul of 100 U/µl SuperScript III Reverse Transcriptase (Life Technologies, cat# 18080-044), and 0.5 µl of 40 U/µl RNaseOUT. Mix 10 times by pipetting. Incubate the sample at 25 ˚C for 5 min, then for 20 min at 42 ˚C, and for another 20 min at 48 ˚C.

Transfer the reverse transcription reaction to a 1.5 ml tube, add 100 µl of TE buffer and 320 µl of Ethanol:3M sodium acetate pH 5.4 (25:1), mix and incubate overnight at -20 ˚C. Centrifuge at 20,000 x g, 4 ˚C for 30 min. Wash the pellet

with 1 ml of 75% Ethanol. Remove the ethanol completely and dissolve the pellet in 5.0 μl of MilliQ-H_2O.

Add 7.5 μl of Formamide containing 10 mM EDTA, Bromophenol Blue and Xylene Cyanol tracking dyes.

16. Size selection of cDNA:

Prepare molecular weight marker in the following way: Mix 2.0 μl of GeneScan 500 LIZ marker (Life Technologies, cat# 4322682), 3.0 μl of MilliQ-H_2O, and 15 μl of Formamide containing 10 mM EDTA, but not tracking dyes.

Thoroughly clean electrophoretic glass plates and cast a 5.5% acrylamide:bisacrylamide (19:1) / 1xTBE / 50% Urea gel using 0.4-0.5 mm spacers. Let the gel polymerize for an additional 30-60 min after becoming solid. Attach a metal plate to one of the glass plates for uniform heat dissipation. Pre-run the gel with 1x TBE running buffer for 15 min at 21 W. Denature the samples and the marker for 5 min at 85 ˚C and then place them immediately on ice for at least one minute. Rinse the wells of the gel by pipetting to remove the urea that migrated there during the pre-run. Load 12.5 μl of sample per lane, then load 10 μl of marker on the left hand side of the first sample and the right hand side of the last sample. Run the electrophoresis for 15-20 min at 21 W.

Rinse with Milli-Q water and wipe/dry the outer surface of the glass plates, but do not remove them from the gel. Scan the gel in a Typhoon scanner, Cy5 channel, +3 mm focal plane to detect the bands of the marker. Print this image in actual size and also print a mirror image. Remove one of the glass plates, put the corresponding printout under the remaining glass plate, align it with the gel and use it as a guide to excise a gel slice from each lane, containing cDNA in the range of 70-120 nt (cDNA length = 56 nt + RNA fragment length).

Chop each piece of gel into 15-25 smaller pieces, transfer them to a 1.5 ml tube, and add 700 μl of TE buffer. Keep on a rotator at room temperature for several hours or overnight.

Centrifuge for 1 min at maximum speed, then carefully transfer the TE buffer into a fresh tube. We use 200 μl flat tips to avoid transferring small acrylamide pieces

with the buffer. Add 0.5 µl of 15 µg/µl GlycoBlue, 75 µl of 3M sodium acetate pH 5.4, and mix. Add 750 µl of isopropanol, mix again, and store overnight at -20 ˚C. Centrifuge for 30 min at 20,000 x g, 4 ˚C. Wash the pellets with 1 ml of 75% Ethanol. Remove the ethanol completely, air dry the pellet for a maximum of 2 min. Add 6.70 µl DEPC treated MilliQ-H_2O and let the tube sit on ice for 5-10 minutes.

17. Circularization of the cDNA:

Add 1.5 µl of the following mix in a PCR tube: 0.8 µl of 10x Circligase II buffer, 0.4 µl of 50 mM $MnCl_2$, and 0.3 µl of 100 U/µl CircLigase II ssDNA ligase (Epicentre, cat# CL9021K). Add 6.5 µl of cDNA solution and mix by pipetting. Incubate in a PCR machine at 60 ˚C for 60 min with lid heating.

18. Linearization at the BamHI site:

Add 30 µl of the following mix: 4 µl of 10x FastDigest buffer, 0.9 µl of 10 µM cut_oligo, and 25.1 µl of MilliQ- H_2O. Incubate for 4 min at 95 ˚C, then decrease the temperature by 1 ˚C every 60 sec until reaching 37 ˚C. Add 2 µl of FastDigest BamHI (Thermo Scientific, cat# FD0054), mix and incubate for 30 min at 37 ˚C.
Transfer the sample to 1.5 ml tube, add 200 µl of TE buffer, add 700 µl of Ethanol:3M sodium acetate pH 5.4 (25:1), and mix and incubate overnight at -20 ˚C. Centrifuge at 20,000 x g, 4 ˚C for 30 min. Wash the pellet with 1 ml of 75% Ethanol. Remove the ethanol completely, dry the pellet briefly, and dissolve in 22 µl of MilliQ-H_2O.

19. Analytical PCR to determine the optimum number of cycles for preparative amplification:

Prepare 42 µl of PCR reaction mix containing 2 µl of single stranded DNA template, 1x Pfu buffer, 0.2 mM dNTP mix, 0.2 µM of each P5solexa and P3solexa primers, and 0.5 units of Pfu polymerase. Split this mix into four aliquots of 10 µl each in PCR tubes.

Prepare a negative control the same way, adding water instead of template.

Amplify one aliquot from each PCR mix for 20, 24, 28, and 32 cycles using the following parameters: initial denaturation: 94 ˚C for 3 min, cycle (denaturation 94 ˚C for 30 sec, annealing 63.5 ˚C for 15 sec, extension 72 ˚C for 30 sec), final extension for 7 min at 72 ˚C, cool down to 4 ˚C.

Using a separate set of pipettes and a separate bench if available, run these samples on a 2% agarose gel containing 0.5x TBE and 0.5 μg/ml Ethidium bromide. For each sample, calculate the number of cycles that will produce about 50-200 ng of PCR product from the rest of the template.

20. Preparative PCR:

Prepare 30 μl of PCR reaction mix the same way as above, using 20 μl of single stranded DNA template. Amplify by PCR using the parameters described above and the number of cycles determined at the analytical step.Run the PCR reactions on a 2% agarose gel and excise gel bands containing products ranging from 150 to 210 bp (PCR product length = 132 bp + RNA fragment length).

Extract DNA from the agarose gel slice using a Zymoclean Gel DNA recovery kit (Zymo Research, cat# D4007). Elute DNA from the column with 10 μl of EB buffer (Qiagen, cat# 19086).

Determine the concentration of DNA by Qubit using the dsDNA BR Assay Kit (Life technologies, cat# Q32850). Prepare 5-20 μl of sequencing library, containing 10 nM DNA and 0.1% Tween-20 in buffer EB. Multiple PCR products can be mixed together if they were prepared with RT primers bearing different barcodes.

Sequence in Illumina HiSeq2000 single end 100 nt.

References

Buckanovich, R.J., and Darnell, R.B. (1997). The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. Molecular and cellular biology *17*, 3194-3201.

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

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

Darnell, R. (2012). CLIP (cross-linking and immunoprecipitation) identification of RNAs bound by a specific protein. Cold Spring Harbor protocols *2012*, 1146-1160.

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

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nature structural & molecular biology *17*, 909-915.

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

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

Roth, M.B., Murphy, C., and Gall, J.G. (1990). A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. The Journal of cell biology *111*, 2217-2223.

Sharma, S., Falick, A.M., and Black, D.L. (2005). Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. Molecular cell *19*, 485-496.

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

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

Table S3. Complete Motif Enrichment Scores near Rbfox iCLIP Crosslink Sites, Related to Figure 7.

Enrichment of sequence motifs within the region from -40 to +40 nucleotides relative to intronic crosslink sites of overlapping Rbfox1, Rbfox2, and Rbfox3 iCLIP clusters in the HMW nuclear fraction						
	Rbfox binding motif					
	Similar to Rbfox binding motif					
	hnRNP M binding motif					
	hnRNP C binding motif					
	hnRNP H binding motif					
	all sites		**sites less than 40 nucleotides away** from the nearest GCAUG or UGCAU		**sites more than 40 nucleotides away** from the nearest GCAUG or UGCAU	
rank	motif	Z-score	motif	Z-score	motif	Z-score
1	GCAUG	648.4	GCAUG	1038.0	UUUUU	358.0
2	UGCAU	645.4	UGCAU	904.3	UUUUG	216.4
3	CAUGU	412.5	CAUGU	592.3	UGUUU	208.9
4	GUAUG	411.3	AUGUG	496.7	CUUUU	195.6
5	AUGUG	405.4	UGUAU	455.7	UUUGU	191.6
6	UGUAU	392.6	GUGCA	430.4	UUUUC	176.0
7	UAUGU	361.7	GUAUG	430.0	AUUUU	168.3
8	UGUGC	336.5	UGUGC	414.7	GUUUU	166.1
9	UGUUU	299.4	UAUGU	408.4	UUGUU	162.9
10	UGUGU	295.6	AUGCA	377.7	UUUUA	154.3
11	GUGCA	289.4	CAUGC	359.0	UGCUU	143.4
12	GUGUA	279.5	GUGUA	285.0	UAUUU	135.9
13	UUUUU	279.4	UGUGU	278.5	UCUUU	129.8
14	UUUGU	268.5	AUGUA	269.1	UAUGU	122.1
15	AUGUA	247.4	UUGCA	267.1	UUUAU	121.9
16	GUUUU	220.6	UAUGC	218.8	UUUCU	120.3
17	CAUGC	206.2	AUAUG	209.8	GUUUG	117.2
18	UUUUG	205.4	GUGUG	192.9	UGUAU	116.8
19	GUUUG	200.3	CUGCA	178.2	UUGCU	115.7
20	AUGCA	198.8	UUUGC	174.8	GUAUG	114.9

21	GUGUG	196.8	GUAUA	169.7	CUGUU	110.8
22	UUGUU	188.6	AUGUU	163.1	UGUUG	106.6
23	UUUGC	177.9	CAUGA	157.2	UGUCU	103.7
24	UUGCA	176.6	CAUGG	153.4	UUAUU	99.5
25	UUUUC	171.9	UUGUG	146.6	UUCUU	99.3
26	UCUGU	164.1	AGCAU	140.1	GCUUG	97.5
27	UGUCU	163.5	UGUAC	138.5	UCUGU	92.6
28	AUGUU	161.9	UGUUU	137.7	GCUUU	87.6
29	CUUUU	159.8	ACAUG	136.5	UGAUG	83.5
30	UGCUU	159.4	UGCAC	128.8	GUGUU	79.5
31	UUUUA	149.0	AUGUC	128.0	UGCCU	76.1
32	AUAUG	147.2	GCAUU	126.9	UUGUG	74.2
33	UAUGC	147.1	GCAUA	126.2	CUUUG	72.2
34	GUGUU	145.3	UGUCU	126.1	UGAUU	70.3
35	AUUUU	145.2	GUGUU	125.6	CUUGU	70.2
36	GUAUA	141.2	UUUGU	121.3	GUUGU	69.1
37	UUGUG	141.1	GUUUG	121.2	CUGUG	68.9
38	UGUAC	139.3	CUGUG	119.8	GUGUA	68.6
39	UGUUG	138.6	CGCAU	115.3	UGUGU	68.4
40	AUGUC	138.4	AUGCU	111.2	UUUGA	67.6
41	UAUUU	132.7	UGCUU	102.8	UUGUA	67.3
42	UUUAU	129.1	GGCAU	102.1	UUUAA	67.2
43	CUGUG	128.0	GUCUG	101.9	UGUGC	66.6
44	GUCUG	123.8	CAUAU	101.4	GUUUA	66.5
45	UUGCU	121.3	UGCGU	90.8	GGUUU	66.0
46	GCUUG	119.0	GCUUG	89.8	GCUGU	65.8
47	UUGUA	106.1	GAAUG	88.2	UGCUG	62.9
48	UUUCU	105.4	UCUGU	86.5	UGGUU	61.4
49	UGCAC	101.4	UUUUG	83.2	AUGUG	61.3
50	UGCCU	101.0	UACAU	82.8	UUUCG	61.1
51	UUAUG	94.7	UUAUG	81.6	GAUUU	60.9
52	AUGCU	94.7	CGUGU	79.1	UUCUG	60.7
53	UCUUU	94.2	UGUGA	77.5	GUCUG	60.5
54	GUAUU	94.1	GUGUC	76.8	UGUAC	59.5
55	CAUGA	93.9	AUGCG	76.3	ACUUU	58.4
56	CUUGU	92.6	UAUAU	75.9	UUUGG	58.4
57	UGUGA	89.7	GUACA	74.6	UUGUC	58.4
58	GAAUG	86.4	UGAAU	73.6	UUUGC	58.4
59	UUAUU	86.0	UGUUG	73.4	AUGUU	58.2
60	CUGUU	84.8	AUGCC	73.0	UUAUG	56.7

61	CUGCA	83.9	AUUGC	72.5	GUAUU	56.3
62	CGUGU	83.7	UUGUA	68.2	UUUCC	55.6
63	GUUGU	83.4	CGUGC	66.6	UUGAU	55.6
64	AGCAU	82.4	GCACG	66.2	GUCUU	54.6
65	GUGUC	81.4	GUUUU	66.1	CGUUU	54.1
66	GCCUG	80.9	UUUAU	64.8	UUUAC	54.0
67	UUGUC	80.7	AAUGC	63.5	CUGAU	53.0
68	GCUUU	80.1	UCAUG	62.5	UUUAG	52.0
69	UGAAU	78.9	GCGUG	60.9	CCUUU	51.7
70	UGCUG	78.7	UUUUU	60.4	CUUGC	51.1
71	CGCAU	78.0	UGAGU	60.3	GUGUC	50.9
72	GCUGU	77.2	GCAUC	59.6	AUGAU	49.7
73	CAUGG	76.9	GUAUU	59.4	UUGGU	49.4
74	ACAUG	76.8	UGUGG	58.1	GCCUG	49.2
75	UUCUU	75.5	CUUGC	53.7	AUGCU	47.2
76	UGCGU	73.8	UGCCU	53.3	GUUUC	47.0
77	UAUAU	70.6	CUUGU	52.7	UUCGU	45.0
78	GUGCU	66.7	AUUUU	52.5	CUAUU	44.6
79	GUUGC	64.9	GUGCU	52.4	UCUUG	44.6
80	CUUGC	64.2	GCCUG	52.0	UUCGC	44.4
81	UGAUU	63.9	UUGUU	51.4	AUGUA	44.3
82	UGAUG	63.5	UCUGC	49.5	AAUUU	44.1
83	AUGCG	63.3	AGUGU	49.1	UUACU	42.2
84	UGUUC	62.3	AUGAA	49.0	UAUCU	42.0
85	CUUUG	62.3	UAUUG	48.7	GUGCU	41.7
86	AUUGU	62.0	CACGU	48.3	AGUUU	40.7
87	UAUUG	60.9	AUGGC	47.9	UGAAU	40.4
88	AAUGU	60.7	GUGCG	46.8	AUUGU	40.3
89	GUUUA	59.2	AUGAU	46.4	UGUUC	40.0
90	AUUGC	59.0	UGUUC	45.8	CGCUU	39.2
91	AUGAU	58.0	AUUUG	45.2	UGCUA	39.1
92	UCGCU	57.9	UGCGC	44.9	GUUGC	39.1
93	GCGUG	57.6	AUGAG	44.6	GAAUG	38.6
94	GCACG	57.4	UAUGA	44.1	UUCUC	38.3
95	UGUGG	56.4	AUUGU	44.1	UGUUA	38.0
96	UUUGG	54.1	UUGCU	43.7	UAAUU	37.6
97	CAUAU	53.7	AAUGU	42.6	AAUGU	37.6
98	GCAUU	53.5	UGCUG	41.9	UCGCU	37.0
99	UGCGC	53.1	ACGUG	41.7	CUGCU	36.7
100	GAUUU	53.0	UCGCU	41.1	GUGUG	36.6

101	GCUGC	52.0	UAUUU	41.0	UACUU	36.3
102	UACAU	51.6	GUUGU	40.8	GUUCU	35.8
103	UGGUU	51.5	CAUUU	40.1	CUUAU	35.6
104	UUCGC	51.3	GUUGC	40.0	UUAAU	35.4
105	GUCUU	51.2	GCUGU	40.0	CCUGU	35.1
106	CUGUA	49.8	GGUAU	40.0	CUGUA	35.0
107	CGUGC	49.4	UUUUA	39.2	UGGCU	34.6
108	UUCUG	49.4	GCGCA	38.8	UUUCA	34.3
109	GCAUA	49.4	GUUUA	38.7	CGUGU	34.1
110	GUACA	49.2	UUGUC	37.9	UCUAU	33.9
111	UUUCG	48.7	GAUGU	37.8	UAGUU	33.4
112	CGUUU	48.2	UACGU	36.9	GUUAU	33.4
113	CACGU	47.8	CACAU	36.8	UUGCC	32.7
114	UGCUA	47.3	GUACG	36.6	UCGUU	32.4
115	UACGU	46.8	UGUAA	36.6	AUUUA	31.8
116	UCAUG	46.6	GCUGC	36.1	CUGCC	31.6
117	AGUGU	45.8	CUGUA	36.0	AUGUC	31.6
118	GUGCG	45.6	ACGCA	35.7	CUUUA	31.0
119	GGUUU	45.6	UCGCA	34.4	UAUUG	30.6
120	UGUUA	45.4	UGUUA	34.1	GAUUG	30.5
121	UUUAC	44.8	GUCAU	33.8	UACGU	30.5
122	GAUGU	44.8	GAGUG	33.6	UGACU	29.4
123	CGUAU	44.5	UGGCA	33.6	GAUGU	29.3
124	CCUGU	43.0	GUCGC	33.5	UUAUA	28.9
125	CCUGC	43.0	CCUGC	33.3	UGUGA	28.6
126	AAUGC	42.3	CUAUG	32.7	GUAUA	28.6
127	UCUAU	42.3	UGUCG	32.1	GCUAU	28.5
128	AUGCC	42.0	GUGCC	32.0	UACCU	28.3
129	UGAGU	40.9	GCACA	31.7	UAUAU	28.1
130	AUUUG	40.9	CCUGU	31.3	UUCCU	27.6
131	UAUGA	40.8	UUUUC	30.7	CCUGC	27.5
132	UUCGU	40.2	CUCGU	30.5	UAUUC	27.4
133	UGGCU	40.1	UGACU	30.4	CUGUC	27.2
134	UUGCC	39.9	GUAUC	30.3	CGUAU	27.0
135	UCUGC	39.3	UUCAU	29.8	GUUGA	26.8
136	GUACG	38.5	UGAUU	29.4	ACUGU	26.7
137	CUGAU	38.4	CUUUG	28.5	UUGAA	26.4
138	UUGAU	38.3	AUGGU	28.4	CCCGU	26.1
139	CAUUU	38.0	GAUUG	28.4	UUGAC	25.3
140	UUCAU	37.3	UUCGC	27.9	AUAUU	25.1

141	UGUCG	36.8	CGUUU	27.7	GGCUG	25.1
142	UAUCU	36.8	UGUCA	27.3	CGUCU	24.9
143	UGACU	36.3	UGUAG	26.5	GUACG	24.5
144	GUUAU	36.2	AGUGC	25.9	GUGCG	24.3
145	UUGGU	35.8	GGUGU	25.9	CCGUU	24.3
146	CGCUU	35.8	UUAUU	25.8	UUGCG	24.2
147	CUAUG	35.8	UGCCG	25.8	CUUCU	24.1
148	GAUUG	35.2	ACUGC	25.7	UCCUU	24.0
149	GGUGU	34.6	CCGUG	25.6	UGCGU	24.0
150	GUAUC	33.8	CUGUU	25.3	UGCGC	23.9
151	UCUUG	33.4	CUUUU	25.2	UUCGG	23.6
152	UACUU	33.1	UGCUA	24.9	CUAUG	23.3
153	GGUAU	32.8	CGUAU	24.4	UUCCG	23.3
154	UUUAA	32.3	CGCUG	24.1	UCGCG	22.8
155	CUUAU	31.8	UCGUG	23.9	ACGUU	22.7
156	UUUGA	31.5	AGUAU	23.4	CUUUC	22.4
157	GUCUA	31.3	CAUUG	23.4	AUUUC	22.3
158	UUACU	30.7	UGAGC	22.8	UUAUC	22.2
159	CUAUU	30.4	ACGUC	22.6	GCACG	22.1
160	CUGUC	30.3	GUCUA	22.5	GUGAU	22.0
161	AGUAU	30.3	UUGCC	22.5	UAUGA	21.8
162	GUGAU	29.7	UGUCC	22.2	UAAUG	21.4
163	UCGUU	29.5	AUAUA	21.5	UCUGA	21.1
164	GCGUU	29.4	GAUGC	21.4	UGAUA	21.1
165	UCGCA	29.2	UGGGU	21.1	UCCGU	21.1
166	UUUCC	29.2	GCGUU	21.0	GCUGC	20.6
167	ACUUU	29.1	UCUUU	20.3	CCUAU	20.6
168	GUGCC	29.1	GCUUU	20.2	AGUGU	20.3
169	ACUGU	29.1	UGAUG	20.0	GGUGU	20.1
170	CCUUU	29.0	CACGC	20.0	UAUGC	19.9
171	CCGUG	28.9	AUGAC	19.9	UUAAC	19.8
172	UCCGU	28.4	AUGGG	19.8	GCGUU	19.8
173	AGUUU	27.9	CGUCA	19.5	GGAUU	19.8
174	UGUCC	27.8	AUCGC	19.5	UGUGG	19.7
175	CUGCU	27.7	CUGCG	19.2	UGUCG	19.7
176	ACGUU	27.6	UGCGA	18.9	AUUCG	19.5
177	GUUUC	27.2	ACACG	18.4	UUACG	19.1
178	GCGCA	27.1	GUUAU	17.7	UUCAU	19.0
179	GCUAU	27.0	UAUGG	17.6	GUACU	18.9
180	UUGCG	27.0	UGGAU	17.2	GCUUA	18.9

181	CGCUG	26.8	UAUCU	17.2	GAUGA	18.7
182	ACGUG	26.8	UUUAC	17.2	CAUUU	18.6
183	UGUAA	26.1	GUGAU	16.9	UGUCC	18.5
184	CUCGU	25.0	UGGCU	16.7	GUCUA	18.4
185	UCGUG	24.4	AUCUG	16.7	AUCUU	18.3
186	CGUCU	24.4	GCUCG	16.7	GUUCG	18.2
187	UGUAG	24.3	CUGUC	16.4	GUCCG	18.2
188	GUCAU	24.3	UCUAU	16.3	CGUUA	18.1
189	GGCAU	24.2	CAUCU	16.2	CUAUC	18.1
190	UGUCA	24.0	UCCGU	15.6	GCUUC	18.0
191	GAUGC	23.9	GUGAG	15.5	ACGCG	18.0
192	UUAUA	23.7	UUUCU	15.0	UGCAC	18.0
193	ACGUC	23.6	CGUUG	14.8	UUAGU	17.9
194	GUCGC	22.3	UGCUC	14.8	CGGAA	17.9
195	UGCGA	22.3	UUGCG	14.3	CACGU	17.8
196	UACCU	22.3	GGUGC	14.2	AUUCU	17.8
197	GUUCU	22.0	UCGUC	14.0	AGUAU	17.7
198	ACGCA	21.0	UCGUU	13.9	CCGGU	17.5
199	AUAUU	21.0	CCAUG	13.7	UAUUA	17.4
200	GUACU	20.9	UGAUC	13.5	CUAGU	17.4
201	UUCUC	20.9	UAGCA	13.5	CGCGA	17.3
202	GUUCG	20.5	GUUCA	13.4	UGACG	17.2
203	UAAUG	20.4	GGUCG	13.4	UCGGU	16.8
204	UUACG	20.0	UAUAC	13.2	ACGUA	16.7
205	AUGGC	19.9	CGGCA	12.9	ACUAU	16.7
206	AAUUU	19.8	CUACG	12.5	UACUG	16.6
207	AUGAC	19.7	ACGUU	12.4	UCUUA	16.5
208	AUGGU	19.7	CGCUU	12.2	AUUUG	15.7
209	UGGGU	19.6	CGUGA	12.1	UAGUG	15.2
210	UUUAG	19.4	GUCGU	12.0	UUGGC	15.0
211	ACGUA	19.3	UACGC	11.9	GGCUU	14.7
212	CUGCG	19.1	GCCGU	11.5	CCGUG	14.7
213	CGUCA	18.9	UAAUG	11.3	CUUCG	14.5
214	AUGAG	18.2	CUCGC	11.1	AUAGU	14.4
215	CGUUG	18.1	ACGUA	10.9	CGUUC	14.3
216	UACGC	18.0	CGCGU	10.8	AGCUU	14.2
217	UUAAU	18.0	UCUCG	10.7	CUCUU	14.0
218	GUCGU	17.9	UAUCG	10.7	CCGUC	13.8
219	CCCGU	17.8	UGGUU	10.4	ACGUC	13.8
220	UAGUU	17.6	ACGCG	10.1	GUAGU	13.8

221	AUCUG	17.6	GCACU	10.0	GUAUC	13.6
222	UGACG	17.4	ACUGU	9.9	CGCGG	13.5
223	CCGUU	17.3	GUUCG	9.4	UUACC	13.4
224	GCAUC	17.2	UUACG	9.2	UCGUG	13.2
225	AUCGC	16.9	GUCUU	9.0	GUCGU	13.1
226	GGCUG	16.9	CGCAA	8.9	UAGUA	13.1
227	ACGCG	16.8	UGACG	8.9	UCAUG	13.1
228	AUGAA	16.8	UAAGU	8.7	GUGGU	13.0
229	AUCUU	16.6	UCGGC	8.7	UGCGA	12.6
230	UGAUC	16.6	CGUCG	8.6	UUCUA	12.6
231	UUGGC	15.3	GUACU	8.4	CUAUA	12.2
232	GCUCG	15.2	UCGUA	8.1	UUCCC	12.1
233	UGGAU	15.1	UACUU	8.1	GAUGC	11.8
234	UAAUU	14.8	GGGUG	7.8	UUAAG	11.7
235	GUCCG	14.7	UGCAA	7.7	CGAGU	11.5
236	CGUUA	14.7	UUCGU	7.6	CUGGU	11.4
237	UACUG	14.4	GUGAA	7.5	AUUGC	11.1
238	UUCGG	14.4	UUGGC	7.2	AUAUG	11.1
239	UCGUC	14.1	CGUCU	7.2	UGGUA	11.1
240	UCGUA	14.1	UACGA	7.2	CGUUG	10.9
241	ACACG	14.1	AUACG	7.2	AACUU	10.9
242	UAUUC	13.5	CGAAU	7.1	UACGC	10.8
243	UUGAC	13.5	CUUAU	7.1	UCGGA	10.8
244	AUUCG	13.4	AUCUU	6.7	AUGCG	10.5
245	CGCAA	13.3	GCGAU	6.6	CUGCG	10.4
246	CGUUC	13.2	AUCGU	6.6	GUGCC	10.3
247	GCCGU	13.2	CAUCG	6.5	UCUUC	10.3
248	UGCCG	13.2	UAGCG	6.5	AUCUG	10.2
249	GACUG	13.0	CUCUG	6.4	UCGUA	10.2
250	CCUAU	12.9	UCGAU	6.4	GGUAU	9.9
251	UUAUC	12.8	GACUG	6.2	GACUG	9.8
252	CAUUG	12.6	GCUAU	6.0	UACUA	9.8
253	CGCGG	11.9	CGUAC	6.0	UGGAU	9.7
254	GUUGA	11.9	AUCCG	5.9	CUUGA	9.6
255	UCGCG	11.8	CAUAC	5.9	UCGAC	9.4
256	CACGC	11.7	CGGAU	5.9	UGAGU	9.3
257	AUUUA	11.5	CGUUC	5.5	UUAGC	9.2
258	UUCCG	11.5	UCAUA	5.4	UCCAU	9.1
259	CGCGU	11.5	GUCCG	5.4	AUUAU	9.0
260	GCUUA	11.3	UCUUG	5.3	UGUAG	9.0

261	UAAGU	11.3	GCGUA	5.3	CCUUG	8.9
262	UAUCG	11.3	CCGCA	5.1	UUGGG	8.8
263	CUUCG	11.2	AUUCG	4.6	AGAUU	8.7
264	CUAUC	10.9	CGCGG	4.4	UAAGU	8.7
265	CGUAC	10.9	GUCUC	4.3	CGAUG	8.7
266	GCGAU	10.8	GAUCG	4.2	GCGUG	8.6
267	AUAUA	10.7	GGGUA	4.2	GCCGU	8.4
268	CUGCC	10.5	CGAUU	4.0	CGCAA	8.3
269	CCAUG	10.5	GCGCG	4.0	ACGGA	8.0
270	UCUCG	10.4	AUGGA	3.8	CCCCU	7.9
271	CGAGU	10.2	CGCCU	3.8	UGUCA	7.8
272	UGAUA	10.0	GUAGC	3.7	GACUU	7.8
273	UAUAC	10.0	UUUCG	3.6	UCCUA	7.8
274	CGCGA	9.9	CCGCG	3.5	GUAAU	7.7
275	CUUUA	9.8	UUCUG	3.5	CAUGU	7.7
276	UUAAC	9.6	GUCGA	3.5	CGCGU	7.5
277	UUUCA	9.4	CCUCG	3.5	UCAUU	7.5
278	UAUGG	9.4	UCCAU	3.3	CGUAC	7.4
279	UAGCG	9.4	UAUAG	3.0	UGAUC	7.4
280	UCCAU	9.3	AUAUU	3.0	UAGCG	7.3
281	UGGUA	9.2	ACUUG	2.9	GUACC	7.3
282	CUCGC	9.0	UGGUA	2.9	CGGAC	7.2
283	GAGUG	8.9	GUACC	2.9	GCCUU	7.0
284	GCUUC	8.9	GUAAG	2.7	UAACG	6.8
285	CAUCG	8.7	CCGUU	2.5	CCGUA	6.7
286	GUGAA	8.7	CGCAC	2.5	AUGAC	6.6
287	UUCCU	8.4	UACCU	2.3	UAUCG	6.6
288	CUACG	8.0	CGAGU	2.3	ACGAC	6.6
289	UACGA	7.5	GAGCA	2.2	CUUGG	6.6
290	GUACC	7.5	UUUGG	2.1	GCGAU	6.5
291	GUGGU	7.4	CCGUA	2.0	AUCGC	5.9
292	CGUGA	7.4	GUCGG	2.0	CGGUG	5.9
293	CUAUA	7.0	CCCGU	1.8	AUCUA	5.8
294	ACUGC	6.6	CGCGC	1.8	CGUCA	5.6
295	CGUCG	6.5	CAACG	1.4	CGGUU	5.6
296	UCGGU	6.4	UACUG	1.4	GUUCC	5.5
297	GUAAU	6.3	CGCUA	1.3	CCCUU	5.5
298	CCGUA	5.9	CGUUA	1.2	ACGAU	5.5
299	GGUCG	5.8	GCGAA	1.2	GUCGC	5.4
300	CCGCA	5.8	ACGAG	1.1	CUGAC	5.2

301	GCGUA	5.8	UUAUA	1.1	GCGCU	5.0
302	UGGCA	5.7	GACGU	0.9	UCGUC	4.9
303	AUUAU	4.7	GUAAU	0.7	CCUAC	4.9
304	UAUUA	4.7	GAUAU	0.0	UGUAA	4.7
305	CGAUG	4.2	AGCGU	0.0	CAUCG	4.5
306	ACUAU	4.2	CUUCG	-0.3	ACCUU	4.5
307	AUUUC	4.2	GUAAC	-0.3	UUCGA	4.5
308	GUAGU	4.1	UUGGU	-0.4	ACUUA	4.4
309	CGAUU	4.0	AUCAU	-0.4	UGACA	4.2
310	UUAGU	4.0	UGGCG	-0.4	GAUAU	4.1
311	CAUCU	3.9	AACGC	-0.7	GUUGG	4.1
312	UAGUG	3.7	UGCCC	-0.7	GUGAA	4.0
313	GCACU	3.7	UACGG	-1.0	CAGUU	3.9
314	AUACG	3.6	ACGCU	-1.4	CUUAA	3.9
315	UCGAC	3.6	GCUUA	-1.5	UCCCG	3.8
316	UGCAA	3.6	CAUAG	-1.7	GCUCG	3.8
317	CCGCG	3.4	UGGUG	-1.7	CUCGU	3.6
318	UUACC	3.3	GUGGC	-1.8	UCCGA	3.6
319	UCAUA	3.1	AUCGA	-1.9	CGCUG	3.5
320	GAUAU	3.1	UUCUU	-2.0	CUACU	3.5
321	CACAU	3.0	CGAUC	-2.1	UGGUG	3.4
322	UCGAU	2.9	CGAGA	-2.1	UACGA	3.4
323	CUCUG	2.6	CUAUA	-2.1	CGUAA	3.4
324	CGGCA	2.5	CGACG	-2.2	CCGCA	3.4
325	AUCGU	2.4	CUGAU	-2.2	GCGUC	3.2
326	AGCGU	2.3	GGCUG	-2.3	GUCAU	3.1
327	GGAUU	2.2	GCGCU	-2.5	UGCGG	3.0
328	CUUUC	2.2	CUAUC	-2.6	AGCGU	2.9
329	CGGAU	2.1	GCGCC	-2.7	GGUUG	2.8
330	AUCCG	2.0	CGUAG	-2.7	UGGGU	2.6
331	CGCUA	2.0	UUACU	-2.8	UCUCG	2.6
332	AUCUA	1.8	UCGAC	-2.9	CGGCU	2.6
333	UCAUU	1.7	GACGC	-2.9	CCAUG	2.6
334	AUAGU	1.6	GUGGU	-2.9	CUAAC	2.5
335	UCCUU	1.6	GCGUC	-3.0	CGAUU	2.4
336	GUCGA	1.4	CGAUG	-3.0	GCUGA	2.3
337	CGAAU	1.4	GCCGC	-3.1	AACGU	2.3
338	GCGCG	1.2	UCACG	-3.2	UAUCA	2.1
339	GCGCU	1.1	GCGGC	-3.5	GCGUA	2.1
340	GAUCG	1.1	UUCUC	-3.6	UCUGC	2.0

341	AUUCU	1.0	AUACA	-3.6	CUGGC	1.9
342	UCGGA	1.0	AUUAU	-3.7	UUAGA	1.9
343	UGGUG	1.0	CUGCU	-3.8	CGUGC	1.8
344	UGAGC	0.9	GAGCG	-3.8	CCGCG	1.8
345	CCGGU	0.6	CGGCU	-3.8	CUCGC	1.8
346	GCACA	0.6	GUUAC	-3.9	CGCUA	1.8
347	ACGAU	0.5	CGUAA	-3.9	AUAUC	1.7
348	CGGUU	0.5	GUGGG	-4.0	UUGGA	1.6
349	CGUAA	0.3	AUCUA	-4.0	CCGGC	1.6
350	GACGU	0.2	GCCCG	-4.1	GAUUA	1.3
351	UACUA	0.0	ACAUA	-4.3	ACUCU	1.3
352	GUUCA	0.0	UGCCA	-4.4	UAUAC	1.3
353	ACGCU	0.0	GAGUA	-4.5	AAUGA	1.2
354	UAACG	0.0	GAUUU	-4.5	UCUAC	1.0
355	GCGUC	0.0	CGCGA	-4.6	ACCGU	1.0
356	UAGUA	0.0	GGCCG	-4.6	CGAUA	0.9
357	AGCUU	-0.2	GAUCU	-4.8	ACGCU	0.9
358	UGCUC	-0.4	CGGUU	-4.8	GUACA	0.7
359	AGUGC	-0.5	UUCGG	-4.9	UCUCU	0.7
360	UAUAG	-0.8	GUUCU	-5.0	CCGAU	0.7
361	GUCUC	-1.2	UUAUC	-5.1	UCGCA	0.7
362	CGGCU	-1.2	CGCAG	-5.2	CGUCG	0.6
363	ACGAC	-1.3	GCGAC	-5.3	ACACG	0.6
364	GAUGA	-1.3	UCAUU	-5.3	UCGAG	0.3
365	UCGGC	-1.3	UCGGU	-5.4	AUGGU	0.2
366	CUAGU	-1.4	UAUCA	-5.7	CUACC	0.2
367	CCUAC	-1.7	CCUAU	-5.8	GAGUU	0.1
368	CGAUC	-1.8	CCGGC	-5.9	CACGA	0.1
369	CGGAC	-1.8	UAACU	-5.9	ACAUU	0.0
370	AUCGA	-1.9	AGCGA	-6.2	UCCGG	0.0
371	GCGAA	-2.0	ACGAU	-6.4	UCCCC	0.0
372	CCUUG	-2.1	AGCGC	-6.4	CGUGA	0.0
373	UAUCA	-2.1	UAAGC	-6.4	GACGA	-0.1
374	CUUCU	-2.3	UUGAC	-6.6	CCAUU	-0.1
375	CGCGC	-2.3	AACGU	-6.7	UAACU	-0.1
376	UCUGA	-2.3	ACCCG	-6.8	GCCCG	-0.2
377	AACGU	-2.4	CAUUC	-6.8	AAUGC	-0.2
378	UUCGA	-2.5	CGGCC	-6.9	GAUUC	-0.2
379	GUUAC	-2.7	GCUUC	-7.0	ACUGA	-0.3
380	CCGGC	-2.7	CGAAC	-7.1	CGAUC	-0.3

381	UCACG	-2.9	UACCG	-7.1	AUCGA	-0.4
382	UCCGA	-3.0	GGCGU	-7.2	CGGUC	-0.4
383	GCCCG	-3.1	UUGAU	-7.2	GCGGA	-0.4
384	GUUGG	-3.1	CGACA	-7.4	GUUAC	-0.5
385	CUUGA	-3.4	CCUAC	-7.4	UCAGU	-0.5
386	GUAAC	-3.7	UGAUA	-7.4	GAUAC	-0.6
387	UUAGC	-3.7	AGUUU	-7.5	UGGUC	-0.6
388	GUGGC	-3.9	UCGCG	-7.5	CUACG	-0.7
389	GGUGC	-4.0	UCGAA	-7.5	CGGGU	-0.7
390	GGCUU	-4.0	UUCCG	-7.6	UACAU	-0.8
391	GAGUA	-4.2	CCGAA	-7.7	CCGAA	-0.8
392	GUAGC	-4.3	ACGGC	-7.7	AUACU	-0.8
393	CGGUG	-4.3	AUAGC	-7.7	AUAAU	-0.8
394	GCCGC	-4.4	UCGGA	-8.1	UUCAC	-0.9
395	UAACU	-4.4	UUCGA	-8.2	GAGUA	-0.9
396	UUGAA	-4.6	AUAUC	-8.3	UCGAU	-1.0
397	AUAUC	-4.6	AACGA	-8.3	GACGU	-1.0
398	CGCAC	-4.7	GGAUG	-8.4	CGACU	-1.0
399	UCCCG	-4.7	CUCGG	-8.4	UUAGG	-1.1
400	GUCGG	-4.7	UAACG	-8.6	UCAUA	-1.2
401	ACUUG	-5.0	GUGAC	-8.7	GUCGA	-1.2
402	AACGC	-5.0	AUACU	-8.7	GACGG	-1.3
403	CCGUC	-5.2	CCGCU	-8.8	CGCAU	-1.4
404	UACGG	-5.2	CGGAC	-8.8	AUACG	-1.5
405	GGUUG	-5.5	UCCGA	-9.0	GUCCU	-1.6
406	CGCCU	-5.5	GGUUU	-9.0	UACCC	-1.6
407	GGGUA	-5.5	UUAAC	-9.1	GCGGG	-1.7
408	GGGUG	-5.6	UUACC	-9.1	UCACG	-1.7
409	UCUUA	-5.6	UCGCC	-9.2	GCGCA	-1.7
410	CGGAA	-5.7	GUUGG	-9.2	GCAAU	-1.8
411	CUGAC	-5.8	UACAC	-9.2	GUGAC	-1.9
412	CCGAA	-5.9	AGUCG	-9.4	UGCAA	-1.9
413	GCCUU	-5.9	CGGUG	-9.5	ACCUG	-1.9
414	CGAUA	-6.0	GCCUA	-9.5	CGUCC	-1.9
415	CAACG	-6.2	UCAUC	-9.6	UUGAG	-1.9
416	UGACA	-6.3	ACGAC	-9.7	GCGCG	-2.2
417	UCUCU	-6.4	CGAUA	-9.7	CGGAU	-2.2
418	CGUAG	-6.5	CCUUG	-9.7	UAGUC	-2.3
419	CCUCG	-6.6	AUCGG	-9.7	GCCUA	-2.5
420	GACGA	-6.6	UCCGC	-9.8	CCUGA	-2.6

421	AUCAU	-6.7	GCGGU	-9.9	UCACU	-2.7
422	UCUAC	-6.8	GGAUA	-9.9	UUCAG	-2.7
423	CGAGA	-6.8	CGGCG	-10.0	ACGUG	-2.8
424	CUGGU	-6.9	UAGAU	-10.1	AUCCG	-2.8
425	GGCGU	-6.9	UAACC	-10.1	GGCGU	-2.9
426	CGGUC	-6.9	GUUUC	-10.2	GAAUU	-2.9
427	AUACU	-7.0	UCUCU	-10.3	AUUGA	-3.0
428	CGCAG	-7.1	UUAAU	-10.3	GAUCG	-3.1
429	GUGAC	-7.2	GUCCA	-10.4	GCGAG	-3.1
430	GACGC	-7.2	AAUCG	-10.5	AUCGU	-3.2
431	UUGGG	-7.2	UCCCG	-10.5	GCCGC	-3.2
432	CGAAC	-7.2	ACCGA	-10.6	UAUAA	-3.2
433	UGGUC	-7.3	GCAAU	-10.7	UAGAU	-3.2
434	GUGAG	-7.6	UGGUC	-10.8	UAUAG	-3.3
435	CGGCG	-7.7	ACGCC	-10.8	GUGGC	-3.3
436	UGCCA	-7.7	GUAGU	-10.8	CGGCG	-3.4
437	ACAUU	-7.8	CCCGC	-10.9	CUCUG	-3.4
438	GCGGA	-8.0	GUCCU	-11.0	GCGAA	-3.4
439	ACGAG	-8.0	GCGGA	-11.0	GGGUU	-3.5
440	GCAAU	-8.1	ACGGU	-11.1	CCUUA	-3.5
441	GGCCG	-8.1	ACUCG	-11.1	CAUUG	-3.6
442	UACCG	-8.3	CGGUC	-11.1	UAACA	-3.6
443	CUUGG	-8.5	CUCCG	-11.2	CCCCG	-3.7
444	GUCCA	-8.6	CGCUC	-11.4	CGAAC	-3.7
445	CUACU	-8.7	CUCAU	-11.4	CUUAG	-3.7
446	ACGGA	-8.7	AAGUG	-11.4	GGUUC	-3.8
447	CGACA	-8.8	CGACC	-11.4	GUCCA	-3.9
448	UCGAG	-8.8	CGAGC	-11.5	CGAAU	-4.0
449	CGACG	-8.8	AAGCG	-11.6	AGGUU	-4.0
450	CUAAC	-9.0	UAUUC	-11.7	UAUGG	-4.1
451	ACCGU	-9.0	GGCGC	-11.7	GGUCG	-4.2
452	CGACU	-9.0	ACAUU	-11.7	GCACU	-4.2
453	ACGGC	-9.1	CUCGA	-11.8	UACCG	-4.3
454	GCCUA	-9.2	ACUAU	-11.8	AUUAC	-4.4
455	UAGCA	-9.4	AGACG	-11.9	UACUC	-4.5
456	AUCGG	-9.5	UGCAG	-12.0	ACGGC	-4.5
457	UGCGG	-9.5	UCUAC	-12.1	UAGCU	-4.5
458	GUCCU	-9.8	GCUAC	-12.1	UGCCG	-4.6
459	UCCGG	-9.8	GACGG	-12.2	AGUUG	-4.8
460	UAGAU	-9.8	CGUGG	-12.2	CUCAU	-4.8

461	GUAAG	-10.0	GGUUG	-12.3	CGACA	-5.0
462	GCUGA	-10.0	GCGAG	-12.3	CGACC	-5.0
463	UCCGC	-10.3	UAAUU	-12.4	UGCCA	-5.2
464	GUUCC	-10.4	GACAU	-12.4	AUCGG	-5.2
465	UGCCC	-10.4	ACUUU	-12.5	AUAUA	-5.3
466	GCGGC	-10.4	AACCG	-12.5	CGCAG	-5.4
467	CACGA	-10.7	UAGUU	-12.6	UAACC	-5.5
468	CUGGC	-10.8	UGAAC	-12.6	GUAAC	-5.6
469	AGCGA	-10.9	CGAAA	-12.7	CGCGC	-5.6
470	GAGCG	-11.0	CUACA	-12.7	GAACG	-5.7
471	GCGCC	-11.0	CAUUA	-12.9	UCCGC	-5.8
472	CAUAG	-11.0	UAGGU	-12.9	AAUUC	-5.8
473	CGACC	-11.1	GACGA	-12.9	ACCUA	-5.9
474	GACGG	-11.2	ACCGC	-13.0	AUGCC	-6.0
475	GAUCU	-11.2	GUUAG	-13.1	AAAUU	-6.1
476	CUCAU	-11.3	UCCGG	-13.2	UUAAA	-6.1
477	CCGAU	-11.4	GAACG	-13.2	CGAGA	-6.1
478	ACGCC	-11.5	CGACU	-13.2	ACCGG	-6.2
479	UGGCG	-11.5	UACUA	-13.3	CACUU	-6.3
480	ACCUU	-11.7	CCUUU	-13.3	UACGG	-6.4
481	UAACC	-11.9	GCCUU	-13.5	GUCUC	-6.4
482	GACUU	-11.9	ACCGU	-13.5	GACGC	-6.5
483	ACCGA	-12.0	UAUUA	-13.6	AACGC	-6.5
484	GCGAC	-12.0	UAGUG	-13.9	GGUAG	-6.6
485	GCGAG	-12.0	CCGAG	-14.0	UCUGG	-6.6
486	GAUAC	-12.2	GGCGG	-14.0	ACGCC	-6.7
487	UAUAA	-12.2	AUUAC	-14.0	UGAAC	-6.8
488	ACCUA	-12.2	CCGAC	-14.1	AGCGA	-6.8
489	AUGGG	-12.3	CACGG	-14.2	GGCCG	-6.8
490	UGAAC	-12.5	UCGAG	-14.2	GUUAG	-7.1
491	UCGCC	-12.7	CCGGU	-14.5	CACGG	-7.2
492	UACCC	-13.0	CUGAC	-14.5	CACCU	-7.2
493	AGUCG	-13.0	CAUCC	-14.6	CGUAG	-7.2
494	UCAUC	-13.2	GCCGG	-14.6	CUGAA	-7.3
495	UAGGU	-13.3	GACCG	-14.6	CACGC	-7.6
496	AACGA	-13.3	CUUUA	-14.7	UAAAU	-7.6
497	CGGGU	-13.3	CGGUA	-14.8	ACGAA	-7.7
498	CGUGG	-13.5	CACCG	-14.8	ACCGA	-7.8
499	UCUUC	-13.5	GUUAA	-15.1	AAGUU	-7.9
500	AUUGA	-13.6	UUAGC	-15.1	UCUAA	-7.9

501	UUCUA	-13.6	AUAGU	-15.2	UAGGU	-7.9
502	AACUU	-13.6	CUAUU	-15.3	GAUCA	-8.1
503	GAACG	-13.8	AAUUG	-15.3	GCUGG	-8.1
504	AUUAC	-14.0	CGCCG	-15.4	CGAAG	-8.2
505	CACGG	-14.1	UGACA	-15.5	ACAUG	-8.2
506	AAUCG	-14.1	UCGGG	-15.5	CUAAU	-8.2
507	GUUAG	-14.2	AUACC	-15.5	CGGUA	-8.4
508	GAUUA	-14.2	UGCGG	-15.5	CGCCU	-8.7
509	CGGUA	-14.3	AGCUU	-15.8	CCCAU	-8.7
510	GGAUG	-14.6	CUAAC	-15.9	CUUAC	-8.7
511	UUGAG	-14.7	UAGUA	-15.9	GUCAA	-8.9
512	UCCUA	-14.8	AAGCA	-16.1	UCAUC	-8.9
513	ACCUG	-15.0	UAUCC	-16.2	CGUGG	-8.9
514	CACCG	-15.2	AAUAU	-16.4	GUUAA	-8.9
515	CCCUU	-15.3	ACCGG	-16.5	AGUCG	-9.0
516	UAGCU	-15.3	CUGGC	-16.5	GGUAA	-9.0
517	CGGCC	-15.4	CGGGU	-16.5	GGACG	-9.1
518	ACCGG	-15.5	AAUUU	-16.5	CACUA	-9.1
519	UCGAA	-15.5	GAAUA	-16.6	UAAAC	-9.1
520	CUCGA	-15.8	AUUUA	-16.7	AGACG	-9.1
521	AGACG	-15.8	ACCUA	-16.7	CUCUA	-9.1
522	UUCAC	-15.9	GAUCA	-16.8	CAAUU	-9.2
523	ACGGU	-15.9	CCCUG	-16.8	GAUCU	-9.2
524	CUCGG	-16.0	CUACU	-16.9	CUCCG	-9.3
525	AACCG	-16.0	GAUAC	-16.9	UCGCC	-9.3
526	AAUGA	-16.1	CCGAU	-16.9	UGACC	-9.3
527	GGUUC	-16.3	UAGCU	-16.9	GUCGG	-9.4
528	UCAGU	-16.4	UAUAA	-17.0	AUCAU	-9.4
529	GGCGC	-16.5	UUAGU	-17.1	ACUUG	-9.4
530	CCCGC	-16.5	ACGGG	-17.1	GACAU	-9.5
531	CUCUU	-16.6	UUACA	-17.1	AGUUA	-9.6
532	GACAU	-16.7	GCUAA	-17.2	CGGCA	-9.6
533	CUAAU	-16.7	GUAGG	-17.3	GUAGC	-9.6
534	CGUCC	-16.8	CAAUG	-17.4	CGCAC	-9.7
535	CGCUC	-16.8	ACGAA	-17.5	GUAGG	-9.7
536	ACUCG	-17.2	CAGCG	-17.5	CACCG	-9.9
537	UAGUC	-17.2	CCGGA	-17.7	CGAAA	-9.9
538	AUAAU	-17.3	CUUGA	-17.7	CAUCU	-10.0
539	CUACC	-17.4	AAACG	-17.7	AGUCU	-10.0
540	CGAAA	-17.4	GGGGU	-17.8	UAAUC	-10.0

541	ACCCG	-17.5	GGACG	-17.8	CAAUG	-10.2
542	UUGGA	-17.6	CCCGA	-17.8	UCGGC	-10.2
543	GAUCA	-17.6	CCCAU	-18.0	GAAAU	-10.2
544	GCGGG	-17.6	GCCGA	-18.0	CUCGA	-10.2
545	AUACA	-17.7	CGAAG	-18.1	GGUCU	-10.5
546	UACAC	-17.8	GGAUU	-18.1	CAACG	-10.7
547	CGAGC	-18.0	GACUA	-18.3	UGAAA	-10.8
548	CUCCG	-18.1	CAGCA	-18.3	GCUAA	-10.8
549	GGAUA	-18.1	AUUGA	-18.3	AGUUC	-10.9
550	CCCCU	-18.2	ACGGA	-18.4	CAGCG	-11.0
551	CCAUU	-18.2	CAAGC	-18.5	AACCG	-11.0
552	CAUAC	-18.3	GUGGA	-18.5	CGACG	-11.1
553	ACGAA	-18.3	GUUGA	-18.5	UCUCC	-11.1
554	CGAAG	-18.4	CCAUA	-18.7	UAGAC	-11.1
555	AGCGC	-18.4	CAUAA	-18.8	GAGCG	-11.1
556	GUUAA	-18.4	CUUGG	-19.3	GCGGC	-11.2
557	CCCAU	-18.5	GGUAA	-19.4	CCCGA	-11.2
558	UUAAG	-19.1	CGGGC	-19.5	ACGAG	-11.3
559	UAACA	-19.2	GCUGA	-19.6	CGCUC	-11.3
560	CGCCG	-19.3	AGGCG	-19.6	GACCG	-11.3
561	GGACG	-19.3	CCGUC	-19.7	GGAUG	-11.3
562	AGAUU	-19.3	CGGGA	-19.7	ACGGU	-11.7
563	GAGUU	-19.4	CAUCA	-20.1	GUCAC	-11.8
564	GAAUU	-19.4	CUAGU	-20.1	UAAUA	-11.8
565	AAUUG	-19.4	GGCGA	-20.2	ACUUC	-11.9
566	ACUUA	-19.5	UGGGC	-20.4	CUCGG	-11.9
567	GGUAA	-19.5	GCAGU	-20.4	CCACG	-11.9
568	UAAGC	-19.9	AUAAU	-20.5	CCUCG	-11.9
569	GUAAA	-19.9	CAGUG	-20.5	CACUG	-11.9
570	GCGGU	-19.9	ACCUG	-20.5	AUUCC	-12.0
571	AAGCG	-20.1	GUAAA	-20.5	AGCGG	-12.0
572	UUCCC	-20.1	UACCC	-20.7	GCUAG	-12.1
573	GUAGG	-20.3	AUUUC	-20.7	UCCCU	-12.2
574	AGUUG	-20.3	GCACC	-20.8	UCUAG	-12.2
575	CCGAG	-20.5	AUUCU	-20.8	GCCAU	-12.2
576	GGGCG	-20.7	UUGGG	-20.9	CAUAU	-12.2
577	GCUAA	-20.7	UUCCU	-20.9	CCCCC	-12.3
578	GACCG	-20.9	UUGAG	-21.0	GUAAA	-12.3
579	GUGGG	-21.0	CCACG	-21.0	UCGAA	-12.3
580	GAUUC	-21.4	UCAGU	-21.1	ACUCG	-12.3

581	CCACG	-21.5	ACCUU	-21.2	AAUUG	-12.4
582	GCCGG	-21.6	AUUCA	-21.2	CGAGC	-12.4
583	UAUCC	-21.7	AUAAG	-21.2	AAUCG	-12.4
584	CAGUU	-21.7	UCCUG	-21.2	AACGG	-12.4
585	CCCUG	-21.7	CACGA	-21.2	UCAAC	-12.4
586	CCCCG	-21.8	CGGAG	-21.3	GAUAA	-12.5
587	ACGGG	-21.9	GGUUC	-21.4	CUUCA	-12.5
588	GCUAC	-21.9	CUGGU	-21.4	ACUAC	-12.6
589	CGCCA	-22.1	CUAAU	-21.6	GUUCA	-12.6
590	CCCGA	-22.4	CUGCC	-21.6	GGGCG	-12.6
591	GUCAA	-22.6	GAUUA	-21.7	GGGUA	-12.6
592	GACUA	-22.8	CGCCC	-21.8	GCGCC	-12.6
593	AUGGA	-22.9	AACUG	-22.1	CGCCA	-12.7
594	GGCGG	-22.9	GGUAC	-22.1	AACGA	-12.7
595	CAGCG	-22.9	ACAUC	-22.2	ACGCA	-12.8
596	CAAUG	-23.0	UUUAG	-22.2	GCGAC	-12.9
597	GGGUU	-23.3	GGCUU	-22.2	CACCC	-13.0
598	GAAUA	-23.5	GGAAU	-22.3	UAUCC	-13.1
599	CGGGA	-23.8	UCCCU	-22.4	AUUGG	-13.2
600	CUCUA	-23.8	AUCUC	-22.6	CCUAG	-13.3
601	CACUU	-23.8	UUCAC	-22.7	CCCGC	-13.3
602	CCUGA	-23.9	GCUAG	-22.7	CAUAG	-13.3
603	AAUAU	-23.9	AGCGG	-22.8	UGCCC	-13.3
604	CCGCU	-24.0	CGUCC	-22.9	AAUCU	-13.4
605	CCGAC	-24.0	AUCAC	-22.9	AGCCG	-13.4
606	GCCGA	-24.1	UCUGA	-23.1	AUCAC	-13.5
607	AUCUC	-24.2	UCAAU	-23.4	GGAUA	-13.6
608	ACUCU	-24.2	GGGCG	-23.4	AUCUC	-13.9
609	CACUG	-24.3	GCAAC	-23.4	CCAUA	-13.9
610	CCGGA	-24.5	CGCCA	-23.5	GACCU	-13.9
611	CAUUC	-24.5	GAUAA	-23.5	GACUA	-13.9
612	UACUC	-24.5	GUUCC	-23.6	ACUGC	-14.1
613	ACUGA	-24.5	UUUAA	-23.8	UGGCG	-14.1
614	CCAUA	-24.6	UUUCA	-23.8	AUGAG	-14.3
615	AUCAC	-24.7	CUAGC	-24.1	GCACC	-14.3
616	AUACC	-25.0	AACGG	-24.1	CGCCG	-14.4
617	UUAGA	-25.1	GCUCA	-24.2	ACCCU	-14.4
618	CUUAC	-25.1	CAAUC	-24.3	CCCUG	-14.6
619	UCGGG	-25.2	GAUGA	-24.4	CCCUA	-14.6
620	CUUAA	-25.3	UUUGA	-24.5	CGCCC	-14.6

621	ACAUA	-25.4	GCUCC	-24.5	GUAGA	-14.7
622	CUACA	-25.5	CCCGG	-24.6	CCUUC	-14.7
623	GCACC	-25.5	CUUUC	-24.7	CUAGA	-14.8
624	UUACA	-25.6	CACUU	-24.8	GAGUG	-14.8
625	GGGGU	-25.8	UACCA	-24.9	CGGGA	-14.9
626	AUUGG	-25.8	CCCCG	-25.0	AGUAC	-15.2
627	GUAGA	-25.8	GGGAU	-25.0	AAUAU	-15.3
628	UCCCU	-25.9	ACAGC	-25.0	GCGGU	-15.3
629	AAGUG	-26.2	CAAGU	-25.1	UAGGA	-15.4
630	GAUAA	-26.3	UUUCC	-25.1	ACCCC	-15.4
631	AAACG	-26.5	GGGCA	-25.2	AUUAG	-15.4
632	UGCAG	-26.5	UAAUA	-25.3	CUUCC	-15.5
633	AGCGG	-26.6	UAGUC	-25.4	GGCGC	-15.5
634	GGAAU	-26.6	UCCUU	-25.4	UACAC	-15.8
635	GCUAG	-26.7	UAGGC	-25.4	GGAAU	-15.8
636	CCCGG	-27.0	ACUAA	-25.5	GUAAG	-15.8
637	GCCAU	-27.0	GAAUU	-25.6	UGCUC	-16.0
638	AUUCA	-27.1	UAAUC	-25.6	CCGAC	-16.0
639	UAAUA	-27.2	CGGAA	-25.7	ACACU	-16.1
640	CGCCC	-27.2	CGAGG	-25.8	AACAU	-16.1
641	UCACU	-27.2	AUAAC	-26.0	CCACC	-16.1
642	UCUGG	-27.3	GUCCC	-26.2	AAACG	-16.2
643	GUGGA	-27.4	CUUCU	-26.2	CCCGG	-16.4
644	UAAAU	-27.5	GCCAU	-26.2	AAGCG	-16.5
645	UAAUC	-27.6	GGUCA	-26.3	UACAA	-16.6
646	ACCGC	-27.7	GUCAG	-26.4	ACCUC	-16.8
647	CGGAG	-27.7	CCCUU	-26.5	CGGCC	-16.8
648	AAGUU	-27.9	CACUG	-26.5	UCACC	-16.9
649	ACAUC	-27.9	UACUC	-26.5	ACUAA	-16.9
650	UCCUG	-27.9	CUACC	-26.5	ACCCG	-17.0
651	CAUUA	-28.1	AUUGG	-26.6	UGGCC	-17.0
652	AACGG	-28.1	AACAU	-26.7	GGGGU	-17.1
653	GGUCU	-28.1	CGGGG	-26.7	UCAAU	-17.1
654	ACUAA	-28.1	CCGCC	-26.7	GAAUA	-17.1
655	UCUAA	-28.2	UGGAC	-26.9	GGGGG	-17.2
656	AACAU	-28.2	GAUUC	-27.0	UAGAG	-17.2
657	UCUCC	-28.2	GUCAA	-27.0	CAAUC	-17.4
658	AUAGC	-28.2	CAACU	-27.0	GAGUC	-17.4
659	UAAAC	-28.3	UCAAC	-27.1	AGGUA	-17.6
660	AGCCG	-28.3	CCUGA	-27.2	GCAGU	-17.7

661	CAUCC	-28.4	CUUAC	-27.2	CCGAG	-17.7
662	GCAGU	-28.6	UCCUA	-27.2	GAACU	-17.7
663	UGGAC	-28.7	AUCCC	-27.3	GAGCU	-17.7
664	UGGCC	-28.7	GUAGA	-27.3	GACUC	-17.8
665	UCCCC	-29.0	GCGGG	-27.3	UCGGG	-17.8
666	AAAUU	-29.2	UGGCC	-27.3	AUGGC	-17.8
667	UCAAU	-29.3	AUCCA	-27.3	GAUAG	-17.9
668	AGUUA	-29.4	CUCUA	-27.4	CUCCC	-18.0
669	AGUCU	-29.6	GGGGG	-27.6	CCCAC	-18.1
670	UCAAC	-29.6	GCUCU	-27.7	GCUAC	-18.1
671	UAGAC	-29.7	ACUAC	-27.7	CAACU	-18.2
672	GAGCA	-30.3	AAAUG	-27.7	GGGUG	-18.3
673	UUAGG	-30.3	ACCAU	-27.7	AUUCA	-18.5
674	GGUAG	-30.5	UAACA	-27.8	GGACU	-18.5
675	ACUAC	-30.7	UCUCC	-28.0	UUACA	-18.5
676	CAGUG	-30.8	GCCCU	-28.0	GAAGU	-18.6
677	AACUG	-30.8	GCCCA	-28.1	GGUGC	-18.6
678	CACCU	-30.8	AGUUG	-28.1	AGAGU	-18.7
679	AUAAC	-30.8	UAAAU	-28.3	UUCCA	-18.7
680	UUCAG	-30.9	GCAGC	-28.3	CCGCC	-18.8
681	CUAGC	-30.9	GGAUC	-28.7	AUGAA	-18.8
682	CGGGC	-31.0	CUCUC	-28.7	AGCGC	-18.9
683	ACUUC	-31.1	CCGGG	-28.8	AUAGG	-18.9
684	UGACC	-31.1	GGGUU	-29.0	AAUAC	-19.0
685	CCCCC	-31.3	AGUCU	-29.1	AACUG	-19.1
686	CCUUC	-31.3	GAGUU	-29.2	AGUAG	-19.1
687	CUAAG	-31.6	AAUGA	-29.3	CUACA	-19.1
688	AGUAC	-31.7	AGUAC	-29.3	AGUAA	-19.2
689	CCGCC	-31.7	CACAC	-29.3	UGGAC	-19.3
690	UGGGC	-31.7	GACCU	-29.5	GGUUA	-19.3
691	CAACU	-31.9	CCCCC	-29.7	UGAGA	-19.3
692	AGGUU	-32.2	GGCAC	-29.7	CCGGA	-19.4
693	GGUAC	-32.2	UCACC	-29.8	UCCUC	-19.5
694	GACCU	-32.3	ACACA	-29.8	AUUAA	-19.5
695	GGGAU	-32.5	UGGGG	-29.9	CUCUC	-19.6
696	CUUCA	-32.5	UUGGA	-30.0	GCCGA	-19.6
697	GGCGA	-32.6	AGAUG	-30.0	UAGCC	-19.7
698	CACCC	-32.8	UCUUA	-30.1	AUACC	-19.8
699	CAAUC	-32.9	GCCUC	-30.1	UAAGC	-19.8
700	AAUUC	-33.0	GACUU	-30.3	AAACU	-19.9

701	GAAGU	-33.1	UUCUA	-30.5	GAUGG	-19.9
702	CUUAG	-33.2	GAACA	-30.5	GGCGA	-19.9
703	UCUAG	-33.3	CUAAG	-30.5	CUGAG	-20.0
704	CGAGG	-33.5	CUCCC	-30.6	CCUCU	-20.1
705	CCCUA	-33.8	UCUAA	-30.6	ACGGG	-20.2
706	GGGGG	-33.9	CCUUC	-30.8	GGCUA	-20.3
707	AAUAC	-33.9	CUUCA	-30.8	GGGAU	-20.3
708	UCACC	-33.9	GGUUA	-31.1	CUAAG	-20.3
709	UACAA	-34.0	AUAGG	-31.2	UAAGA	-20.4
710	GCAAC	-34.2	AGUUA	-31.3	AUCCU	-20.5
711	CAAUU	-34.3	AACCU	-31.3	UAGAA	-20.5
712	CUAGA	-34.4	AGCCG	-31.3	AGAAU	-20.6
713	UAGGC	-34.6	AGCUG	-31.4	AGUGA	-20.6
714	AAAUG	-34.6	CCAUU	-31.6	UCCAC	-20.6
715	UACCA	-34.7	AACAC	-31.6	CCCUC	-20.9
716	CCUUA	-34.7	CACUC	-31.8	GAAUC	-20.9
717	GUCAC	-34.7	AAUAC	-31.9	UGAAG	-21.0
718	AAUCU	-34.9	CUUCC	-32.2	AGACU	-21.1
719	UGAAA	-34.9	AUAGA	-32.3	CGGAG	-21.1
720	AAGUA	-35.1	CCCCU	-32.3	AAAUG	-21.1
721	GCUGG	-35.2	ACCCC	-32.3	GGUCC	-21.2
722	GGUUA	-35.2	AGCUA	-32.3	AAGGU	-21.2
723	CCUAG	-35.2	UCUGG	-32.5	GACAC	-21.2
724	GAUAG	-35.4	ACCUC	-32.7	AACUA	-21.2
725	AUUCC	-35.6	UCUUC	-32.8	CGGGC	-21.4
726	UCCAC	-35.8	GAGCU	-32.8	ACAUC	-21.4
727	AGGUA	-35.8	ACCAC	-32.9	GCCGG	-21.4
728	AUAGG	-36.0	CCUAG	-32.9	GCCUC	-21.4
729	CACUA	-36.2	CAGUU	-33.0	CUAGC	-21.5
730	GAGUC	-36.2	AACUU	-33.0	UGAGC	-21.6
731	ACCCC	-36.2	AUCAG	-33.1	AAUGG	-21.6
732	CUCCC	-36.4	GCAAG	-33.1	CCAUC	-21.7
733	CUCUC	-36.6	GGUGG	-33.2	CAGAU	-21.7
734	ACCAU	-36.7	GAUCC	-33.3	AAUUA	-21.8
735	AUCCC	-36.7	CACCU	-33.4	UAGGG	-21.8
736	AGGCG	-36.8	AAGUU	-33.4	GGUAC	-21.9
737	AACUA	-37.1	UGAGA	-33.7	CGAGG	-22.1
738	CGGGG	-37.2	CACUA	-33.7	CCGCU	-22.2
739	GGUCC	-37.3	GCCAC	-33.8	GCCCC	-22.2
740	CUUCC	-37.3	GGGUC	-33.8	CAGUA	-22.4

741	ACCCU	-37.9	GAGAU	-33.8	UAGCA	-22.4
742	GAUCC	-38.4	ACUAG	-33.9	GAGAU	-22.5
743	AACCU	-38.4	CCAUC	-34.1	AGGAU	-22.6
744	GCUCC	-38.4	UCUAG	-34.1	ACCGC	-22.7
745	GAAAU	-38.8	ACUGA	-34.3	CUCCU	-22.8
746	AUAAG	-38.8	GAGUC	-34.6	UCUCA	-22.9
747	AGUUC	-38.9	UAAAC	-34.7	CCUCC	-23.1
748	UCCUC	-39.0	AAAGC	-34.8	AUACA	-23.2
749	GCAGC	-39.0	GAAUC	-34.9	ACUAG	-23.2
750	CCAUC	-39.0	AGGUG	-34.9	AGAUC	-23.3
751	GACUC	-39.1	GCCCC	-34.9	CCUAA	-23.3
752	ACACU	-39.2	GUCAC	-35.0	GAAAC	-23.3
753	GGUGG	-39.4	UCCAC	-35.2	CACUC	-23.5
754	GAAUC	-39.5	GGUCU	-35.4	AAGUA	-23.5
755	GUCAG	-39.6	CUAGA	-35.4	UCCUG	-23.5
756	CACUC	-39.8	UAGAC	-35.4	ACACC	-23.6
757	CUCCU	-39.8	AAGUA	-35.4	UACCA	-23.8
758	AGUAA	-39.9	GGUGA	-35.7	AAGUC	-23.9
759	GCCUC	-39.9	UGACC	-35.7	AAGUG	-24.0
760	AAUGG	-40.0	ACUCU	-35.8	AGUCA	-24.0
761	CCACC	-40.1	ACAAC	-36.0	UGGCA	-24.1
762	GAGAU	-40.1	GAACC	-36.1	CCACU	-24.2
763	UGAGA	-40.3	AAUUC	-36.2	GUGGA	-24.3
764	CCAAU	-40.9	GCUGG	-36.3	GUGCA	-24.5
765	AUCCA	-41.0	AGAUU	-36.4	CCAAU	-24.6
766	ACCUC	-41.1	GGUCC	-36.5	GGCGG	-24.6
767	CAGUA	-41.4	AACUA	-36.6	AUAAC	-24.6
768	GUCCC	-41.4	AAGCU	-36.7	UUGCA	-24.7
769	AUUAG	-41.4	UCACU	-36.7	AACCU	-24.9
770	AGAUG	-41.5	CCCUA	-36.8	GCAAC	-24.9
771	AACAC	-41.5	UACAA	-36.9	CAAAU	-25.0
772	GAACU	-41.6	CUGAG	-36.9	UAGGC	-25.1
773	AGUAG	-41.7	UCCUC	-37.0	CAUUA	-25.2
774	CCUAA	-41.7	GACUC	-37.0	CAUUC	-25.3
775	GGUGA	-41.8	GGCCC	-37.0	UUCAA	-25.3
776	GAGCU	-41.9	UCACA	-37.0	ACAAU	-25.3
777	GCCCC	-42.1	ACUUA	-37.0	AGUGC	-25.3
778	GGGUC	-42.3	UGAAA	-37.2	AGGCU	-25.4
779	GCAAG	-42.4	GGCUC	-37.2	CAGUG	-25.4
780	CCACU	-42.4	CACCC	-37.3	AUAGA	-25.4

781	CCGGG	-42.7	GCCAA	-37.5	UGGGC	-25.4
782	CAUAA	-42.8	GAGAG	-37.6	GGUGG	-25.5
783	AUAGA	-42.9	CCUUA	-37.7	CCAGU	-25.6
784	GCCCU	-43.0	CUCUU	-37.7	AAUCC	-25.7
785	UUCAA	-43.1	GAAGU	-37.7	CAGUC	-25.8
786	CAAGU	-43.1	CAAUA	-37.8	AAUAG	-25.9
787	ACACC	-43.1	AAAUU	-37.8	GACCC	-26.0
788	CACAC	-43.2	GAUAG	-37.9	GUCAG	-26.3
789	CUGAA	-43.4	ACAGU	-38.0	ACAGU	-26.4
790	GGUCA	-43.4	CCUCC	-38.0	ACAUA	-26.5
791	GAACA	-43.5	UUGAA	-38.1	GGACC	-26.6
792	UAGAG	-43.5	CAGCU	-38.2	AUCCC	-26.9
793	ACUAG	-43.7	CUCCU	-38.3	ACUCC	-26.9
794	AUCCU	-44.1	UCCCC	-38.3	CAUCC	-27.1
795	CAGAU	-44.2	AGGUU	-38.4	ACUGG	-27.4
796	GGAUC	-44.3	GGCCU	-38.6	CAUGA	-27.5
797	AAUUA	-44.4	ACACU	-38.9	GAUCC	-27.6
798	AGAAU	-44.4	UGAAG	-38.9	UGCAG	-27.6
799	CCUCC	-44.6	AGCCU	-39.0	GUGGG	-27.9
800	UAGAA	-45.4	AUUAG	-39.1	GCAAG	-28.3
801	GCUCU	-45.5	UUCAG	-39.2	AGUCC	-28.3
802	CCCUC	-45.9	GACCA	-39.3	GCCCU	-28.4
803	GAUGG	-45.9	GCCAG	-39.3	ACCAU	-28.4
804	ACAAC	-46.0	AGAAU	-39.3	AAAUC	-28.6
805	CAAGC	-46.0	CAGUA	-39.5	GAGGU	-28.7
806	CCUCU	-46.0	AUUCC	-39.5	GGGCU	-28.7
807	UGAAG	-46.1	CUGAA	-39.5	CAACC	-28.8
808	GCCAA	-46.1	CAAAU	-39.6	CAGCC	-28.8
809	AAUCC	-46.2	CCACU	-39.7	CACAU	-29.1
810	AGAUC	-46.3	UUAGA	-39.9	CGGGG	-29.3
811	AAGGU	-46.4	AAUGG	-39.9	AGGCG	-29.3
812	ACUGG	-46.4	UCUCA	-40.0	GGAUC	-29.3
813	UAAGA	-46.6	CCUAA	-40.0	CUCAA	-29.4
814	GGACU	-46.6	UUAAG	-40.1	GGCAC	-29.4
815	UAGCC	-46.8	UCAGC	-40.3	CUCAC	-29.4
816	UCUCA	-47.2	ACACC	-40.3	GCACA	-29.6
817	CAACC	-47.3	ACAAG	-40.3	AACAC	-29.9
818	UGGGG	-47.4	CAAUU	-40.6	UCCCA	-29.9
819	ACACA	-47.4	CUUAA	-40.6	GGCCU	-30.2
820	ACAGC	-47.7	CCUCU	-40.7	GACAA	-30.2

821	CCCAC	-47.8	CAGAU	-40.8	UAAGG	-30.3
822	CCAGU	-47.9	AGUAA	-40.8	AACUC	-30.3
823	CUGAG	-48.1	ACUUC	-40.8	GCCAA	-30.7
824	AGGUG	-48.2	CCAAU	-40.9	GGGAC	-30.8
825	GCCCA	-48.2	CAGUC	-41.0	GAGAG	-30.8
826	CAGUC	-48.2	AAUCU	-41.0	AGCAU	-30.8
827	ACCAC	-48.3	AGUAG	-41.0	GGUGA	-30.9
828	CACCA	-48.4	UAGCC	-41.2	AUCCA	-31.0
829	GGACC	-48.4	UACAG	-41.2	AGCUA	-31.0
830	GAGAG	-48.6	GACAA	-41.2	UCACA	-31.2
831	AGUGA	-48.6	AGUCC	-41.4	AUCAG	-31.4
832	GGCAC	-48.6	AUUAA	-41.5	GGAGU	-31.5
833	AGUCC	-49.0	CUCAC	-41.5	CAUGC	-31.6
834	ACAAU	-49.1	AUCCU	-41.7	AGAUA	-31.8
835	AUCAG	-49.1	GGGGC	-41.8	CCUCA	-31.9
836	UUCCA	-49.1	CUUAG	-41.8	CAGGU	-32.2
837	CAAAU	-49.3	CCAGC	-41.8	ACAAC	-32.4
838	CAUCA	-49.5	AGGUA	-41.9	GGUCA	-32.4
839	GAGGU	-49.5	AGUGA	-42.0	CAUAC	-32.6
840	CAGCA	-49.5	AGAUC	-42.1	ACCAC	-32.7
841	UUAAA	-49.6	GAGGU	-42.1	GUGAG	-32.9
842	AAGUC	-49.9	GCAAA	-42.2	GGCAA	-32.9
843	ACAGU	-50.1	CCCUC	-42.2	AUAGC	-33.1
844	UAGGA	-50.3	GGAGU	-42.2	AUAAG	-33.1
845	AGAGU	-50.3	CCAAC	-42.3	CACCA	-33.2
846	AGGUC	-50.4	GAAGC	-42.5	CAAGU	-33.3
847	GACAC	-50.6	AUCAA	-42.5	GCCAC	-33.4
848	ACUCC	-51.0	AGGUC	-42.6	CCAAC	-33.5
849	GAAAC	-51.1	AGAGC	-42.6	GCUCC	-33.7
850	AGCCU	-51.4	CACCA	-42.7	CACAC	-33.8
851	AUUAA	-51.4	CCUCA	-42.9	ACCCA	-33.8
852	GGCCU	-51.5	ACUGG	-43.0	GCAGC	-34.1
853	GGGCA	-51.8	GAGCC	-43.0	CUAAA	-34.3
854	GGACA	-52.1	AGAUA	-43.0	GCAUC	-34.6
855	GCCAC	-52.3	GCAGA	-43.1	CAAUA	-34.6
856	CCAAC	-52.3	ACCCU	-43.1	GUCCC	-34.6
857	CUCCA	-52.3	GACAC	-43.2	GCUCU	-34.8
858	CUCAC	-52.4	GGCUA	-43.3	GCAAA	-34.8
859	GACAA	-52.5	ACUCA	-43.8	AGGAC	-35.0
860	AAUAG	-52.6	UAAGG	-43.9	ACUCA	-35.1

861	AGUCA	-52.7	AAGUC	-44.1	AGUGG	-35.2
862	AAGCU	-52.8	AGCAC	-44.3	CCGGG	-35.3
863	AGAUA	-52.9	AAGAC	-44.5	GCAUU	-35.3
864	UCACA	-52.9	UCCAA	-44.9	CUCCA	-35.4
865	GCCAG	-53.4	ACAAU	-45.1	GGACA	-35.4
866	AGCUG	-53.5	CAGGU	-45.1	GGGUC	-35.4
867	CCUCA	-53.5	AGUUC	-45.1	AAGCU	-35.5
868	GGCUA	-53.6	GGUAG	-45.2	CCCCA	-35.5
869	UACAG	-53.8	GGACU	-45.2	GGCCC	-35.6
870	AAGCA	-53.8	AAGGC	-45.3	UAAAG	-35.6
871	ACCCA	-53.8	CUCCA	-45.3	UGGGA	-35.8
872	CAGCU	-54.0	AGAGA	-45.5	AGCCU	-35.8
873	AACUC	-54.0	GGCAA	-45.5	CUAGG	-35.9
874	AAACU	-54.2	GGACA	-45.5	UCAGC	-36.1
875	GCUCA	-54.2	UUCAA	-45.6	AGCUG	-36.1
876	GGCCC	-54.3	CAACC	-45.8	GCCCA	-36.3
877	AACCC	-54.6	AACCC	-45.8	AGCAC	-36.5
878	GAACC	-54.8	AAGCC	-45.9	AGCCC	-36.6
879	UAAGG	-54.9	UUCCA	-45.9	AACCA	-36.7
880	AGACU	-55.2	ACCAA	-46.0	CUGGA	-36.9
881	AGGAU	-55.3	UUCCC	-46.2	AGAAC	-37.0
882	GACCC	-55.6	AAUUA	-46.2	UCAAG	-37.1
883	CCCCA	-56.1	UAAGA	-46.4	CAUAA	-37.1
884	UAGGG	-56.1	UGGGA	-46.9	ACCAG	-37.1
885	AGCUA	-56.4	GACAG	-47.0	GACAG	-37.7
886	GGAGU	-56.4	UGAGG	-47.2	ACAGC	-37.7
887	CCAGC	-56.5	CUCAA	-47.2	CCUGG	-37.8
888	CAAUA	-56.9	GAACU	-47.9	AACCC	-37.9
889	GGGAC	-57.0	CCAGU	-47.9	GACCA	-38.1
890	AGCCC	-57.3	AGGGU	-47.9	AGGGG	-38.3
891	GGCAA	-57.6	GAGAC	-48.0	UGGAA	-38.4
892	CUCAA	-57.9	GGACC	-48.1	GAACA	-38.5
893	AGAAC	-58.3	CCACA	-48.3	AGAGA	-38.6
894	UCAGC	-58.4	AAGGU	-48.3	CCAGC	-38.7
895	GGCUC	-58.4	GGGAG	-48.5	UCAGA	-38.7
896	CAACA	-58.5	CAGAC	-48.6	AGGUC	-38.8
897	UCAGA	-58.6	GAAAG	-48.6	CAACA	-39.2
898	GGGGC	-58.8	AAAAA	-48.8	AGGUG	-39.2
899	GAGCC	-58.8	GAAGG	-49.1	GAGAC	-39.3
900	CCUGG	-59.0	UUAGG	-49.1	ACACA	-39.4

901	UGGGA	-59.4	AGCCC	-49.1	UCCAA	-39.6
902	AUCAA	-59.7	CUAGG	-49.1	AAAUA	-39.6
903	UGGAA	-60.0	CACAA	-49.1	UACAG	-39.7
904	CAGGU	-60.4	GGGGA	-49.2	GCUCA	-39.8
905	AGAGC	-60.4	AGACC	-49.2	CAGCU	-40.1
906	AGAGA	-60.5	UAGGG	-49.4	AUGGA	-40.1
907	UCCCA	-60.9	GAUGG	-49.5	AGGGC	-40.2
908	UCCAA	-61.6	GGAGC	-49.6	CAAGG	-40.4
909	GGGCU	-61.8	ACUCC	-49.7	AAAGU	-40.4
910	AGGGU	-61.8	CAAAC	-49.9	AAGAU	-40.4
911	GAAGG	-62.0	CCACC	-49.9	UGAGG	-40.5
912	ACUCA	-62.0	AGACU	-50.2	AGAUG	-40.6
913	UCAAG	-62.2	UAGAG	-50.3	AUGGG	-40.6
914	GCAAA	-62.2	UGGAG	-50.5	GCCAG	-40.7
915	AGCAC	-62.3	AACUC	-50.6	AGGGU	-40.8
916	UGAGG	-62.6	GAAAC	-50.6	AAUCA	-40.9
917	AAUCA	-62.9	AAUCC	-50.6	AGAGC	-41.0
918	GAGAA	-63.0	CCUGG	-50.7	GGCUC	-41.1
919	ACAAG	-63.3	AAUAG	-50.7	GAAGG	-41.4
920	GACAG	-63.4	CAACA	-50.8	UGGGG	-41.7
921	AACCA	-63.7	AGUGG	-50.9	GGGGC	-41.7
922	CUAAA	-63.9	GGCCA	-51.1	GGGCA	-42.1
923	AAAUC	-63.9	AGAAC	-51.1	UCAGG	-42.2
924	AGUGG	-64.7	AGGAU	-51.2	GGGCC	-42.2
925	GAGAC	-64.8	UGGAA	-51.5	AAAAU	-42.2
926	CAGAC	-65.6	AGCUC	-51.5	CAAAC	-42.3
927	CUAGG	-65.7	GGGCC	-51.9	AAGGG	-42.5
928	GACCA	-65.9	AGAGU	-51.9	GAGAA	-42.7
929	GGCCA	-66.2	CCCAC	-52.1	GGCAG	-42.9
930	GAAGC	-66.6	AGGCA	-52.2	GAACC	-42.9
931	AAAGC	-66.6	CUAAA	-52.3	CAUCA	-43.2
932	GGGAG	-66.6	AGUCA	-52.5	CAAGC	-44.0
933	AAAAA	-66.7	GGAAC	-52.5	GGAAC	-44.3
934	GGGCC	-66.7	AGGCC	-52.6	GGAGC	-44.5
935	AAGAC	-66.8	AAUCA	-52.8	CAAGA	-44.6
936	AGACC	-67.1	AGGGG	-53.1	AAACC	-44.6
937	GAAGA	-67.4	CACAG	-53.1	UCCAG	-44.6
938	ACAGG	-67.5	UAGGA	-53.2	GGGGA	-44.8
939	GGAGC	-67.9	GAGAA	-53.7	GGGAA	-44.8
940	CCCAA	-68.0	UCAAG	-53.8	AGGGA	-45.1

941	GGAGG	-68.1	CAGCC	-54.2	AGACC	-45.1
942	GCAGA	-68.2	GGCAG	-54.4	ACAAG	-45.7
943	UCAAA	-68.4	AAACC	-54.9	UCAAA	-45.7
944	AGCUC	-68.6	GGGAC	-54.9	GAAAG	-45.9
945	AAACC	-68.7	ACCCA	-54.9	AAGCA	-45.9
946	CAAAC	-68.7	CCAAG	-55.2	CCACA	-46.2
947	CACAA	-69.0	CCAAA	-55.3	AAGGC	-46.3
948	ACCAG	-69.1	GAGGG	-55.3	CACAG	-46.3
949	AAGCC	-69.7	UAGAA	-55.3	GAGGG	-46.3
950	CUGGG	-69.7	UCAGA	-55.4	AAUAA	-46.3
951	CAGCC	-69.8	AGGAC	-55.4	CUGGG	-46.3
952	CAAGG	-69.9	AACCA	-55.4	GAAGA	-46.5
953	AACAG	-70.6	UCCCA	-55.4	CUGCA	-46.6
954	GGGGA	-70.7	AAACU	-55.5	GAGGA	-46.7
955	AUAAA	-71.0	GGGCU	-55.7	AACAA	-46.8
956	CUGGA	-71.4	ACCAG	-55.9	GAGCA	-46.9
957	AAUAA	-71.5	GAAAU	-56.7	GAAGC	-46.9
958	AGGAC	-71.6	GGAGG	-57.0	AAGCC	-47.7
959	GAGGG	-71.6	AAAGU	-57.2	GGCCA	-47.9
960	AAGGC	-72.1	GACCC	-57.2	GGCAU	-47.9
961	AAAGU	-72.4	GAGGA	-57.2	AUCAA	-48.2
962	AGGCU	-72.5	CCCCA	-57.4	UAAAA	-48.3
963	UAAAG	-73.5	CAGGC	-57.4	CCCAA	-48.4
964	AGGGC	-73.6	AUAAA	-57.6	ACAGG	-48.5
965	UCCAG	-73.8	CCCAA	-57.6	GGAAG	-48.8
966	ACCAA	-74.1	CCAGA	-57.6	CACAA	-48.9
967	AAAUA	-74.5	AACAG	-57.9	GCAGA	-48.9
968	GAGGC	-74.8	ACAGG	-58.2	GAAAA	-49.2
969	CCAAG	-75.0	AGACA	-58.3	UGGAG	-49.4
970	GGCAG	-75.1	GAGGC	-58.8	AAGAC	-49.4
971	GGGAA	-75.2	CUGGA	-59.0	AAAGC	-49.5
972	UCAGG	-75.5	UCAGG	-59.3	CAGGG	-49.8
973	CCACA	-75.7	UAAAG	-59.3	AGCAA	-49.9
974	UGGAG	-76.5	GGAAG	-59.4	AGCUC	-50.2
975	AGGGA	-76.5	GGGAA	-59.5	AACAG	-50.3
976	GGAAG	-77.0	GCAGG	-59.6	CAGGC	-50.3
977	AAGAU	-78.6	AAUAA	-59.6	AGCCA	-50.5
978	AGCAA	-79.0	GGAGA	-60.7	ACCAA	-50.5
979	AAACA	-79.5	UUAAA	-60.8	AAAAA	-50.6
980	GGAAC	-79.5	AAAAC	-61.2	GGGAG	-50.6

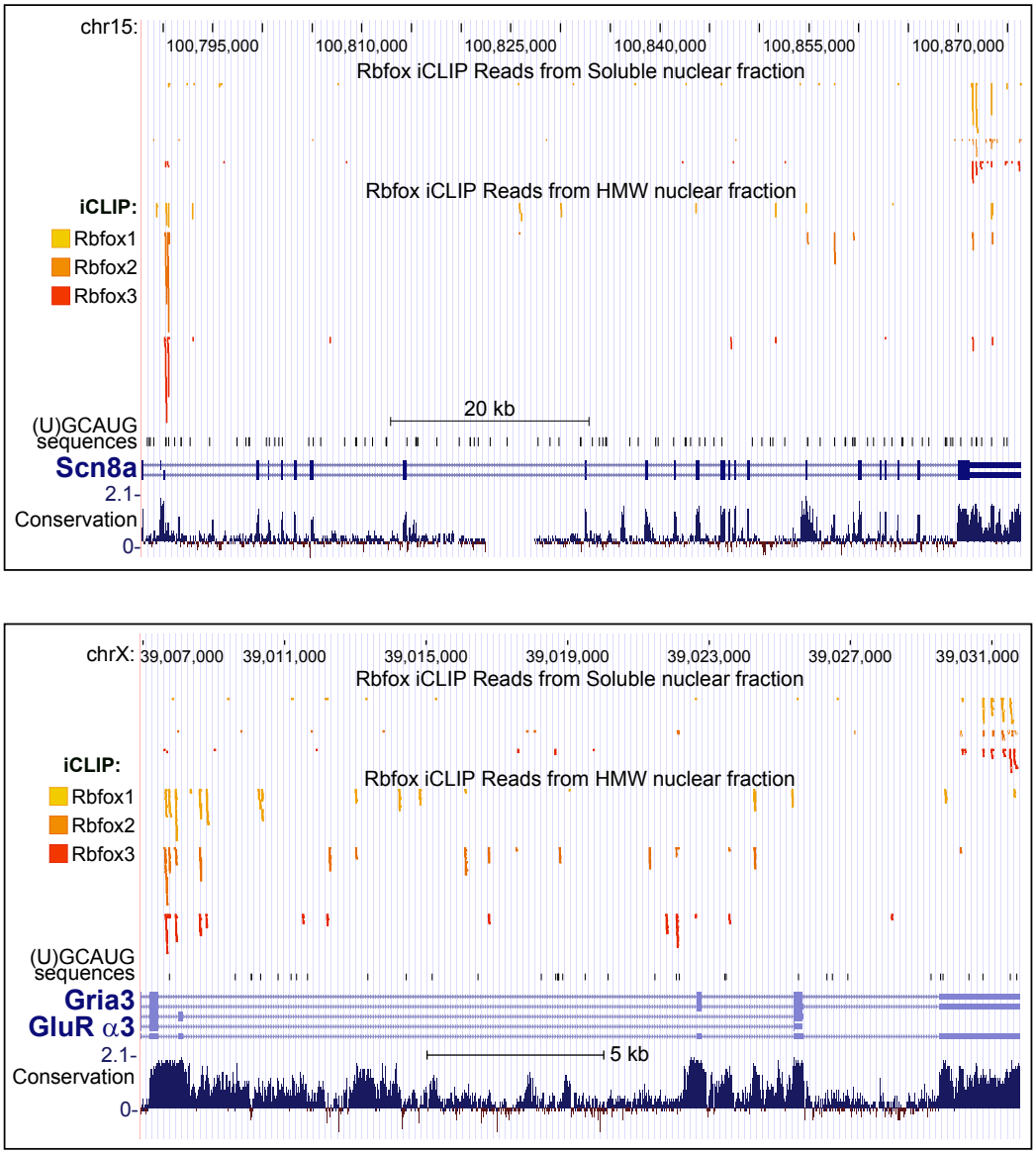
981	AGAAG	-80.2	AACAA	-61.2	GAGCC	-50.8
982	AGGCC	-80.2	AGAAG	-61.7	CAUGG	-50.8
983	UAAAA	-80.4	AGCAG	-61.9	GGAGG	-50.9
984	AAAAU	-80.6	CCCAG	-62.0	CAGCA	-51.3
985	CACAG	-80.9	AAAUC	-62.0	AAGAA	-51.4
986	AAGAA	-81.1	CUCAG	-62.3	AGACA	-51.6
987	AAGGG	-81.2	AGGGA	-62.3	AAGGA	-51.6
988	AACAA	-81.4	AGGCU	-63.0	AUGCA	-52.0
989	GGAGA	-81.4	CAGAG	-63.6	GAGGC	-52.3
990	AGGAG	-81.6	AGGGC	-63.8	AUAAA	-52.3
991	GAAAG	-81.6	AAGGG	-64.6	CUCAG	-52.5
992	CAGGC	-83.5	CAAAG	-64.6	AAGAG	-52.6
993	GCAGG	-84.3	CUGGG	-64.8	GCAGG	-53.0
994	CAAAG	-84.9	CAAAA	-64.9	AGGAG	-53.1
995	CCAAA	-85.3	GAAAA	-65.2	GGAGA	-53.6
996	AAAAC	-85.4	AAAUA	-65.3	CCAAA	-53.9
997	AGGGG	-85.8	GAAGA	-66.0	CAGAC	-54.1
998	CAAGA	-86.3	CCAGG	-66.5	CAAAA	-54.7
999	AGCAG	-86.4	AAGAU	-66.6	CCAAG	-54.7
1000	CCCAG	-86.4	AGCAA	-67.9	AAACA	-55.2
1001	GAGGA	-87.7	AAGAA	-68.0	GCAUA	-55.7
1002	AGGCA	-88.5	UCAAA	-68.5	AAAAC	-56.1
1003	GGAAA	-88.6	ACAAA	-68.6	ACAAA	-56.2
1004	ACAAA	-88.9	CAAGG	-68.7	GGAAA	-56.4
1005	CCAGA	-89.4	AGGAA	-68.9	AGAAA	-56.8
1006	AGACA	-89.7	UCCAG	-69.3	AGGCC	-56.9
1007	AAGGA	-89.9	AGCCA	-69.7	CCAGA	-57.8
1008	AGCCA	-90.0	AAAGA	-70.0	AGAAG	-58.7
1009	CAGGG	-90.5	AAAAG	-70.1	ACAGA	-59.1
1010	ACAGA	-90.7	AAACA	-70.5	CCCAG	-60.1
1011	CAAAA	-91.0	ACAGA	-70.6	CAGGA	-60.2
1012	GAAAA	-91.4	CAGGA	-70.6	AGGAA	-61.1
1013	AAGAG	-94.0	AGGAG	-71.1	AAAGG	-63.6
1014	CAGAA	-94.4	CAGAA	-73.6	AAAGA	-63.6
1015	AAAGG	-95.5	CAAGA	-74.1	CAGAG	-64.6
1016	AGGAA	-95.6	AGAAA	-74.2	AGAGG	-66.1
1017	CUCAG	-96.4	AGAGG	-74.3	CCAGG	-67.0
1018	CAGAG	-99.2	GGAAA	-74.5	AGCAG	-68.0
1019	AAAAG	-99.7	CAGGG	-76.3	AAAAG	-68.3
1020	AGAGG	-99.8	AAAAU	-76.6	CAAAG	-69.0

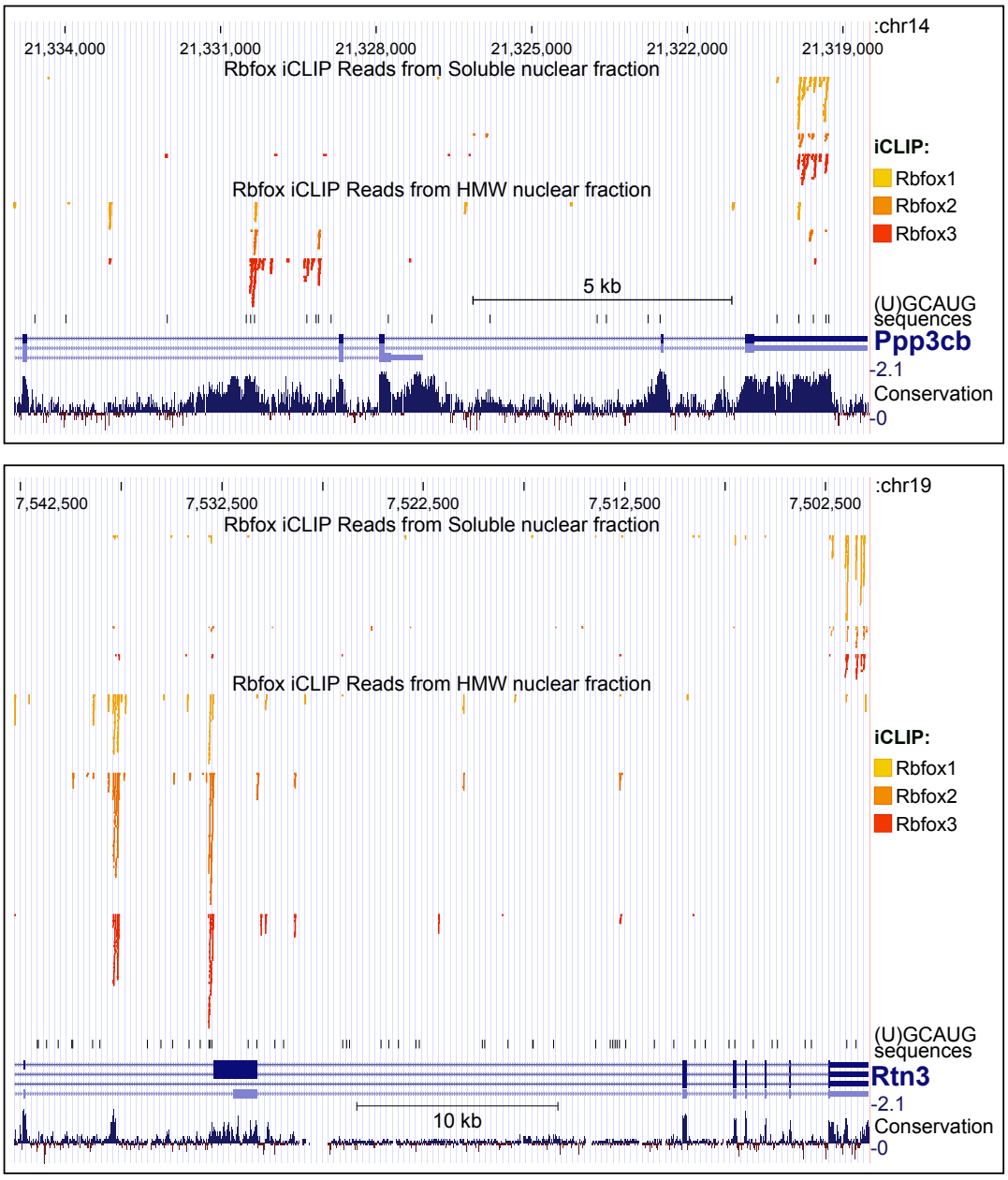
1021	AGAAA	-102.8	AAGAG	-78.1	AGGCA	-69.6
1022	AAAGA	-103.8	AAGGA	-80.8	CAGAA	-72.0
1023	CCAGG	-103.8	UAAAA	-82.5	GCAUG	-101.5
1024	CAGGA	-109.4	AAAGG	-93.9	UGCAU	-108.8

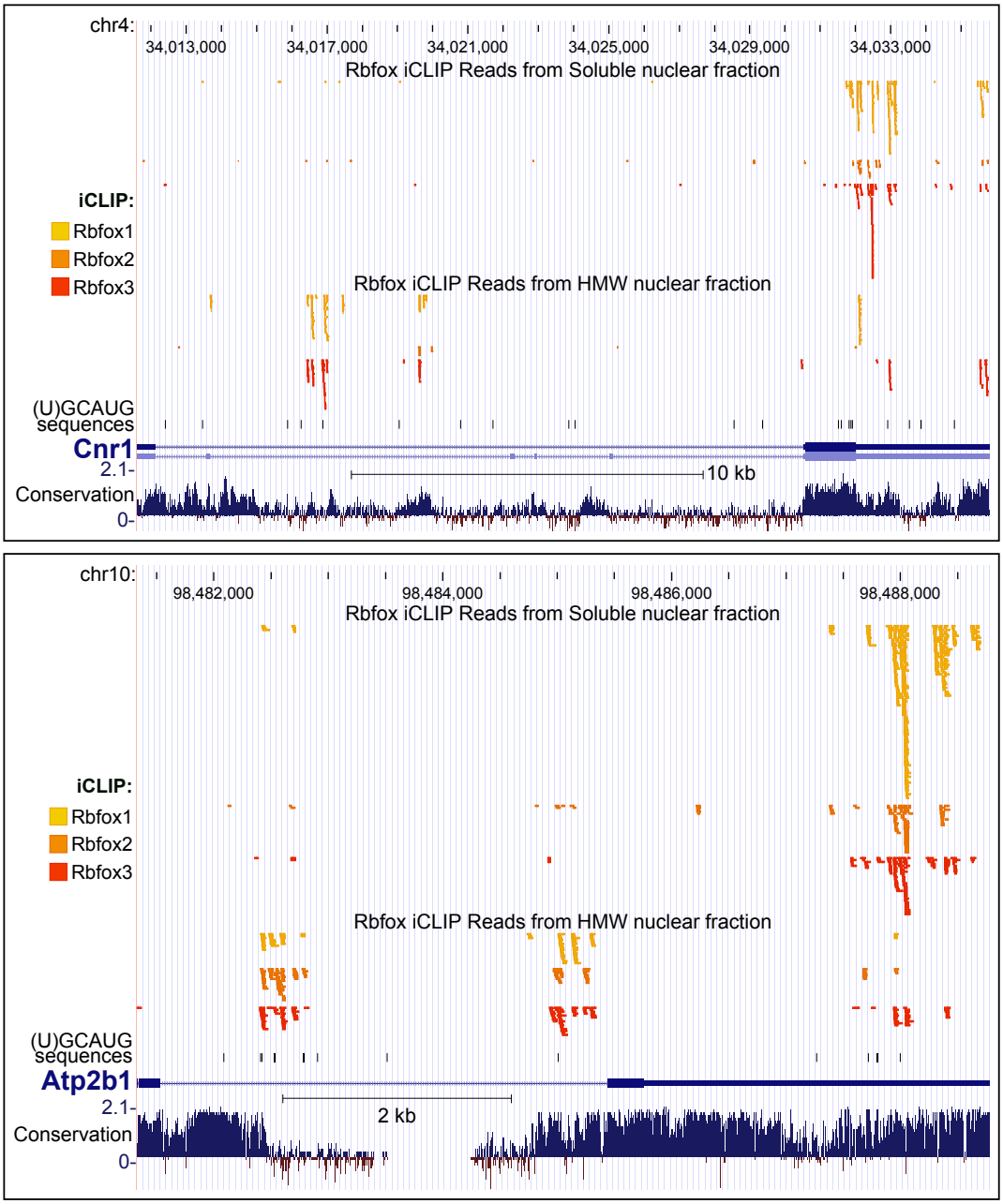
Data S1. UCSC Genome Browser Views of the

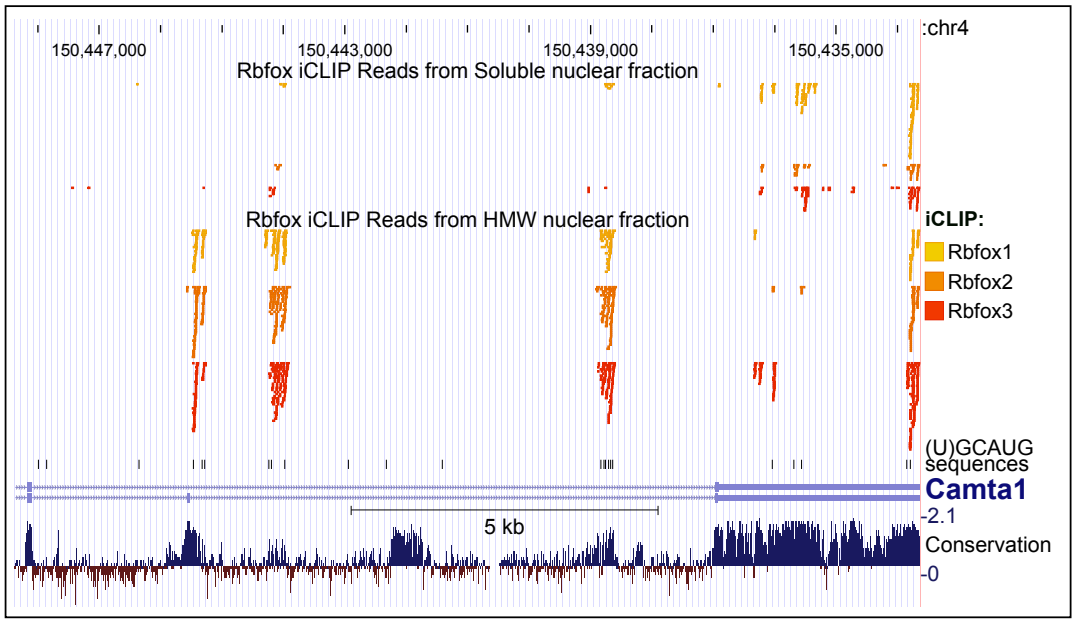
Scn8a, Gria3, Ppp3cb, Rtn3, Cnr1, Atp2b1, Camta1, and Aplp2 Genes, Related to

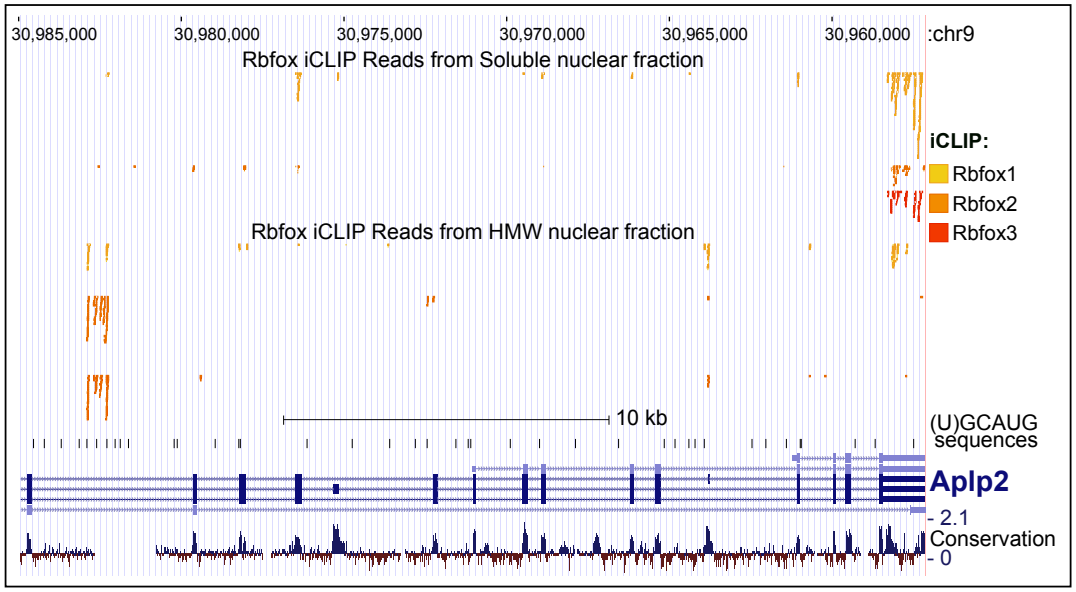
Figure 2.




:chr14
21,334,000
21,331,000
21,328,000
21,325,000
21,322,000
21,319,000
Rbfox iCLIP Reads from Soluble nuclear fraction
Rbfox iCLIP Reads from HMW nuclear fraction
5 kb
iCLIP:
Rbfox1
Rbfox2
Rbfox3
(U)GCAUG sequences
Ppp3cb
-2.1
Conservation
-0
:chr19
7,542,500
7,532,500
7,522,500
7,512,500
7,502,500
Rbfox iCLIP Reads from Soluble nuclear fraction
Rbfox iCLIP Reads from HMW nuclear fraction
10 kb
iCLIP:
Rbfox1
Rbfox2
Rbfox3
(U)GCAUG sequences
Rtn3
-2.1
Conservation
-0


chr4:
34,013,000
34,017,000
34,021,000
34,025,000
34,029,000
34,033,000
Rbfox iCLIP Reads from Soluble nuclear fraction
iCLIP:
Rbfox1
Rbfox2
Rbfox3
Rbfox iCLIP Reads from HMW nuclear fraction
(U)GCAUG sequences
Cnr1
2.1-
10 kb
Conservation
0-
chr10:
98,482,000
98,484,000
98,486,000
98,488,000
Rbfox iCLIP Reads from Soluble nuclear fraction
iCLIP:
Rbfox1
Rbfox2
Rbfox3
Rbfox iCLIP Reads from HMW nuclear fraction
(U)GCAUG sequences
Atp2b1
2.1-
2 kb
Conservation
0-


150,447,000
150,443,000
150,439,000
150,435,000
:chr4
Rbfox iCLIP Reads from Soluble nuclear fraction
Rbfox iCLIP Reads from HMW nuclear fraction
iCLIP:
Rbfox1
Rbfox2
Rbfox3
(U)GCAUG sequences
Camta1
5 kb
-2.1
Conservation
-0


30,985,000
30,980,000
30,975,000
30,970,000
30,965,000
30,960,000
:chr9
Rbfox iCLIP Reads from Soluble nuclear fraction
Rbfox iCLIP Reads from HMW nuclear fraction
iCLIP:
Rbfox1
Rbfox2
Rbfox3
10 kb
(U)GCAUG sequences
Aplp2
- 2.1
Conservation
- 0

CHAPTER 3

INTRODUCTION TO NON-RNA-BINDING LOW COMPLEXITY AND DISORDERED REGIONS IN RNA-BINDING PROTEINS

RNA-binding proteins (RBPs) control RNA metabolism homeostasis from biogenesis to degradation in the cell. RBPs contain one or more RNA-binding domains (RBDs) that can recognize RNA in a sequence-specific manner. Multiple RBDs extend the surface, often β-sheet, contacting the RNA and hence confer higher binding affinities and specificities. RBDs consist of several different structural classes, including RNA recognition motif (RRM), hnRNP K-homology (KH), double-stranded RBD (dsRBD), Zinc finger domains and others. RRM is by far the most common and best-studied class of RBDs, and it is known to bind single-stranded RNA. The typical RRM is composed of about 90 amino acids that form a four-stranded anti-parallel β-sheet with two helices packed on each side, giving the domain βαββαβ topology. The dsRBD is a smaller proteins domain of 65-70 amino acids adopting an αβββα structure that binds to double-stranded RNA. Zinc fingers are generally considered as DNA-binding domains, however, several classes of zinc finger proteins, including the common C2H2 zinc fingers, exhibit RNA-binding activity. Structure determination of RNA:RBD interactions by co-crystallography and nuclear magnetic resonance (NMR) provides detailed information on residues in RBDs that are critical in interacting with RNA bases. Furthermore, functional characterization of RBDs has been benefited tremendously from the

development of technologies in identifying their binding targets and binding sequences. For example, CLIP-seq enabled us to discover numerous binding targets of RBPs and their binding sites globally *in vivo* (Ule et al., 2003).

Besides the RNA-binding domains, RBPs also contain other domains that are less well characterized. These auxiliary domains are often composed of intrinsically disordered regions (IDRs) or of low amino acid complexity (LC), and both are highly enriched in the RBPs compared to non-RBPs (Calabretta and Richard, 2015; Kato et al., 2012). Although disordered regions are historically viewed as hubs for protein-protein interaction networks (Haynes et al., 2006), the interacting partners and the nature of protein-protein interactions remain elusive for majority of the RBPs.

The assemblies of RBPs and RNA can form RNP granules in the nucleus and the cytoplasm, which are microscopically visible membrane-less bodies. Recent observations of germ line P granules in *C. elegans* suggested a mechanism for structuring membrane-less bodies depending on the ability of proteins to transition between a soluble form and a condensed phase (Brangwynne et al., 2009). When the concentration of P granule components is high, these components are condensed to microscopically visible bodies. Conversely, lowering the concentration will cause the dissolution of P granules and the redistribution of protein throughout the cells. Together with the observations that P granules flowed off nuclei, dripped, and often fused into one larger drop, P granules were proposed to behave like liquid droplets, which diffuse, dissolve and condense during the cycles of assembly and disassembly of P granules

(Brangwynne et al., 2009). Multivalent interactions by repeated interaction domains, often seen in RNA-binding proteins and RNA, could yield sharp liquid-liquid-demixing phase separations, suggesting that multivalency may be a ubiquitous driving force for liquid droplet condensation (Li et al., 2012b).

Work from McKnight's group showed that many RNA-binding proteins precipitated by biotinylated isoxazole (b-isox) were components of RNP granules (Han et al., 2012; Kato et al., 2012). Many of them contain LC sequences. These LC sequences can undergo concentration-dependent phase transition *in vitro* into a hydrogel state, which consists of uniformly polymerized amyloid-like fibers (Kato et al., 2012). These observations suggest that LC sequences of RBPs are capable of driving RBPs compartmentalization into non-membrane bound structures. Among these LC sequences, repetitive tripeptide [G/S] Y [G/S] motif is highly enriched. Mutations of tyrosine to serine impair the ability to be retained in a preformed hydrogel (Han et al., 2012; Kato et al., 2012). Dysregulation of assembly and disassembly of LC sequence domains might result in pathogenic fibrillar inclusions of RBPs like FUS and TDP-43, seen in neurodegenerative disorders (Maekawa et al., 2009; Vance et al., 2009).

A systematic survey of multiple RBPs containing IDRs or LC sequences suggested that disordered regions of RBPs have a propensity to form fibers or phase-separated liquid droplets *in vitro*, and this process could be triggered by low temperature, high protein concentration, low salt concentration or RNA-binding (Lin et al., 2015). Over time these structures mature to more stable higher ultra-structures, which often coincides with

formation of fibrous structures or hydrogel droplets wherein an amyloid-like cross-beta structure was observed (Kato et al., 2012; Lin et al., 2015). Different disordered domains could co-assemble into phase-separated droplets, suggesting coexistence of homotypic and heterotypic interactions.

RNA strongly promotes assemblies of RBPs. RNA was trapped in the precipitates by b-isox via bounding by RBPs in the assemblies (Han et al., 2012). Moreover, RNA was shown to trigger higher-order assemblies of FUS (Schwartz et al., 2013) and promote liquid droplet formation (Lin et al., 2015). Post-translational modification of RBPs could also be utilized to regulate the propensity of higher-order assembly. Phosphorylation of the MEG proteins promotes granule disassembly and dephosphorylation of MEG promotes granule assembly (Wang et al., 2014). Phosphorylation of LC sequences in FUS impedes retention to a pre-formed hydrogel *in vitro* (Han et al., 2012), suggesting that phosphorylation disassembles FUS assembly.

However, the role of IDRs and LC sequences in RBPs is poorly understood and the relationship between propensity of aggregation and their physiological function remains unresolved. Several studies reported that LC sequences allow localizing RBPs in and out of organized, membrane-free subcellular organelles like stress granules or RNA granules in the cytoplasm (Jain et al., 2016; Kato et al., 2012). A typical component of stress granules hnRNP A1, form liquid droplets, relying on its LC sequences. The phase separation promotes the fibrillization of disease-causing mutant hnRNPA1-D262V, probably by increasing local molecular concentration and hence nucleation (Molliex et al.,

2015). Disease-causing mutations in LC sequences of FUS induce further phase transition into poorly soluble fibrillar hydrogels that trap other RNP granule components, disrupt RNP granule function and cause neurodegeneration (Murakami et al., 2015). Besides RNP granules, stress granules enriched in RBPs and RNA have been proposed to be dynamic, liquid-like structures. IDRs and LC sequences have been shown to contribute to the formation of stress granules.

In addition to RNP granule and stress granule regulation, LC sequences have been shown to play roles in other biological processes. In yeast, the RNA-binding protein Rim4 forms amyloid-like aggregates and mediates translational repression of numerous mRNA transcripts by binding to their 3' UTR. Blocking the aggregation of Rim4 releases RNA transcripts for active translation (Berchowitz et al., 2015). When fused to a DNA-binding domain, LC sequences of the FET (FUS/EWS/TAF15) could interact with the carboxyl terminal domain (CTD) of RNA polymerase II through polymerization to activate RNA transcription (Kwon et al., 2013). Phosphorylation of CTD regulates its ability to bind to the hydrogel of FET protein *in vitro*, suggesting a new mechanism of how phosphorylation and dephosphorylation of CTD modulate the recruitment of RNA polymerase II to transcription start site. SR repeats of SRSF2 do not form hydrogels by themselves, but they bind to hydrogel droplets formed by the LC domains of hnRNPA2. SRSF2 with SR-to-GR mutations becomes trapped in the nucleoli at an early stage of nuclear speckles formation and impedes pre-mRNA splicing (Kwon et al., 2014). Understanding the physiochemical rules governing RNP granule assembly and

properties promises to shed light on the important but still poorly understood aspects of gene regulation in the cell.

In the course of investigating the assembly of Rbfox with other LASR components, unexpectedly, we found that the LC sequence at the C-terminal part of Rbfox serves as an interface between Rbfox and LASR and mediates the higher-order assembly of Rbfox. Interestingly, as observed in many other RBPs like FUS and hnRNPA1, purified LC sequence of Rbfox is prone to aggregate and form hydrogel, thus proposing an intriguing question whether the assembly of Rbfox and LASR utilizes the aggregation property of LC sequence of Rbfox. Indeed, we found that tyrosine repeats similar to tripeptide motif [G/S] Y [G/S] are nearly evenly distributed across this domain and they are required for high-order assembly of Rbfox. We also found that the higher-order assembly is required for Rbfox-dependent splicing activation, but dispensable for splicing repression, which provides a clue for understanding a splicing factor as a dual actor. Our findings on how the LC sequences of Rbfox promote higher-order assembly with LASR to regulate alternative splicing are described in Chapter 4.

References

Berchowitz, L.E., Kabachinski, G., Walker, M.R., Carlile, T.M., Gilbert, W.V., Schwartz, T.U., and Amon, A. (2015). Regulated Formation of an Amyloid-like Translational Repressor Governs Gametogenesis. Cell.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. Science *324*, 1729-1732.

Calabretta, S., and Richard, S. (2015). Emerging Roles of Disordered Sequences in RNA-Binding Proteins. Trends Biochem Sci *40*, 662-672.

Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G.*, et al.* (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell *149*, 768-779.

Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., Uversky, V.N., Vidal, M., and Iakoucheva, L.M. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. Plos Comput Biol *2*, 890-901.

Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., and Parker, R. (2016). ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. Cell *164*, 487-498.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J.*, et al.* (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753-767.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell *155*, 1049-1060.

Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S.L. (2014). Poly-dipeptides encoded by the C9ORF72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science.

Li, P.L., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F.*, et al.* (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336-U129.??

Lin, Y., Protter, D.S., Rosen, M.K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Molecular cell *60*, 208-219.

Maekawa, S., Leigh, P.N., King, A., Jones, E., Steele, J.C., Bodi, I., Shaw, C.E., Hortobagyi, T., and Al-Sarraj, S. (2009). TDP-43 is consistently co-localized with ubiquitinated inclusions in sporadic and Guam amyotrophic lateral sclerosis but not in familial amyotrophic lateral sclerosis with and without SOD1 mutations. Neuropathology *29*, 672-683.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. Cell *163*, 123-133.

Murakami, T., Qamar, S., Lin, J.Q., Schierle, G.S., Rees, E., Miyashita, A., Costa, A.R., Dodd, R.B., Chan, F.T., Michel, C.H.*, et al.* (2015). ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. Neuron *88*, 678-690.

Patel, S.S., Belmont, B.J., Sante, J.M., and Rexach, M.F. (2007). Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. Cell *129*, 83-96.

Schwartz, J.C., Wang, X., Podell, E.R., and Cech, T.R. (2013). RNA seeds higher-order assembly of FUS protein. Cell reports *5*, 918-925.

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P.*, et al.* (2009). Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. Science *323*, 1208-1211.

Wang, J.T., Smith, J., Chen, B.C., Schmidt, H., Rasoloson, D., Paix, A., Lambrus, B.G., Calidas, D., Betzig, E., and Seydoux, G. (2014). Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically-disordered proteins in C-elegans. eLife *3*.

CHAPTER 4

LOW-COMPLEXITY SEQUENCE DOMAINS OF RBFOX FORM HIGHER-ORDER COMPLEXES WITH LASR TO REGULATE ALTERNATIVE SPLICING

Introduction

RNA-binding proteins (RBPs) control all aspects of RNA metabolism from biogenesis to decay. An RBP may have RNA-binding domains (RBDs) of several structural classes each recognizing short RNA element. Modern sequencing methods now allow relatively facile identification of large numbers of binding sites for these proteins across the transcriptome. However, these proteins frequently have auxiliary non-RNA binding domains, whose protein-protein interactions and functions are mostly not known. These auxiliary domains are often predicted to contain intrinsically disordered regions (IDR) and sequences of low amino acid complexity (LC) (Kato et al., 2012; Neelamraju et al., 2015).

Recent studies have found that certain IDR and LC domains have a propensity to form fibers or phase-separated liquid droplets *in vitro* (Elbaum-Garfinkle et al., 2015; Lin et al., 2015; Molliex et al., 2015; Nott et al., 2015). Over time these structures mature into highly stable assemblies containing an amyloid-like cross-beta structure within a hydrogel matrix (Kato et al., 2012; Murakami et al., 2015; Patel et al., 2015). These aggregation properties and the formation of amyloid-like fibrils have garnered great interest from the recognition that in ALS and other neurological pathologies RNA-binding proteins such as

FUS and TDP-43 form very stable cellular inclusions containing amyloid-like structures. Recent work has also found that IDR or LC domains and their aggregation may function in normal mRNA metabolism to allow the reversible localization of RBPs in and out of membrane-free subcellular organelles such as cytoplasmic stress granules or RNP granules (Jain et al., 2016; Kato et al., 2012). The nuclear RBPs involved in splicing and other processes also have extensive LC sequences, but the roles for these sequences and their aggregation properties in normal function are not yet defined.

The regulation of alternative pre-mRNA splicing involves a very large number of RBPs that bind nascent transcripts to alter spliceosome assembly and splice site choice. One important family of splicing regulators is the Rbfox proteins that control networks of spliced isoform expression in brain, muscle and during early embryonic development (Nakahata and Kawamoto, 2005; Underwood et al., 2005; Yeo et al., 2009b). The Rbfox proteins are of particular interest from their associations with neurological diseases. Rbfox1 in particular was found to be mutated in rare patients with autism spectrum disorders and epilepsy, and changes in Rbfox1 expression and Rbfox-dependent splicing have been observed in brains of ASD patients. There are three mammalian Rbfox genes (Rbfox1, Rbfox2, and Rbfox3) each containing a single highly conserved RBD of the RNA recognition motif type (RRM) that binds the short RNA element GCAUG or often UGCAUG (Auweter et al., 2006; Jin et al., 2003). Alternative promoters and alternative splicing diversifies the protein structures derived from each gene, including generating both nuclear and cytoplasmic isoforms (Damianov and Black, 2010; Lee et al., 2009).

The nuclear Rbfox proteins promote alternative exon inclusion when bound to a downstream UGCAUG element, or skipping of an exon when bound upstream or within the exon (Jin et al., 2003). The N and C terminal domains have segments of low amino acid complexity and presumably engage in protein-protein interactions (Figure 4.S 4.1A). The C terminal domain was found to be required for splicing regulation by an MS2-tethered protein (Sun et al., 2012). However, besides determining the subcellular localization, the extended N and C terminal domains flanking the RRM are of unknown function.

Recently we showed that Rbfox proteins regulate splicing in association with the Large Assembly of Splicing Regulators, LASR, a multi-protein complex of RNA-binding proteins (Damianov et al., 2016). The interaction of Rbfox with LASR components was observed to affect their binding to nascent RNA and their activity in splicing. LASR contains eight proteins hnRNPs C, H, M, and UL, NF110, NF45, Matrin3, and DDX5, all approximately equimolar with Rbfox and each other. Virtually all of the intron-bound Rbfox was associated with LASR and could be extracted from the high-molecular-weight (HMW) material of cell nuclei by nuclease digestion. Despite extensive nuclease digestion, the Rbfox/LASR complex was found to sediment on density gradients at 55S and was thus forming complexes of a higher-order than a single Rbfox bound to a single LASR. The nature of the Rbfox/LASR contacts or the interactions leading to their higher-order assembly remained unresolved.

Here we report that the Rbfox C-terminal domain mediates its interaction with LASR, and further that an LC sequence within this domain mediates higher-order assembly of Rbfox/LASR complexes. Examining the solution properties of this purified domain, we find that the LC region causes the protein to form soluble aggregates and to form fibrous structures and hydrogels. Mutations that specifically block the higher-order assembly of Rbfox but not its interactions with LASR also block Rbfox-dependent splicing activation and establish a link between the biophysical properties of Rbfox aggregation and its function in splicing regulation.

Results

The C-terminal domain of Rbfox mediates interaction with LASR and formation of higher-order assembly.

To identify regions of Rbfox1 responsible for interacting with LASR, we generated Flp-In™ T-REx™ 293 cell lines stably expressing deletion mutants of Rbfox1 tagged with the HA-FLAG epitopes and the SV40 NLS (Figure 4.1A). We isolated Rbfox/LASR complexes from these cells by FLAG immunoprecipitation. The complete set of LASR proteins was co-immunoprecipitated with Rbfox missing either the N-terminal domain (NT) or the RNA-binding domain (RBD), indicating that these two regions are not essential for the LASR interaction (Figure 4.1B). Some LASR subunits were isolated in lower amounts in these samples, suggesting that NT and RBD, while not essential, may provide contacts

for particular proteins. In contrast, deletion of C-terminal domain (ΔCT) abolished the interaction with LASR, and the CT fragment alone co-immunoprecipitated the LASR subunits with similar efficiency to full-length Rbfox1 (Figure 4.1A). Thus the CT provides the primary contact with LASR. We previously found that Rbfox1 sediments on glycerol gradients as a larger assembly than predicted for a single Rbfox/LASR complex. Applying the glycerol density gradient assays to the mutant proteins, we found that the CT fragment sedimented as a higher-order complex similar to the full-length protein, while ΔCT remained at the top of the glycerol gradient as expected (Figure 4.1C). These results demonstrate that the Rbfox CT domain is necessary and sufficient both to interact with LASR and to form higher-order complexes.

Repetitive tyrosine residues within the CT region are essential for higher-order assembly of Rbfox1

We previously found that all the brain-expressed variants of the Rbfox proteins including Rbfox1, Rbfox2, and Rbfox3 form higher-order complexes seen in gradients. These were observed both endogenously in mouse brain and with ectopically expressed proteins in HEK293 cells (Damianov et al., 2016). Both Rbfox1 and Rbfox2 have muscle-specific variants derived from the inclusion of muscle-specific exon M43 instead of exon B40 (Nakahata and Kawamoto, 2005). Exon B40 was first described in neuronal cells although it was included in other non-neuronal cells as well. Exons M43 and B40 encode related but not identical amino acid sequences within the CT domain (Figure

4.1D). We examined whether the M43 Rbfox variants also form higher-order complexes. Comparing the glycerol gradient sedimentation of these variants, we found that Rbfox1_M43 exhibited a similar sedimentation profile to Rbfox1_B40 and Rbfox2_B40, with a peak in the 55S region. Strikingly, Rbfox2_M43 did not form higher-order complexes, but instead sedimented as a much smaller species near the top of the gradient (Figure 4.1D). Rbfox2 exon M43 could also modify the behavior of Rbfox1. A hybrid protein (Rbfox1_2M43) also failed to form higher-order complexes (Figure 4.1D). These data suggested that Rbfox2_M43 might be missing residues needed for the higher-order assembly. Aligning the exon encoded amino acid sequences, it was notable that of three tyrosine residues present in both B40 and M43 exons, two were conserved in Rbfox1_M43 but all were missing from Rbfox2_M43.

Looking more broadly in the CT region, we found additional tyrosine residues upstream and particularly downstream of exon B40, with the downstream residues more closely spaced and conserved across the Rbfox paralogs (Figure 4.1E). To examine the role of these tyrosines in the higher-order assembly, we created a series of mutant Rbfox1 proteins with increasing numbers of tyrosines changed to serine or alanine. A mutant Rbfox1 missing three tyrosine residues sedimented partially at 55S, but a substantial fraction of the protein shifted to the top of the gradient. For mutants with additional tyrosines changed to serine (6, 7, or 10 residues), higher-order assembly was nearly eliminated (Figure 4.1F). Mutations to serine showed a slightly stronger effect than alanine mutations. In contrast, changing three tyrosines to phenylalanine residues did not

impair the higher-order assembly of Rbfox1 (Figure 4.1F), suggesting that aromatic interactions mediate the assembly.

Given that the CT domain was required for the Rbfox interaction with LASR, it was possible that the higher-order assembly involved interactions of LASR proteins and that the effect of the tyrosine mutations reflected a loss of the Rbfox1 and LASR interaction. This proved to not be the case. All the tyrosine-to-serine mutants as well as Rbfox2_M43 retain their interaction with LASR, pulling down LASR in a FLAG immunoprecipitation assay with equal efficiency to wild type proteins (Figure 4.1G and data not shown). These data indicated that the property of higher-order assembly by Rbfox proteins could be separated from the LASR interaction.

Multiple interfaces between the CT region of Rbfox1 and LASR

To further define the interactions between Rbfox and LASR, we made additional mutations. We divided the CT domain into three fragments, C1, C2, and C3 (Figure S4.1A). C1 included the sequence upstream of the ten mutated tyrosines. C2 contained the mutated tyrosines. C3 included the sequence downstream of the mutated tyrosines. Each fragment was fused to HA-FLAG-SV40NLS-tagged Rbfox1 ΔCT which failed to interact with LASR (Figure 4.1B). Adding C1 to Rbfox1 ΔCT had little effect, with the protein pulling down only small amounts of LASR. This protein also did not form higher-order complexes (Figure 4.2B). In contrast, the proteins containing either C2 or C3 pulled down LASR relatively efficiently (Figure 4.2A). C2 was more prone to aggregation

into higher-order complexes than C3 (Figure 4.2B). The medium-sized complexes in fraction 7 to 9 formed by C3 may result from tyrosines not tested by mutation. Mutation of the tyrosines in C2 eliminated the higher order assembly (Figure 4.2B), and unlike the full length protein also reduced the interaction with LASR (Figure 4.2A).

These data indicate that C2 and C3 can independently interact with LASR. Unlike the full length CT domain, the interaction of the C2 fragment was dependent on the tyrosines.

Aggregation of the CT domain alone

A variety of self-aggregation properties have been reported for the LC domains of RNA-binding proteins, including the formation of hydrogels, amyloid-like fibrils, and phase-separated liquid droplets (Kato et al., 2012; Lin et al., 2015). The LC domain of the FUS protein contains the repeated tripeptide [G/S] Y [G/S] that mediates its assembly into an amyloid-like cross-beta structure (Kato et al., 2012). Although not exact matches to the FUS tripeptide motif, many of the tyrosines in the Rbfox1 C2 region are preceded or followed by glycine or serine. Hot spots of aggregation in polypeptides of the CT domain were identified by AGGRESCAN (Conchillo-Sole et al., 2007) (Figure S4.1A). We examined whether the Rbfox CT domain exhibited similar aggregation properties to FUS and other RBPs. We purified bacterial recombinant 6xHis-tagged wildtype CT domain and the CT mutant with ten tyrosines changed to serines (CT-YS), each fused to mCherry. Wild type mCherry-CT fusion protein eluted in the void volume of a size-exclusion column as large soluble nucleic-acid-free aggregates. In contrast, a portion of the CT-YS mutant

eluted as monomeric protein, indicating that it is less prone to aggregation (Figure S4.1B).

We used a fluorescent liquid-liquid droplet assay to examine aggregation of the wildtype and mutant proteins. CT or CT-YS was fused to the SNAP tag which was fluorescently labeled with SNAP-surface 649. Protein solutions containing 2% of fluorescently labeled proteins were used in droplet assembly assays as described (Lin et al., 2015). Changes in phase were monitored by fluorescence microscopy. The labeled solutions remained monophasic and clear at room temperature in 150 mM NaCl. However, when the NaCl concentration was dropped to 37.5 mM, the CT protein at 12 μM was seen to form fluorescent fibrous structures by epifluorescent microscopy after 21 hours. Less concentrated solutions of the CT protein (6 μM) also formed fibrous structures with longer incubation (48 hours). In contrast, the CT-YS mutant remained in solution for days under same conditions (Figure 4.3A). The CT formed more fibrous structures than the round liquid droplets reported with the FUS protein. Fluorescent aggregates were only observed by reducing the salt concentration from the 150 mM used in purification, although it is possible that smaller aggregates of protein eluted in the void from the size exclusion column may seed the larger visible structures forming in low salt.

To test this possibility, we carried out similar experiments using the monomeric fractions of the C2 and its ten-tyrosine-to-serine mutant C2-YS (Figure S4.1B). Performing the same droplet assembly assay, we observed that C2 but not C2-YS formed fibrous aggregates in low salt (Figure 4.3B).

We also tested the crowding reagent polyethylene glycol (PEG) to promote phase separation of proteins at low concentration. In the presence of 10% PEG, SNAP-C2 labeled with SNAP-surface 549 formed spherical fluorescent structures of 0.5-1 μM in 150 mM NaCl (Figure S4.1D). An equivalent solution of SNAP-C2-YS remained clear (Figure S4.1D). The SNAP-surface 549 labeled C2 or C2YS were incubated with C2 or C2-YS fused to monomeric EGFP (mEGFP). mEGFP-C2 but not mEGFP-C2-YS was found to label the SNAP-surface 549 labeled C2 droplet as well as form droplets on its own (Figure 4.3C), indicating that homotypic interactions of the C2 region require the tyrosines.

In other assays we compared C2 and C2YS for their ability to form hydrogels and fibers. After incubation at room temperature for 5 days, the concentrated mEGFP-C2 (1 mM) formed hydrogel droplets, whereas the mEGFP-C2-YS did not (Movie S1). The large fibrous structures formed by the C2 protein were observed by both fluorescent microscopy and transmission electron microscopy (Figure S4.1EF). These structures were stained with Thioflavin T, possibly indicated an amyloid-like assembly (Figure S4.1E). The assembly of pure Rbfox protein or the CT domain required the tyrosine residues. This homophilic assembly may serve to concentrate Rbfox/LASR in structures larger than the unit complex of one Rbfox and one LASR complex.

The Rbfox Tyrosine-rich domain is required for GCAUG-dependent exon activation

In earlier work, we showed that nearly all intron bound Rbfox protein was in complex with LASR. Rbfox proteins affected the splicing activity of the LASR component hnRNP M, and conversely hnRNP M affected Rbfox activity on some exons. The identification of mutations in Rbfox that eliminate its interaction with LASR, or alter its higher-order assembly allowed us to test the requirement for these interactions in Rbfox mediated splicing. We used two Rbfox-regulated minigene reporters in *in vivo* splicing assays: Dup-E33 and Dup-E9* (Tang et al., 2009). Dup-E33 contains CaV1.2 exon 33, which has a downstream UGCAUG site required for Rbfox enhanced exon inclusion. Dup-E9* contains a modified CaV1.2 exon 9*, which has an upstream UGCAUG site and is repressed by Rbfox (Figure 4.4AB). These reporters were transiently expressed in a HEK293 cell line that has had endogenous Rbfox2 expression eliminated by CRISPR knock-out. As seen previously, E33 is largely skipped in Rbfox2$^{-/-}$ HEK293 cells, but its splicing is strongly stimulated by coexpressed wildtype Rbfox1. Conversely, E9* is spliced into the DUP-E9* mRNA in the Rbfox2$^{-/-}$ HEK293 cells and this splicing is strongly repressed by Rbfox1. In contrast, the ΔCT mutant of Rbfox1 neither activated E33 nor repressed E9* (Figure 4.4B). Note that the ΔCT protein carries the SV40 T antigen NLS to replace C-terminal nuclear localization signal of Rbfox1, and the ΔCT protein has not lost its nuclear localization (Figure 4.1B). The CT domain, which is required for LASR association, is also required for splicing regulation by Rbfox.

To test the requirement for the Rbfox tyrosine residues in splicing regulation, we expressed the minigene reporters in cells with series of mutant Rbfox proteins.

Interestingly, splicing repression and splicing activation showed different responses to these mutations. Activation of E33 was reduced by the tyrosine to serine mutations, and this effect depended on the number of tyrosines altered. Loss of three tyrosines led to a partial loss of activity, but the protein missing 10 or more tyrosines was only slightly more active than the F126A mutation that knocks out RNA binding by Rbfox1 (Figure 4.4D). In contrast, the repression of E9* splicing was only minimally affected by the tyrosine mutations; the protein missing 13 tyrosines was still much more active that F126A (Figure 4.4E and 4.1E). For these minigenes, the higher-order assembly of the Rbfox1/LASR complexes is needed for splicing activation, but dispensable for splicing repression.

To assess the effects of the tyrosine mutations on a larger number of exons and on exons expressed from endogenous genes, we applied RASL-seq (RNA-mediated oligonucleotide annealing, selection and ligation with next-generation sequencing) (Li et al., 2012a) to profile 5530 alternative splicing events in Rbfox2$^{-/-}$ HEK293 cell lines expressing Rbfox1 and its mutants. RNA was isolated from three biological replicates of cells expressing wildtype Rbfox1 (WT), the ten-tyrosine-to-serine mutant (10Y), the RNA-binding-defective F126A mutant (FA), and control cells (Ctrl). For each assayed alternative splicing event, the ratio of included isoform to excluded isoform (In/Ex ratio) was calculated. For all the exons assayable in HEK293 cells, these ratios were compared between the different cell lines and replicates. Exons exhibiting greater than a 1.5-fold change in In/Ex ratio between WT and control ($p<0.05$), and less than a 1.5-fold change between FA and control were defined as Rbfox1 regulated. By these criteria, 206

cassette exons were regulated by Rbfox1 (Figure 4.5A, Table S4.1). The In/Ex ratios for each exon across all the cell lines were subjected to unsupervised hierarchical clustering. As expected, replicates of cells expressing the same protein were most similar and the values for the FA cells were most similar to the control. The cells expressing WT protein were most different from the other cells. The cells expressing the 10Y protein showed the most variability across replicates. Looking at individual exons across the different cell lines, the exons clustered into three groups with distinct splicing behavior. Exons in Group 1 (29) were repressed by Rbfox1, while the 49 exons in Group 2 and the 128 exons in Group 3 were activated by Rbfox1 (Figure 4.5A and S4.2B, Table S4.1). As seen for DUP-E9*, the group 1 exons that were repressed by wildtype Rbfox1 are also largely repressed by the 10Y mutant proteins. For exons activated by Rbfox1, those in group 3 lost regulation by the 10Y mutation, as seen for DUP-E33 (Figure 4.5A). With some variation between replicates, exons in Group 2 were at least partially activated by 10Y, and many were activated to a similar level as by WT (Figure 4.5A). The median fold change of In/Ex ratio of WT to 10Y for exons in group 2 is close to 1, while there is more than 1.5-fold change between WT and 10Y for exons in group 3 (Figure 4.5B). As seen with the reporter gene exons and many endogenous exons (Figure 4.4DE, 4.5 and S4.2B), splicing repression and splicing activation showed different dependencies on the tyrosine residues of the CT domain. Exons repressed by Rbfox were not strongly affected by the Y-S mutations, but a substantial subset of exons activated by Rbfox requires these

residues and presumably the higher-order assembly of the Rbfox/LASR for splicing regulation.

DISCUSSION

Repetitive tyrosine residues in LC domains trigger higher-order assembly of Rbfox proteins with LASR

We showed that C-terminal tail of Rbfox proteins containing conserved LC sequences interact with LASR and this interaction is required for maintaining Rbfox activity in splicing regulation. A series of repetitive tyrosine residues within this LC domain could serve as one of the interfaces interacting with LASR, but more importantly, they are required in high-order assembly of Rbfox proteins. Tripeptide motif [G/S] Y [G/S] has been reported to be involved in forming amyloid-like fibers with cross-beta sheet structures (Kato et al., 2012). Not all the tyrosine residues in the repeats that we identified follow this pattern, but most of them either match or partially match. We also noticed an interesting spacing pattern of repetitive tyrosine residues, which might play a role in maintaining β-sheet structures. Structural determination of this domain will help to determine the role of tyrosines in forming such structures and spacing effect of these residues.

We find that many proteins in LASR consist of LC domains or disordered domains and all the major LASR subunits were in the core subset of proteins precipitated by b-isox

(Kato et al., 2012). Since we found that LASR could exist and form higher-order assembly by itself without Rbfox proteins, we would suspect that the mechanism we discovered for Rbfox1 higher-order assembly could be applied to other RNA-binding proteins to assemble LASR.

Association of Phase-separable and fiber forming properties with higher-order assembly

We demonstrated that polymerization of Rbfox1 through repetitive tyrosine residues in CT region *in vivo* correlates with its capability of forming fibers at high protein concentration and phase-separable droplets at low protein concentration with crowding reagent *in vitro*. We also observed that Rbfox1 LC domain could fibrillize over time and form phase-separable droplet structures, while its ten-tyrosine-to-serine mutant could not. The droplets formed at the condition with crowding reagent were able to from fibers when incubated for longer time. LC domain of Rbfox proteins could form fibrous structures without forming liquid droplets. However, phase-separation at low protein concentration promoted by crowding reagent could increase the local concentration of protein, leading to fiber formation ultimately.

We found that short RNA fragments protected by Rbfox-LASR complexes, which turned out to be the binding sites for Rbfox and other RNA-binding proteins in LASR (Unpublished, Ying, Y. & Black, D.L.). After nuclease digestion, these RNA were trimmed to short fragments, which are not likely to be the organizer of the complexes. However,

RNA could potentially initiate the assembly process to bring proteins closer, as it could promote phase separation and fibrous structure formation *in vitro* (Lin et al., 2015). Meanwhile, traditional protein-protein interactions as well as potential fiber-like polymerization by disordered and LC domains could occur to form higher-order complexes.

Higher-order assembly of Rbfox1 in alternative splicing regulation

Most studies of RBPs primarily focus on the RBD and their RNA recognition properties. Little is known about the role of non-RBD regions. Here we showed that non-RBD region at C-terminal of Rbfox proteins modulates the splicing activity of Rbfox proteins by interacting with LASR. Interestingly, the splicing activation activity could be further tuned by the capability to form higher-order assembly of Rbfox proteins. Specifically, Rbfox proteins fully capable of higher-order assembly activate splicing potently, mutants mildly impeded in higher-order assembly activate to an intermediate level, and mutants that are incapable of higher-order assembly fail to activate.

The observations from minigene reporter assays hold true in general for many exons that identified by RASL-seq. This add-on tunable layer of splicing regulation by higher-order assembly could precisely control the ratio of transcript abundance between different splice variants. Dysregulation of the higher-order assembly and disassembly might result in aberrant splicing regulation.

EXPERIMENTAL PROCEDURES

Plasmids

Rbfox1 and its mutants were cloned into destination vector pcDNA5-FRT/TO-3xHA-3xFLAG (a gift from James A. Wohlschlegel, UCLA) via the Gateway cloning system (Life Technologies) for expression in Flp-In™ T-Rex™ 293 cells. Tyrosine-to-serine mutations were introduced by QuickChange™ site-directed mutagenesis. CT or C2 domain of Rbfox1 and their tyrosine-to-serine mutants tagged with an N-terminal SNAP (NEB), mCherry or mEGFP were cloned into pET28aTev vector for expression in *E. coli*. Details are available in the Supplemental Experimental Procedures.

Cell culture

Flp-In™ T-Rex™ 293 cells stably expressing 3xHA-3xFLAG-SV40NLS-Rbfox1 and its deletion mutants, 3xHA-3xFLAG-Rbfox1 and its tyrosine mutants, 3xHA-3xFLAG-Rbfox1 and Rbfox2 B40 or M43 variants, were generated using the Flp-In™ system (Life Technologies) according to the manufacturer's instructions. All cell lines were cultured in Dulbecco's Modification of Eagle's Medium (Mediatech) containing L-glutamine and 10% FBS. Protein expression was induced with 500 ng/ml doxycycline for two days.

Antibodies

Primary antibodies used for western blot were as follows: FLAG (Sigma, F3165-1MG), GAPDH (Ambion, AM4300), snRNP70 (Sharma et al., 2005). The secondary antibodies were goat anti-mouse or anti-rabbit IgG conjugated with cy3, cy5 (GE Healthcare) or IRDye 680LT, 800CW (LI-COR). Typhoon imager 9410 (GE Healthcare) and Odyssey Infrared Imaging System (LI-COR) were used for detection.

Protein complexes analysis

Subcellular fractionation from cells was performed as described (Damianov et al., 2016). HMW protein fractions were subjected to glycerol gradient ultracentrifugation or FLAG-immunoprecipitation as described (Damianov et al., 2016). Proteins were analyzed by western blot or stained by SYPRO Ruby protein gel stain (Life Technologies).

Recombinant protein purification

Proteins were expressed in *E. coli* Rosetta2 (DE3) and purified with a HisTrap HP 5ml column (GE Healthcare), followed by a Hiload Superdex 200 column (GE Healthcare). Purified proteins were concentrated and frozen at -80 °C. Details are available in Supplemental Experimental Procedures.

Droplet Assembly

Droplet assembly was performed as describe previously (Lin et al., 2015). Proteins were fluorescently labeled with SNAP-Surface 549 or SNAP-Surface 649 (NEB) according to

the manufacturer's protocols. Purified bacterial protein solutions (2% of the proteins are fluorescently labeled) were diluted to 37.5 mM NaCl, 50 mM Tris (pH 7.5), and 1mM DTT. Reactions were placed in 96-well glass bottom plate (MatTek) coated with 3% BSA for 15 minutes and sealed with PCR plate film (USA scientific) to minimize evaporation. Images were acquired on LSM 510 Meta Confocal Microscope (Zeiss) or Eclipse TE2000 (Nikon).

Hydrogel formation

Hydrogel droplets of mEGFP-C2 of Rbfox1 were prepared as described previously (Kwon et al., 2013). In brief, proteins were concentrated to 1mM in gelation buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT). A droplet (2 µl) of protein solution was placed on a glass-bottomed dish (MatTek) sealed with parafilm and incubated at room temperature for 5 days.

Transmission Electron Microscopy

The protein solutions (5 µl) were loaded onto a glow-discharged TEM grid (Electron Microscopy Sciences, FCF400-Cu-SB) and stained with 0.8% uranyl formate. TEM images were obtained at 120 kV on T12 quick cryoEM and cryoET (FEI).

In vivo splicing assay

Transfections were performed with BioT (Bioland Scientific) according to the manufacturer's protocol. 48 hours after transfection, RNA was extracted from cells using Trizol (Life Technologies), and reverse transcribed with random hexamers and Superscript III (Life Technologies) according to manufacturer's instructions. Minigene reporter spliced products were amplified by PCR (22 cycles) and separated by denaturing urea-PAGE. Endogenous gene splice products were amplified by PCR (22-25 cycles) with addition of SYBR Green I dye (Bio-rad) in the reaction and resolved by native PAGE. The amplified splice products were detected by Typhoon imager 9410 (GE Healthcare) and quantified by ImageQuant^{TM} TL. Primer sequences are listed in Table S1.

RASL-seq splicing analysis

RASL-seq was performed as described with modifications (Li et al., 2012a). In brief, ligated oligos were subjected to two rounds of PCR amplification to include adapters compatible with the flow cell on Miseq. The library was sequenced on Miseq using custom sequencing primers. Sequenced reads were mapped to the oligo sequence pool by Blat (Kent, 2002) allowing two mismatches. Significantly changed splicing events were identified by average fold change and t-test ($p < 0.05$, comparing wildtype Rbfox1 to control). Details are available in Supplemental Experimental Procedures.

References

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

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

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R.*, et al.* (2012). The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science *338*, 1587-1593.

Barnby, G., Abbott, A., Sykes, N., Morris, A., Weeks, D.E., Mott, R., Lamb, J., Bailey, A.J., Monaco, A.P., and International Molecular Genetics Study of Autism, C. (2005). Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT. American journal of human genetics *76*, 950-966.

Berchowitz, L.E., Kabachinski, G., Walker, M.R., Carlile, T.M., Gilbert, W.V., Schwartz, T.U., and Amon, A. (2015). Regulated Formation of an Amyloid-like Translational Repressor Governs Gametogenesis. Cell.

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

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annual review of biochemistry *72*, 291-336.

Bouvet, P. (2001). Determination of nucleic acid recognition sequences by SELEX. Methods in molecular biology *148*, 603-610.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. Science *324*, 1729-1732.

Calabretta, S., and Richard, S. (2015). Emerging Roles of Disordered Sequences in RNA-Binding Proteins. Trends Biochem Sci *40*, 662-672.

Carreira-Rosario, A., Bhargava, V., Hillebrand, J., Kollipara, R.K., Ramaswami, M., and Buszczak, M. (2016). Repression of Pumilio Protein Expression by Rbfox1 Promotes Germ Cell Differentiation. Developmental cell *36*, 562-571.

Conchillo-Sole, O., de Groot, N.S., Aviles, F.X., Vendrell, J., Daura, X., and Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC bioinformatics *8*, 65.

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

Damianov, A., Ying, Y., Lin, C.H., Lee, J.A., Tran, D., Vashisht, A.A., Bahrami-Samani, E., Xing, Y., Martin, K.C., Wohlschlegel, J.A.*, et al.* (2016). Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. Cell *165*, 606-619.

Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C., Eckmann, C.R., Myong, S., and Brangwynne, C.P. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. Proceedings of the National Academy of Sciences of the United States of America *112*, 7189-7194.

Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet *15*, 689-701.

Fukumura, K., Kato, A., Jin, Y., Ideue, T., Hirose, T., Kataoka, N., Fujiwara, T., Sakamoto, H., and Inoue, K. (2007). Tissue-specific splicing regulator Fox-1 induces exon skipping by interfering E complex formation on the downstream intron of human F1gamma gene. Nucleic acids research *35*, 5303-5311.

Fukumura, K., Taniguchi, I., Sakamoto, H., Ohno, M., and Inoue, K. (2009). U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. Nucleic acids research *37*, 1907-1914.

Gao, C., Ren, S., Lee, J.H., Qiu, J., Chapski, D.J., Rau, C.D., Zhou, Y., Abdellatif, M., Nakano, A., Vondriska, T.M.*, et al.* (2016). RBFox1-mediated RNA splicing regulates cardiac hypertrophy and heart failure. The Journal of clinical investigation *126*, 195-206.

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

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

Han, H., Irimia, M., Ross, P.J., Sung, H.K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N.*, et al.* (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature *498*, 241-245.

Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G.*, et al.* (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell *149*, 768-779.

Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., Uversky, V.N., Vidal, M., and Iakoucheva, L.M. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. Plos Comput Biol *2*, 890-901.

Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., and Parker, R. (2016). ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. Cell *164*, 487-498.

Jangi, M., Boutz, P.L., Paul, P., and Sharp, P.A. (2014). Rbfox2 controls autoregulation in RNA-binding protein networks. Genes & development *28*, 637-651.

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

Kalsotra, A., Xiao, X., Ward, A.J., Castle, J.C., Johnson, J.M., Burge, C.B., and Cooper, T.A. (2008). A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. Proceedings of the National Academy of Sciences of the United States of America *105*, 20333-20338.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J.*, et al.* (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753-767.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome research *12*, 656-664.

Kim, K.K., Adelstein, R.S., and Kawamoto, S. (2009). Identification of neuronal nuclei (NeuN) as Fox-3, a new member of the Fox-1 gene family of splicing factors. The Journal of biological chemistry *284*, 31052-31061.

Kuroyanagi, H. (2009). Fox-1 family of RNA-binding proteins. Cellular and molecular life sciences : CMLS *66*, 3895-3907.

Kuroyanagi, H., Ohno, G., Mitani, S., and Hagiwara, M. (2007). The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo. Molecular and cellular biology *27*, 8612-8621.

Kuwasako, K., Takahashi, M., Unzai, S., Tsuda, K., Yoshikawa, S., He, F., Kobayashi, N., Guntert, P., Shirouzu, M., Ito, T.*, et al.* (2014). RBFOX and SUP-12 sandwich a G base to

cooperatively regulate tissue-specific splicing. Nature structural & molecular biology *21*, 778-786.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell *155*, 1049-1060.

Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S.L. (2014). Poly-dipeptides encoded by the C9ORF72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science.

Lee, J.A., Damianov, A., Lin, C.H., Fontes, M., Parikshak, N.N., Anderson, E.S., Geschwind, D.H., Black, D.L., and Martin, K.C. (2016). Cytoplasmic Rbfox1 Regulates the Expression of Synaptic and Autism-Related Genes. Neuron *89*, 113-128.

Lee, J.A., Tang, Z.Z., and Black, D.L. (2009). An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. Genes & development *23*, 2284-2293.

Li, H., Qiu, J., and Fu, X.D. (2012a). RASL-seq for massively parallel and quantitative analysis of gene expression. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Chapter 4*, Unit 4 13 11-19.

Li, P.L., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F*., et al.* (2012b). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336-U129.

Lin, Y., Protter, D.S., Rosen, M.K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Molecular cell *60*, 208-219.

Lovci, M.T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T.Y., Stark, T.J., Gehman, L.T., Hoon, S*., et al.* (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. Nature structural & molecular biology *20*, 1434-1442.

Maekawa, S., Leigh, P.N., King, A., Jones, E., Steele, J.C., Bodi, I., Shaw, C.E., Hortobagyi, T., and Al-Sarraj, S. (2009). TDP-43 is consistently co-localized with ubiquitinated inclusions in sporadic and Guam amyotrophic lateral sclerosis but not in familial amyotrophic lateral sclerosis with and without SOD1 mutations. Neuropathology *29*, 672-683.

Martin, C.L., Duvall, J.A., Ilkin, Y., Simon, J.S., Arreaza, M.G., Wilkes, K., Alvarez-Retuerto, A., Whichello, A., Powell, C.M., Rao, K*., et al.* (2007). Cytogenetic and

molecular characterization of A2BP1/FOX1 as a candidate gene for autism. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics *144B*, 869-876.

Mauger, D.M., Lin, C., and Garcia-Blanco, M.A. (2008). hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. Molecular and cellular biology *28*, 5403-5419.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. Science *338*, 1593-1599.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. Cell *163*, 123-133.

Murakami, T., Qamar, S., Lin, J.Q., Schierle, G.S., Rees, E., Miyashita, A., Costa, A.R., Dodd, R.B., Chan, F.T., Michel, C.H.*, et al.* (2015). ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. Neuron *88*, 678-690.

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

Neelamraju, Y., Hashemikhabir, S., and Janga, S.C. (2015). The human RBPome: From genes and proteins to human disease. J Proteomics *127*, 61-70.

Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D.*, et al.* (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Molecular cell *57*, 936-947.

Nutter, C.A., Jaworski, E.A., Verma, S.K., Deshmukh, V., Wang, Q., Botvinnik, O.B., Lozano, M.J., Abass, I.J., Ijaz, T., Brasier, A.R.*, et al.* (2016). Dysregulation of RBFOX2 Is an Early Event in Cardiac Pathogenesis of Diabetes. Cell reports.

Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M.*, et al.* (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell *162*, 1066-1077.

Ponthier, J.L., Schluepen, C., Chen, W., Lersch, R.A., Gee, S.L., Hou, V.C., Lo, A.J., Short, S.A., Chasis, J.A., Winkelmann, J.C.*, et al.* (2006). Fox-2 splicing factor binds to a

conserved intron motif to promote inclusion of protein 4.1R alternative exon 16. The Journal of biological chemistry *281*, 12468-12474.

Schwartz, J.C., Wang, X., Podell, E.R., and Cech, T.R. (2013). RNA seeds higher-order assembly of FUS protein. Cell reports *5*, 918-925.

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

Singh, R., and Valcarcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. Nature structural & molecular biology *12*, 645-653.

Singh, R.K., Xia, Z., Bland, C.S., Kalsotra, A., Scavuzzo, M.A., Curk, T., Ule, J., Li, W., and Cooper, T.A. (2014). Rbfox2-coordinated alternative splicing of Mef2d and Rock2 controls myoblast fusion during myogenesis. Molecular cell *55*, 592-603.

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

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

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212-1215.

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

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P.*, et al.* (2009). Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. Science *323*, 1208-1211.

Venables, J.P., Brosseau, J.P., Gadea, G., Klinck, R., Prinos, P., Beaulieu, J.F., Lapointe, E., Durand, M., Thibault, P., Tremblay, K.*, et al.* (2013a). RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. Molecular and cellular biology *33*, 396-405.

Venables, J.P., Lapasset, L., Gadea, G., Fort, P., Klinck, R., Irimia, M., Vignal, E., Thibault, P., Prinos, P., Chabot, B.*, et al.* (2013b). MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. Nature communications *4*, 2480.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wang, J.T., Smith, J., Chen, B.C., Schmidt, H., Rasoloson, D., Paix, A., Lambrus, B.G., Calidas, D., Betzig, E., and Seydoux, G. (2014). Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically-disordered proteins in C-elegans. eLife *3*.

Wei, C., Qiu, J., Zhou, Y., Xue, Y., Hu, J., Ouyang, K., Banerjee, I., Zhang, C., Chen, B., Li, H.*, et al.* (2015). Repression of the Central Splicing Regulator RBFox2 Is Functionally Linked to Pressure Overload-Induced Heart Failure. Cell reports.

Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q.*, et al.* (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell reports *6*, 1139-1152.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009a). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nature structural & molecular biology *16*, 130-137.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009b). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nature structural & molecular biology *16*, 130-137.

Zhou, H.L., Baraniak, A.P., and Lou, H. (2007). Role for Fox-1/Fox-2 in mediating the neuronal pathway of calcitonin/calcitonin gene-related peptide alternative RNA processing. Molecular and cellular biology *27*, 830-841.

Zhou, H.L., and Lou, H. (2008). Repression of prespliceosome complex formation at two distinct steps by Fox-1/Fox-2 proteins. Molecular and cellular biology *28*, 5507-5516.

FIGURE LEGENDS

Figure 4.1: Repetitive tyrosine residues in CT domain of Rbfox mediate its higher-order assembly.

(A) Schematic diagram of Rbfox1 and its deletion mutants.

(B) Immunoprecipitation of HA-FLAG-SV40NLS-Rbfox1 deletion mutants with LASR. Soluble and HMW nuclear fractions were prepared from Flp-In T-REx 293 cells stably expressing HA-FLAG-SV40NLS-Rbfox1 deletion mutants. Protein extracts were immunoprecipitated with antibodies to FLAG. Immunoprecipitates were separated by 4-12% SDS-PAGE and stained by SYPRO Ruby. Arrowheads indicate HA-FLAG-SV40NLS-Rbfox1 deletion mutants. LASR subunits are indicated on the right.

(C) Sedimentation of Rbfox1 deletion mutants through 10-50% glycerol density gradients. HMW fractions prepared from Flp-In T-REx 293 cells stably expressing HA-FLAG-SV40NLS-Rbfox1 deletion mutants were loaded onto 10-50% glycerol density gradients. Gradient fractions from top to bottom run from left to right. Proteins from odd gradient fractions were immunoblotted with antibody to FLAG. 40S and 60S markers are indicated below.

(D) Amino acid sequences encoded by exon B40 and M43 in Rbfox1 or Rbfox2 (top). Sedimentation of Rbfox1 and Rbfox2 splice variants through 10-50% glycerol density gradients (bottom). HMW fractions prepared from Flp-In T-REx 293 cells stably expressing HA-FLAG-Rbfox1 and HA-FLAG-Rbfox2 splice variants were loaded onto 10-50% glycerol density gradients. Gradient fractions from top to bottom run from left to right. Proteins from odd gradient fractions were immunoblotted with antibody to FLAG. 40S and 60S markers are indicated below.

(E) Amino acid sequence alignments of CT domain in Rbfox1, Rbfox2 and Rbfox3 variant

with exon B40 included. Tyrosine residues that have been examined by mutagenesis are shown in red.

(F) Sedimentation of Rbfox1 tyrosine mutants through 10-50% glycerol density gradients. HMW fractions prepared from Flp-In T-REx 293 cells stably expressing HA-FLAG-Rbfox1 tyrosine mutants were loaded onto 10-50% glycerol density gradients. Gradient fractions from top to bottom run from left to right. Proteins from odd gradient fractions were immunoblotted with antibody to FLAG. 40S and 60S markers are indicated below.

(G) Immunoprecipitation of HA-FLAG-Rbfox1 tyrosine-to-serine mutants with LASR. Soluble and HMW nuclear fractions were prepared from Flp-In T-REx 293 cells stably expressing HA-FLAG-Rbfox1 tyrosine-to-serine mutants. Protein extracts were immunoprecipitated with antibody to FLAG. Immunoprecipitates were separated by 4-12% SDS-PAGE and stained by SYPRO Ruby. HA-FLAG-Rbfox1 tyrosine-to-serine mutants and LASR subunits are indicated on the right.

Figure 4.2: Rbfox1 interacts with LASR through multiple interfaces of CT domain.

(A) Schematic diagram of Rbfox1 CT mutants (top). Immunoprecipitation of HA-FLAG-SV40NLS-Rbfox1 CT mutants with LASR (bottom). Soluble and HMW nuclear fractions were prepared from Flp-In T-REx 293 cells stably expressing HA-FLAG-SV40NLS-Rbfox1 CT mutants. Protein extracts were immunoprecipitated with antibody to FLAG. Immunoprecipitates were separated by 4-12% SDS-PAGE

and stained by SYPRO Ruby. Arrowheads indicate HA-FLAG-SV40NLS-Rbfox1 CT mutants. LASR subunits are indicated on the right. Asterisks mark non-specific bands.

(B) Sedimentation of Rbfox1 CT mutants through 10-50% glycerol density gradients. HMW fractions prepared from Flp-In T-REx 293 cells stably expressing HA-FLAG-SV40NLS-Rbfox1 CT mutants were loaded onto 10-50% glycerol density gradients. Gradient fractions from top to bottom run from left to right. Proteins from odd gradient fractions were immunoblotted with antibody to FLAG. 40S and 60S markers are indicated below.

Figure 4.3: LC domain of Rbfox1 form fibrous aggregates *in vitro*.

(A) Schematic diagram of SNAP-tagged Rbfox1 CT domain with a C-terminal 6xHis tag and its ten-tyrosine-to-serine mutant (top). Fluorescence microscopy images of the macroscopic structures formed by SNAP-CT and SNAP-CT-YS in 37.5 mM and 150 mM NaCl (bottom). 2% of proteins were labeled with SNAP-surface 649. Images were taken 21 hours and 48 hours after the initiation of phase separation by lowering the salt concentration. Scale bar: 5 μm.

(B) Schematic diagram of SNAP-tagged Rbfox1 C2 fragment with a C-terminal 6xHis tag and its ten-tyrosine-to-serine mutant (top). Confocal fluorescence microscopy images of the macroscopic structures formed by SNAP-C2 and SNAP-C2-YS in 37.5 mM and 150 mM NaCl (bottom). 2% of proteins were labeled with SNAP-surface

549. Images were taken 24 hours after the initiation of phase separation by lowering the salt concentration. Scale bar: 5 μm.

(C) Schematic diagram of SNAP or mEGFP-tagged Rbfox1 C2 fragment with a C-terminal 6xHis tag and its ten-tyrosine-to-serine mutant (top). Confocal fluorescence images showing the co-partitioning of C2 into spherical fluorescent droplets (bottom). 0.5 μM of each protein were mixed with addition of 10% PEG8000. Images were taken 24 hours post-incubation. Scale bar: 5 μm.

Figure 4.4: The tyrosine-rich region in CT domain of Rbfox1 is required for GCAUG-dependent exon activation.

(A) Schematic diagram of DUP-E33 and the modified DUP-E9* minigene reporters. The downstream UGCAUG site and its mutation of E33 are indicated. The exonic and downstream UGCAUG from the original DUP-E9* were mutated to generate the modified DUP-E9* with upstream UGCAUG intact. This modified DUP-E9* minigene was used in this study.

(B) *In vivo* splicing assays with Rbfox1 CT deletion mutant on E33WT, E33mt and E9*WT, E9*mt minigene reporters. Denaturing gel electrophoresis of RT-PCR products for minigene reporters upon co-expression with HA-FLAG-SV40NLS-Rbfox1 deletion mutants. Spliced products are indicated on the right. A graph showing PSI (percentage spliced in) calculated from four independent experiments is shown above each gel. Error bar dictates SEM. **** $p < 0.0001$ by

unpaired, one-tailed Student's t test between Rbfox1 CT deletion (ΔCT) and full length Rbfox1 (FL).

(C) Representative western blot of protein expression level in splicing assays in (B). Whole cell lysates were prepared from Flp-In T-REx 293 cells transiently expressing HA-FLAG-SV40NLS-Rbfox1 deletion mutants and probed with antibodies to FLAG and snRNP70.

(D) *In vivo* splicing assays with Rbfox tyrosine mutants on E33WT and E33mt (right) minigenes. Denaturing gel electrophoresis of RT-PCR products for E33WT and E33mt minigenes upon co-expression with HA-FLAG-Rbfox1 tyrosine-to-serine mutants. Spliced products are indicated on the right. A graph showing PSI calculated from three independent experiments is shown above each gel. Error bar dictates SEM. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ by unpaired, one-tailed Student's t test between each Rbfox1 tyrosine-to-serine mutant and wildtype Rbfox1.

(E) *In vivo* splicing assays with Rbfox tyrosine mutants on E9*WT and E9*mt minigenes. Denaturing gel electrophoresis of RT-PCR products for E9* WT and E9*mt minigenes upon co-expression with HA-FLAG-Rbfox1 tyrosine-to-serine mutants. Spliced products are indicated on the right. A graph showing PSI calculated from three independent experiments is shown above each gel. Error bar dictates SEM. n.s., not significant by unpaired, one-tailed Student's t test between each Rbfox1 tyrosine-to-serine mutants and wildtype Rbfox1.

(F) Representative western blot of protein expression level in splicing assays in (D) and

(E). Whole cell lysates were prepared from Flp-In T-REx 293 cells transiently expressing HA-FLAG-Rbfox1 tyrosine mutants and probed with antibodies to FLAG and GAPDH.

Figure 4.5: Higher-order assembly of Rbfox1 is needed for splicing activation of many endogenous exons.

(A) Heatmap was generated by hierarchical clustering of cassette-exon splicing (206 events) significantly changed by Rbfox1 from RASL-seq.

(B) Boxplot showing the fold change of WT In/Ex ratio versus 10Y In/Ex ratio in Group 1, 2 and 3, respectively. Each dot represents a cassette-exon splicing event.

Supplementary Figure 4.1: CT domain of Rbfox1 aggregated in vitro.

(A) Plots of “hot spots” of aggregation in CT of Rbfox1 predicted by AGGRESCAN. Wildtype Rbfox1 CT and its ten-tyrosine-to-serine mutant were shown in red and blue, respectively. Amino acid sequences of Rbfox1 CT domain were shown below. Low complexity sequences predicted by SEG program were shaded in grey. Sequences in red were “hot spots” for aggregation. Ten tyrosine residues examined by mutagenesis were underlined in C2.

(B) Chromatographs of recombinant proteins purified from E. coli. Molecular weight markers are indicated on top of the graphs.

(C) Protein gels of recombinant proteins purified from E. coli stained by Coomassie blue.

(D) Confocal fluorescence microscopy images of SNAP-Surface 549 labeled SNAP-C2 but not SNAP-C2-YS forming droplets at low protein concentration in the presence of 10% PEG 8000. Scale bar: 5 μm.

(E) Fluorescence microscopy images of SNAP-C2 forming fibrous structures stained with Thioflavin T.

(F) Transmission electron micrograph of fibrous structures formed by SNAP-C2.

Supplementary Figure 4.2: Significantly changed cassette exons identified from RASL-seq, related to Figure 4.5.

(A) Western blot showing protein expression level of 12 samples used in RASL-seq. Whole cell lysate were probed with antibodies to FLAG and GAPDH.

(B) RT-PCR validation of exons in Group 1, 2 and 3 identified from RASL-seq. Spliced products are indicated on the right. PSI values were shown below the gels.

Table S4.1: related to Figure 4.4 and 4.5, Supplementary Figure 4.2.

(A) RASL-seq sequencing reads mapping information.

(B) Splicing of cassette exon significantly changed by Rbfox1 from RASL-seq.

(C) PCR Primers used in *in vivo* splicing assays.

Supplemental Experimental Procedures

Plasmids

The SNAP-pET28aTev, mCherry-pET28aTev or mEGFP-pET28aTev vectors were constructed by two-step cloning. First, the thrombin cleavage site of the original pET28a vector (Novagen) was replaced with a Tev cleavage site by QuickChange mutagenesis generating the pET28aTev vector. Then, the SNAP tag (NEB), mCherry, or mEGFP was inserted into the vector pET28aTev before the Tev cleavage site by Gibson Assembly (NEB). CT or C2 domain of Rbfox1 and their ten-tyrosine-to-serine mutants were cloned into the SNAP-pET28aTev, mCherry-pET28aTev or mEGFP-pET28aTev vectors after the Tev cleavage site. All constructs contain an N-terminal SNAP, mCherry or mEGFP tag and a C-terminal 6xHis tag. The monomeric mutation (A206K) was made for EGFP. The sequences of all resulting vectors were confirmed by DNA sequencing.

Recombinant protein purification

The proteins were expressed in *E. coli* strain Rosetta2 (DE3). Bacteria were cultured in LB medium supplemented with 50 μg/ml Kanamycin and 34 μg/ml chloramphenicol at 37 °C and induced at OD600 of 0.6–0.8 with 0.5 mM IPTG at 25 °C for 4 hours. The harvested cells were resuspended in equilibration buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl) and lysed by sonication. After centrifugation, the soluble fractions of the cell lysates were loaded onto a HisTrap HP 5ml column (GE Healthcare) equilibrated with equilibration buffer. The columns were washed with equilibration buffer containing 50 mM imidazole, and the target proteins were eluted with a linear gradient of equilibration buffer

containing 50-500 mM imidazole. The proteins were further purified by a Hiload Superdex 200 column (GE Healthcare) equilibrated in the buffer containing 50 mM Tris-HCl pH 7.5, and 150 mM NaCl. The purified proteins were concentrated with Amicon Ultra centrifugal filters (Millipore), and stored at -80 °C. The purities of purified proteins were confirmed by SDS-PAGE, and the concentrations were determined by absorbance at UV280.

RASL-seq splicing analysis

RASL-seq was performed as described with modifications (Li et al., 2012a). Total RNA from Flp-In T-REx 293 cells stably expressing Rbfox1 mutants were extracted with Trizol (Life Technologies) and treated with DNase I. RASL-seq oligos were annealed to 1 μg of total RNA. After ligation, 5 μl eluted ligated oligos were used for 10 cycles of PCR amplification using primers F1: 5'-CCGAGATCTACACTCTTTCCCTACACGACGGCGA CCACCGAGAT-3' and R1: 5'- GTGACTGGAGTTCAGACGTGTGCGCTGATGCTACGA CCACAGG-3'. Half of the resulting PCR products were used in the second round of PCR amplification (10-15 cycles) using primers F2: 5'- AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACG-3' and D701-D712 adapters (Illumina). The indexed PCR products were pooled and sequenced on Miseq with a custom sequencing primer 5'-ACACTCTTTCCCTACACGACGGCGACCACCGAGAT-3' and a custom index sequencing primer 5'-TAGCATCAGCGCACACGTCTGAACTCCAGTCAC-3'. Sequencing data were mapped to RASL-seq oligo pool sequences with Blat allowing for two mismatches. On average, 1.3 million reads were obtained for each sample. Splicing

events were filtered for a minimum of 5 reads averaged across all samples. Ratios of the counts of long to short isoforms (In/Ex ratio) were calculated. The significantly changed events were identified by average fold change and the t-test (p-value < 0.05, comparing wildtype Rbfox1 to control). Heatmaps were generated by hierarchical clustering using R.

Figure 4.1

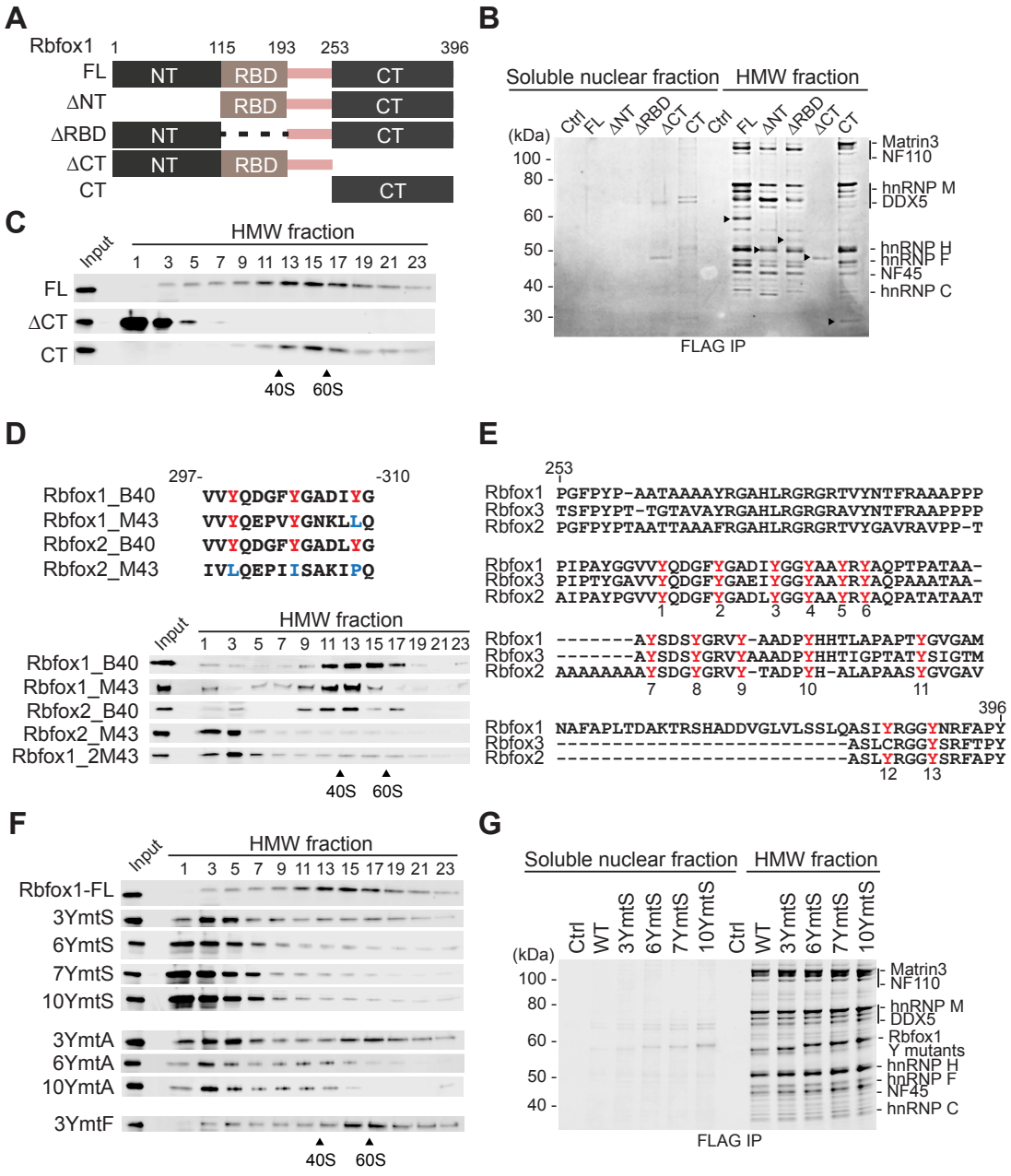

A
Rbfox1
1
115
193
253
396
FL
NT
RBD
CT
ΔNT
ΔRBD
ΔCT
B
Soluble nuclear fraction
HMW fraction
Ctrl
FL
ΔNT
ΔRBD
ΔCT
CT
(kDa)
100
80
60
50
40
30
Matrin3
NF110
hnRNP M
DDX5
hnRNP H
hnRNP F
NF45
hnRNP C
FLAG IP
C
Input
HMW fraction
1 3 5 7 9 11 13 15 17 19 21 23
FL
ΔCT
CT
40S
60S
D
297-
-310
Rbfox1_B40 VVYQDGFYGADIYG
Rbfox1_M43 VVYQEPVYGNKLLQ
Rbfox2_B40 VVYQDGFYGADLYG
Rbfox2_M43 IVLQEPIISAKIPQ
Input
HMW fraction
1 3 5 7 9 11 13 15 17 19 21 23
Rbfox1_B40
Rbfox1_M43
Rbfox2_B40
Rbfox2_M43
Rbfox1_2M43
40S
60S
E
253
Rbfox1 PGFPYP-AATAAAAYRGAHLRGRGRTVYNTFRAAAPPP
Rbfox3 TSFPYPT-TGTAVAYRGAHLRGRGRAVYNTFRAAPPPP
Rbfox2 PGFPYPTAATTAAAFRGAHLRGRGRTVYGAVRAVPP-T
Rbfox1 PIPAYGGVVYQDGFYGADIYGGYAAYRYAQPTPATAA-
Rbfox3 PIPTYGAVVYQDGFYGAEIYGGYAAYRYAQPAAATAA-
Rbfox2 AIPAYPGVVYQDGFYGADLYGGYAAYRYAQPATATAAT
1 2 3 4 5 6
Rbfox1 -------AYSDSYGRVY-AADPYHHTLAPAPTYGVGAM
Rbfox3 -------AYSDSYGRVYAAADPYHHTIGPTATYSIGTM
Rbfox2 AAAAAAAAYSDGYGRVY-TADPYH-ALAPAASYGVGAV
7 8 9 10 11
396
Rbfox1 NAFAPLTDAKTRSHADDVGLVLSSLQASIYRGGYNRFAPY
Rbfox3 ----------ASLCRGGYSRFTPY
Rbfox2 ----------ASLYRGGYSRFAPY
12 13
F
Input
HMW fraction
1 3 5 7 9 11 13 15 17 19 21 23
Rbfox1-FL
3YmtS
6YmtS
7YmtS
10YmtS
3YmtA
6YmtA
10YmtA
3YmtF
40S
60S
G
Soluble nuclear fraction
HMW fraction
Ctrl
WT
3YmtS
6YmtS
7YmtS
10YmtS
(kDa)
100
80
60
50
40
Matrin3
NF110
hnRNP M
DDX5
Rbfox1
Y mutants
hnRNP H
hnRNP F
NF45
hnRNP C
FLAG IP

Figure 4.2

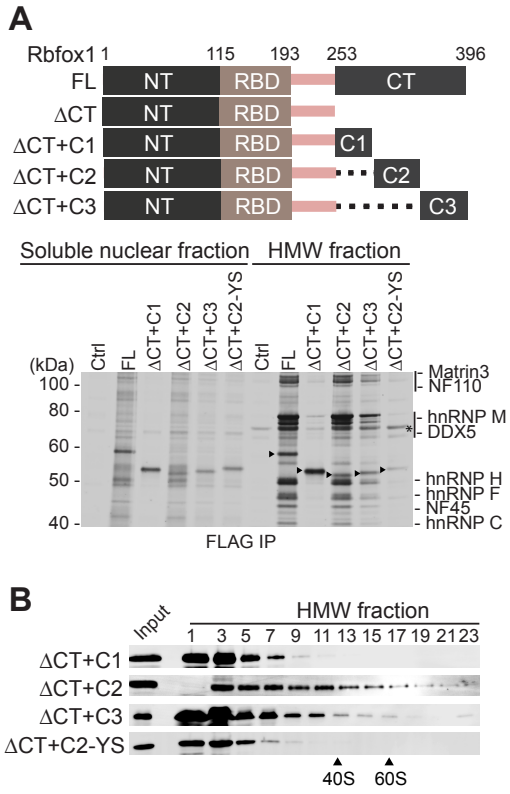

A
Rbfox1 1
115
193
253
396
FL
NT
RBD
CT
ΔCT
ΔCT+C1
C1
ΔCT+C2
C2
ΔCT+C3
C3
Soluble nuclear fraction
HMW fraction
Ctrl
FL
ΔCT+C1
ΔCT+C2
ΔCT+C3
ΔCT+C2-YS
(kDa)
100 -
80 -
60 -
50 -
40 -
Matrin3
NF110
hnRNP M
DDX5
hnRNP H
hnRNP F
NF45
hnRNP C
FLAG IP
B
Input
HMW fraction
1 3 5 7 9 11 13 15 17 19 21 23
ΔCT+C1
ΔCT+C2
ΔCT+C3
ΔCT+C2-YS
40S
60S

Figure 4.3

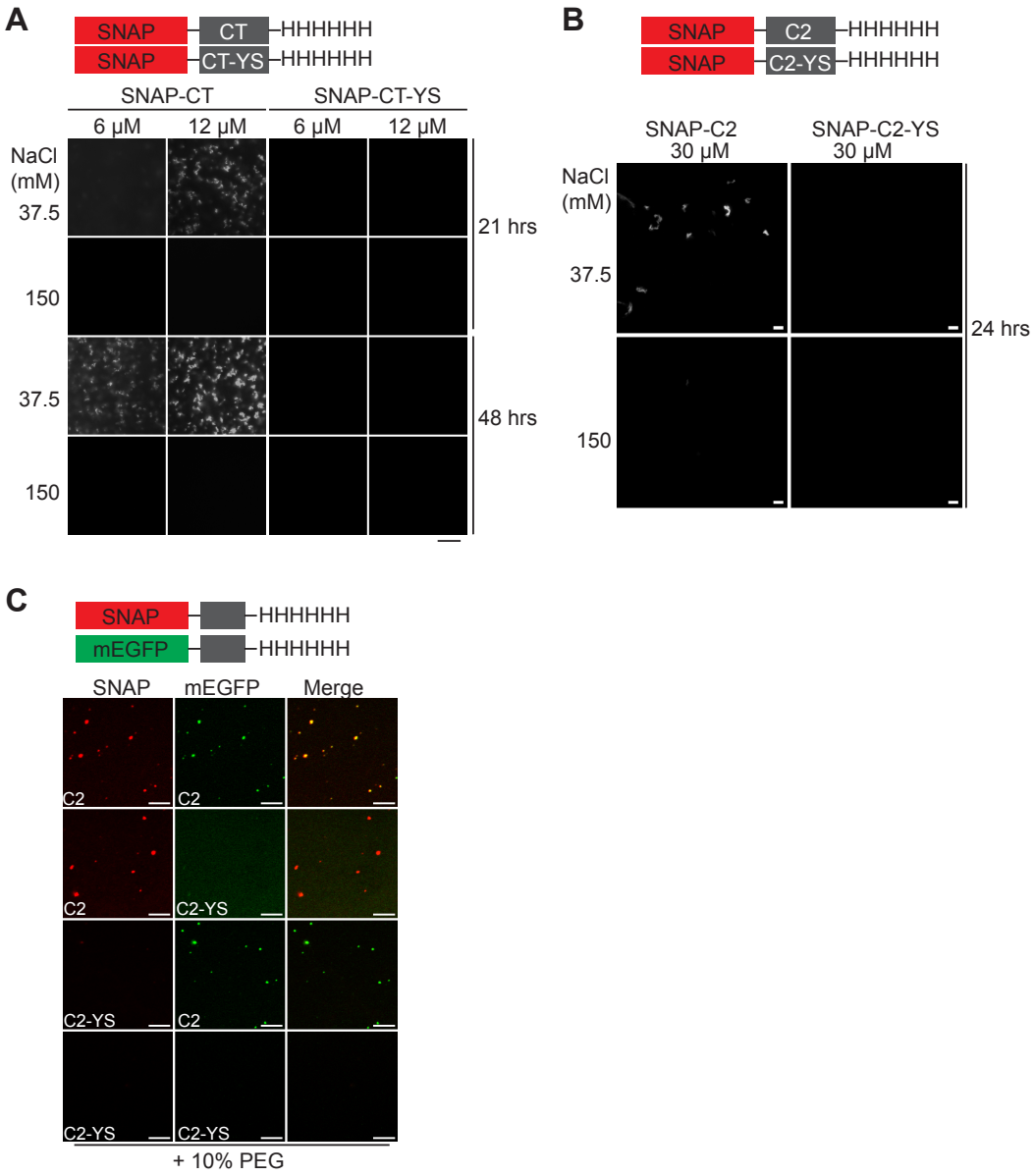

A
SNAP
CT
HHHHHH
SNAP
CT-YS
HHHHHH
SNAP-CT
SNAP-CT-YS
6 μM
12 μM
6 μM
12 μM
NaCl (mM)
37.5
150
37.5
150
21 hrs
48 hrs
B
SNAP
C2
HHHHHH
SNAP
C2-YS
HHHHHH
SNAP-C2 30 μM
SNAP-C2-YS 30 μM
NaCl (mM)
37.5
150
24 hrs
C
SNAP
HHHHHH
mEGFP
HHHHHH
SNAP
mEGFP
Merge
C2
C2
C2
C2-YS
C2-YS
C2
C2-YS
C2-YS
+ 10% PEG

Figure 4.4

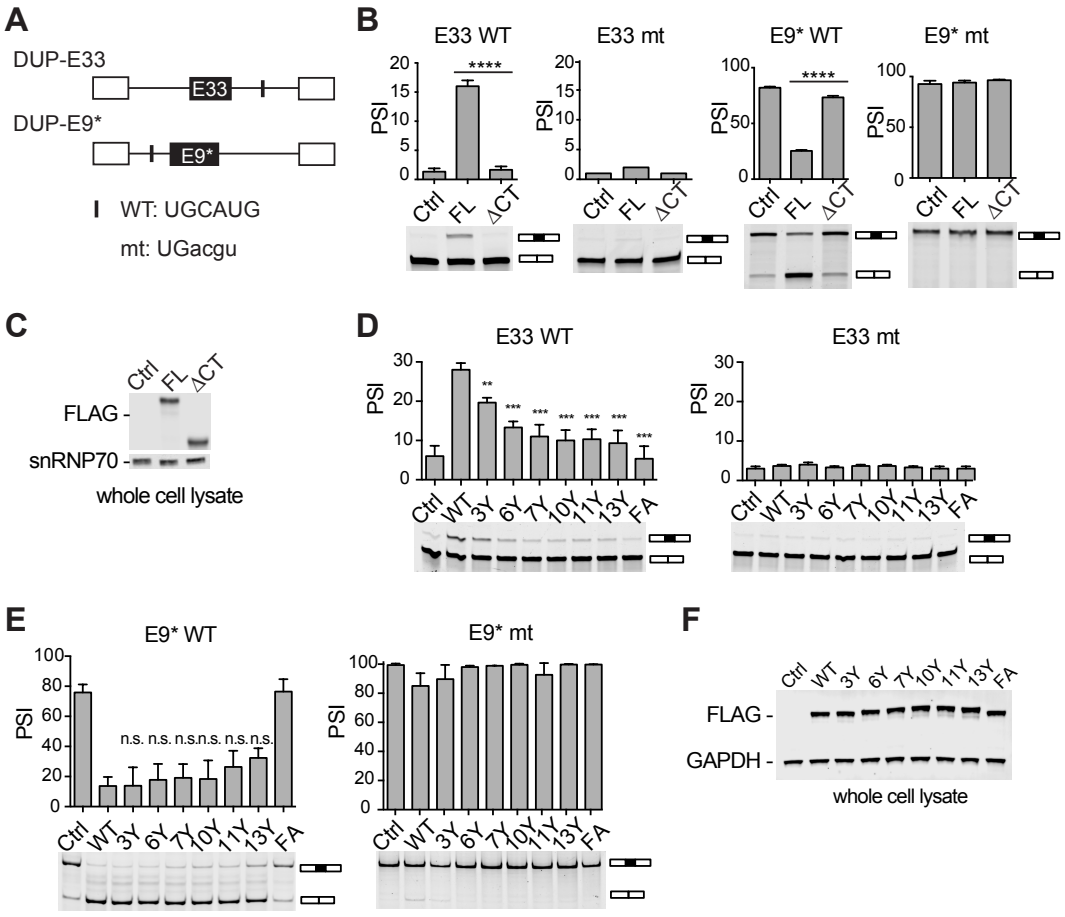
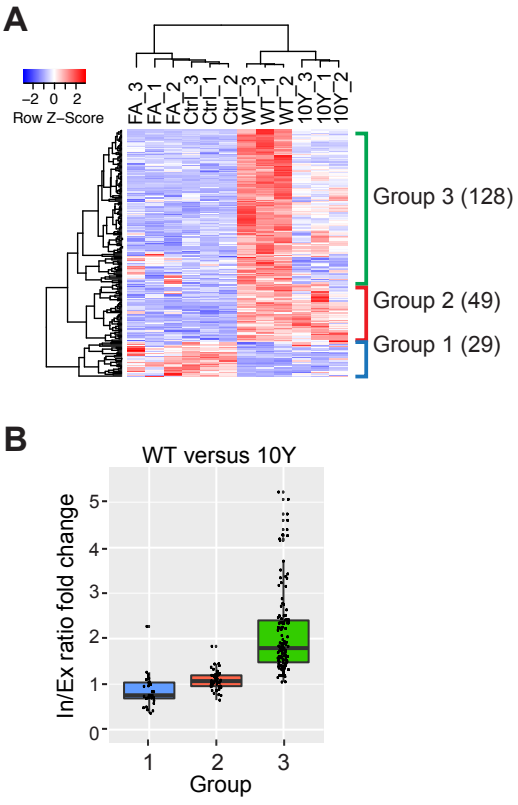
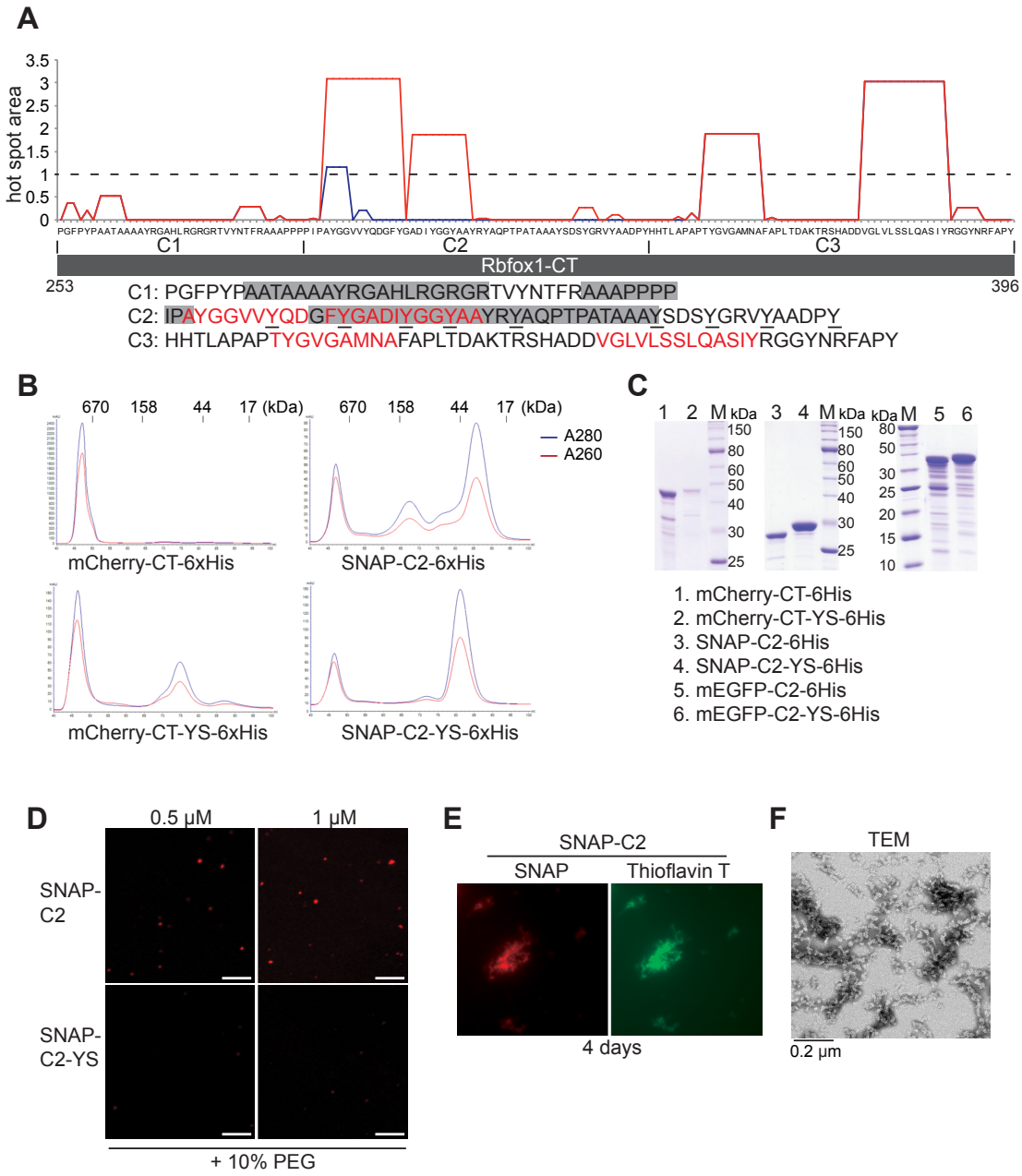

A
DUP-E33
E33
DUP-E9*
E9*
WT: UGCAUG
mt: UGacgu
B
E33 WT
E33 mt
E9* WT
E9* mt
PSI
****
Ctrl
FL
ΔCT
C
Ctrl
FL
ΔCT
FLAG -
snRNP70 -
whole cell lysate
D
E33 WT
E33 mt
PSI
**
***
Ctrl
WT
3Y
6Y
7Y
10Y
11Y
13Y
FA
E
E9* WT
E9* mt
PSI
n.s.
F
FLAG -
GAPDH -
whole cell lysate

Figure 4.5


A
−2 0 2
Row Z−Score
FA_3
FA_1
FA_2
Ctrl_3
Ctrl_1
Ctrl_2
WT_3
WT_1
WT_2
10Y_3
10Y_1
10Y_2
Group 3 (128)
Group 2 (49)
Group 1 (29)
B
WT versus 10Y
ln/Ex ratio fold change
0
1
2
3
4
5
1
2
3
Group

Supplementary Figure 4.1



Supplementary Figure 4.2

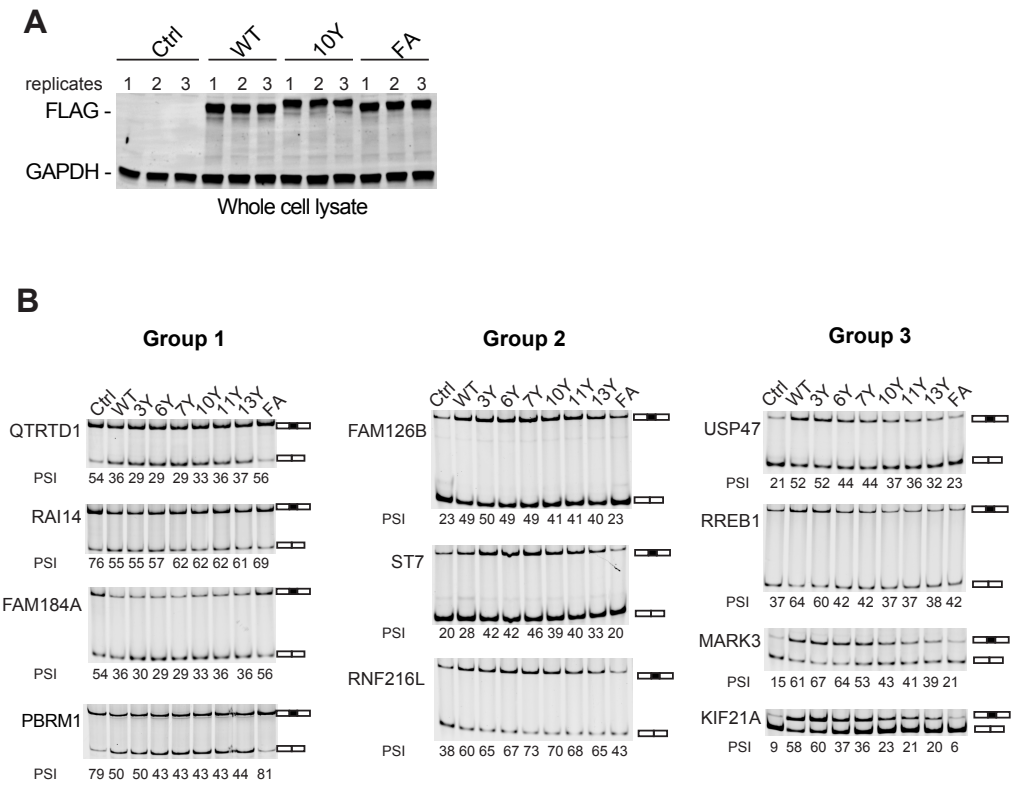


Table S4.1A. RASL-seq sequencing reads mapping information.

Sample	Ctrl1	Ctrl2	Ctrl3
raw reads	2,107,845	2,302,560	1,883,733
blat mapped(2 mismatches)	1,851,866	1,943,012	1,531,724
%mapped	88	84	81

Sample	WT1	WT2	WT3
raw reads	2,125,022	1,628,396	1,642,182
blat mapped(2 mismatches)	1,700,616	1,250,668	1,315,270
%mapped	80	77	80

Sample	10Y1	10Y2	10Y3
raw reads	1,705,217	1,597,497	1,299,457
blat mapped(2 mismatches)	1,414,190	1,247,133	1,063,517
%mapped	83	78	82

Sample	FA1	FA2	FA3
raw reads	1,630,762	1,050,440	943,323
blat mapped(2 mismatches)	908,078	789,440	621,655
%mapped	56	75	66

Table S4.1B. Splicing of cassette-exon significantly changed by Rbfox1 from RASL-seq.

event_ID	gene_symbol	Ctrl_mean	WT_mean	YS_mean	FA_mean	pval	FDR	group
6731	MELK	3.71	8.42	2.93	5.01	0.041340	0.045299	3
6729	MATR3	9.99	17.88	10.30	14.52	0.038940	0.043596	3
6703	RREB1	0.77	3.58	0.81	0.90	0.013520	0.026680	3
6684	IQCK	1.17	2.11	1.39	0.83	0.009125	0.021999	3
6654	MLL5	0.55	1.21	0.70	0.48	0.000478	0.008626	3
6587	PRUNE	1.33	2.85	1.99	1.52	0.035640	0.040956	3
6524	SETD5	0.11	0.36	0.23	0.09	0.005347	0.017767	3
6478	ZNF131	26.48	61.43	34.75	24.09	0.021136	0.032638	3
6417	hMLH1	0.24	0.82	0.45	0.21	0.033420	0.039687	3
6397	FSD1L	0.17	0.34	0.11	0.25	0.029057	0.036722	3
6388	C1orf58	22.09	37.15	19.49	24.40	0.007594	0.020056	3
6348	ZNF221	0.26	0.51	0.32	0.33	0.004045	0.015637	3
6231	PCM1	0.58	1.21	0.67	0.42	0.011661	0.025554	3
6004	DDB2	0.62	1.87	0.99	0.90	0.024293	0.034182	3
5933	MDM2	0.04	0.09	0.06	0.04	0.006571	0.019223	3
5928	HAUS2	7.85	12.88	12.31	6.55	0.004466	0.016385	3
5869	C14orf179	0.02	0.05	0.02	0.03	0.013215	0.026680	3
5868	C20orf72	2.19	5.74	2.57	3.09	0.023114	0.033531	3
5853	BC037884	0.30	0.70	0.47	0.43	0.003811	0.015394	3
5847	ZMYM1	0.26	0.42	0.36	0.18	0.012398	0.026359	3
5826	C5orf32	14.21	27.98	18.93	11.66	0.042437	0.045532	3
5796	HELLS	0.02	0.03	0.02	0.01	0.017316	0.029975	3
5778	CDKN2A	0.00	0.00	0.00	0.00	0.001427	0.010543	3
5601	UBE2F	0.01	0.06	0.05	0.01	0.045999	0.047617	3
5507	TANK	0.08	0.23	0.12	0.08	0.010964	0.024549	3
5501	HMBOX1	1.50	2.69	1.73	1.58	0.013867	0.026831	3
5458	ZNF384	0.04	0.19	0.05	0.05	0.001795	0.010877	3
5406	METT5D1	0.11	0.25	0.11	0.12	0.040843	0.044993	3

5306	CARS	60.99	203.06	85.91	45.26	0.006625	0.019223	3
5278	VAP	0.20	0.56	0.35	0.19	0.005087	0.017761	3
5129	LCLAT1	0.13	0.46	0.19	0.15	0.005826	0.018606	3
5064	CHD1L	55.28	128.00	97.00	61.00	0.042797	0.045679	3
3661	RAB6A	1.30	8.29	5.25	0.89	0.000486	0.008626	3
3646	FIP1L1	0.21	0.37	0.24	0.17	0.006936	0.019572	3
3645	FNBP1	0.03	0.11	0.07	0.03	0.003253	0.015216	3
3608	MARK3	0.09	3.34	0.90	0.12	0.042132	0.045455	3
3603	SENP6	1.90	3.41	2.08	1.46	0.021269	0.032638	3
3601	RBM5	0.30	0.68	0.49	0.35	0.003646	0.015394	3
3539	FHL1	0.04	0.11	0.03	0.04	0.045680	0.047525	3
3530	MICAL3	0.00	0.07	0.00	0.00	0.035221	0.040956	3
3500	TPIP	60.17	137.33	67.33	58.33	0.021864	0.032638	3
3469	FSTL4	0.39	0.69	0.50	0.33	0.019797	0.031879	3
3403	LOC400927	0.44	1.04	0.44	0.38	0.023000	0.033531	3
3394	RBM16	0.10	0.15	0.08	0.10	0.026894	0.035743	3
3359	PRKDC	20.73	37.22	25.49	29.25	0.040274	0.044666	3
3356	RORA	0.08	0.29	0.15	0.09	0.001087	0.010543	3
3333	KIF21A	0.03	0.72	0.14	0.04	0.000303	0.008560	3
3329	DIAPH1	0.08	1.52	0.79	0.10	0.019826	0.031879	3
3287	NGLY1	15.09	34.26	18.87	12.33	0.000798	0.010272	3
3228	EXT2	136.50	322.00	99.72	158.00	0.047159	0.048092	3
3189	NF2	1.64	11.87	8.21	1.18	0.026006	0.034787	3
3188	USP47	0.10	1.20	0.28	0.11	0.001537	0.010543	3
3168	MACF1	0.04	0.07	0.04	0.05	0.032579	0.039020	3
3121	USO1	0.02	0.06	0.03	0.03	0.012623	0.026359	3
3114	CSDE1	0.01	0.03	0.02	0.01	0.018988	0.031801	3
3055	LPHN2	0.01	0.03	0.01	0.01	0.003324	0.015216	3
2935	KIF23	0.86	1.46	1.03	0.67	0.019312	0.031879	3
2903	NOVA1	19.18	49.33	27.06	26.83	0.029751	0.037370	3
2881	MSI1	6.53	13.11	5.42	9.33	0.035526	0.040956	3
2864	GOLIM4	0.00	0.04	0.02	0.01	0.034650	0.040556	3
2843	EIF4G1	4.43	12.20	5.20	5.44	0.014751	0.026950	3
2806	RAD17	0.13	0.25	0.19	0.13	0.014500	0.026910	3
2793	CHTF8	0.74	1.60	1.13	0.52	0.014066	0.026831	3
2715	PDGFC	0.05	0.08	0.05	0.04	0.045469	0.047525	3
2660	BM009	1.03	1.71	1.45	0.76	0.032093	0.038889	3
2646	PBX3	0.82	1.29	1.05	0.80	0.032351	0.038972	3
2564	MARK2	0.92	1.90	1.23	0.94	0.012843	0.026359	3

2559	PHF20	0.02	0.05	0.03	0.02	0.006267	0.018985	3
2514	TPM3	0.16	0.47	0.23	0.23	0.031349	0.038439	3
2477	STRADA	0.20	0.65	0.26	0.18	0.004099	0.015637	3
2465	SMARCA2	4.68	7.43	5.30	5.51	0.002802	0.014429	3
2436	FIP1L1	2.57	4.13	3.55	3.14	0.004534	0.016385	3
2369	USP37	2.49	6.72	3.02	2.63	0.046850	0.048015	3
2329	PTBLP	0.13	0.58	0.21	0.09	0.045422	0.047525	3
2294	DCUN1D5	1.64	2.53	1.08	1.81	0.011436	0.025332	3
2065	CAST	0.01	0.05	0.03	0.01	0.003133	0.015060	3
2018	TJP2	6.82	13.71	11.97	5.19	0.023637	0.034050	3
1959	RBM27	0.15	0.33	0.24	0.15	0.000526	0.008626	3
1950	HMGN3	0.01	0.03	0.01	0.00	0.001419	0.010543	3
1878	MPRIP	0.00	0.05	0.04	0.00	0.003765	0.015394	3
1858	ST7	0.01	0.26	0.06	0.01	0.038239	0.043045	3
1766	NUMB	0.01	0.02	0.01	0.01	0.001137	0.010543	3
1733	NACA	0.00	0.01	0.00	0.00	0.001569	0.010543	3
1683	KIF21A	0.07	0.12	0.05	0.11	0.001470	0.010543	3
1603	PLCH1	0.01	0.28	0.02	0.01	0.021416	0.032638	3
1596	CASP8	0.12	0.43	0.26	0.18	0.005339	0.017767	3
1549	DKFZp566L241	0.34	1.40	0.76	0.38	0.000047	0.002408	3
1525	VEGFA	0.45	1.13	0.50	0.58	0.006001	0.018660	3
1513	TSC22D2	0.03	0.37	0.09	0.04	0.009364	0.022173	3
1483	LRRFIP2	0.06	0.23	0.10	0.04	0.024145	0.034182	3
1438	OSBPL9	1.37	3.35	2.51	1.49	0.013440	0.026680	3
1362	NF2	1.64	11.87	8.21	1.18	0.026006	0.034787	3
1333	SPTAN1	0.31	0.58	0.39	0.25	0.021441	0.032638	3
1314	DAZAP1	3.17	6.71	4.91	3.86	0.000632	0.008677	3
1311	TCERG1	0.49	1.66	0.96	0.52	0.002716	0.014348	3
1266	DPP8	0.19	0.34	0.19	0.17	0.010130	0.023240	3
1201	RSRC2	0.02	0.05	0.01	0.03	0.033669	0.039687	3
1185	BCAS3	0.06	0.13	0.07	0.06	0.004001	0.015637	3
1172	EIF4G1	2.74	4.72	3.70	1.87	0.025169	0.034386	3
1139	MYO6	0.00	0.04	0.01	0.01	0.001001	0.010543	3

1065	EIF4A2	0.39	0.79	0.66	0.51	0.004964	0.017631	3
1051	DST	15.34	33.70	18.59	13.30	0.001183	0.010543	3
1045	CDK5RAP2	0.45	0.78	0.56	0.36	0.022914	0.033531	3
1034	SNX14	0.37	0.75	0.55	0.36	0.008130	0.020935	3
1020	TSR1	7.58	19.76	8.33	8.50	0.001309	0.010543	3
1010	SLMAP	0.07	0.12	0.08	0.09	0.042145	0.045455	3
992	MTA1	0.58	1.06	0.89	0.57	0.015791	0.027803	3
876	EIF4H	0.09	0.44	0.22	0.06	0.002275	0.012667	3
843	EHBP1	4.09	6.48	6.17	6.11	0.035803	0.040956	3
785	CHTF8	0.74	1.60	1.13	0.52	0.014066	0.026831	3
767	RP5	0.43	1.99	0.79	0.32	0.001746	0.010877	3
743	PPFIBP1	0.04	0.11	0.02	0.05	0.043726	0.046431	3
730	IQCH	15.87	42.56	17.00	14.64	0.036767	0.041615	3
627	ARHGAP21	25.75	55.00	31.50	19.22	0.020335	0.031978	3
574	GOLGA2	0.01	0.13	0.05	0.02	0.000419	0.008626	3
504	BRD4	0.06	0.21	0.08	0.06	0.005814	0.018606	3
432	DNM1L	0.00	0.08	0.02	0.00	0.013599	0.026680	3
421	SORBS1	0.33	0.74	0.55	0.42	0.016466	0.028746	3
414	NCAPD2	11.24	19.00	10.75	14.00	0.028219	0.036706	3
400	DOCK9	0.04	0.20	0.08	0.04	0.022866	0.033531	3
388	BNIP2	0.03	0.27	0.11	0.03	0.014914	0.026950	3
333	UHRF2	82.74	171.67	108.00	102.33	0.015754	0.027803	3
325	TIAM1	11.82	35.68	14.24	10.75	0.024721	0.034386	3
266	MYO9A	0.01	0.03	0.01	0.00	0.017927	0.030520	3
265	JMJD6	0.03	0.09	0.05	0.04	0.005288	0.017767	3
162	ANGEL2	0.25	0.44	0.28	0.25	0.003143	0.015060	3
152	PLEKHM2	1.53	6.25	2.56	1.73	0.020118	0.031879	3
141	FUBP1	0.02	0.06	0.02	0.02	0.025947	0.034787	3
6907	TSC1	14.27	23.56	36.29	14.02	0.008517	0.021397	2
6504	FGFR1OP	2.63	5.27	4.43	2.14	0.001633	0.010543	2
6326	CSNK1G2	0.38	0.89	0.93	0.51	0.012924	0.026359	2

6067	WDR20	0.05	0.20	0.17	0.04	0.009139	0.021999	2
5860	KLC1	2.75	4.70	3.80	3.61	0.006564	0.019223	2
5751	RNF216L	0.32	0.85	0.92	0.36	0.028331	0.036706	2
5331	RPS24	0.58	2.43	2.83	0.74	0.049633	0.049633	2
3482	SRPK2	12.23	19.18	21.71	13.85	0.007430	0.020056	2
3475	CCDC88A	2.09	5.60	3.06	1.46	0.007373	0.020056	2
3397	IMAA	0.65	1.16	1.17	0.81	0.018902	0.031801	2
3220	UBQLN1	10.29	24.66	22.69	8.21	0.002517	0.013643	2
3154	RNPS1	1.53	2.66	2.54	1.07	0.014429	0.026910	2
3024	MAPT	0.20	0.48	0.35	0.18	0.008974	0.021999	2
3005	ZNF195	0.33	0.62	0.66	0.37	0.000025	0.002408	2
2964	CNOT10	24.76	100.07	88.18	22.88	0.017480	0.030007	2
2331	SLMAP	5.31	11.06	9.27	4.14	0.000128	0.004378	2
2314	PHF21A	1.56	3.15	2.68	1.28	0.031163	0.038439	2
2195	MBTD1	2.41	3.70	3.88	2.08	0.004469	0.016385	2
2179	SMAD9	0.41	0.90	0.64	0.31	0.006931	0.019572	2
2106	MAX	0.32	0.55	0.47	0.35	0.003599	0.015394	2
1998	ZNF507	0.16	0.26	0.24	0.16	0.009798	0.022936	2
1956	FOXM1	0.01	0.04	0.03	0.01	0.007986	0.020825	2
1917	MRPS18A	526.33	854.81	1068.83	362.48	0.024392	0.034182	2
1890	ST7	0.36	1.13	1.19	0.35	0.029011	0.036722	2
1875	GNAS	1.00	1.52	1.46	1.21	0.007536	0.020056	2
1796	MAP4K4	1.59	3.64	2.93	1.28	0.009184	0.021999	2
1744	EPB41	0.02	0.43	0.30	0.01	0.008432	0.021397	2
1701	ATXN2	6.64	22.49	29.89	4.87	0.000586	0.008626	2
1648	PROM1	0.04	0.16	0.13	0.03	0.001425	0.010543	2
1569	FBXW11	0.21	0.32	0.33	0.28	0.024943	0.034386	2
1300	CAMK2G	0.05	0.17	0.19	0.05	0.044241	0.046736	2
1268	RAD51L3	1.27	2.08	1.93	1.48	0.021607	0.032638	2
1252	ODF2	3.35	7.31	6.85	4.88	0.002877	0.014454	2
1074	BCLAF1	0.13	0.26	0.26	0.14	0.010432	0.023616	2
943	ATXN2	1.43	2.55	2.38	1.63	0.006069	0.018660	2
755	ROBO1	0.09	0.13	0.12	0.09	0.025205	0.034386	2

736	FGFR1OP2	1.89	4.53	4.08	2.55	0.012793	0.026359	2
672	KIAA0182	3.80	8.92	11.61	3.95	0.027351	0.036117	2
658	CIP29	0.31	3.63	3.02	0.46	0.015562	0.027803	2
550	UBE2L3	38.35	92.12	91.96	53.14	0.035985	0.040956	2
538	KLF6	0.01	0.04	0.03	0.01	0.049295	0.049535	2
531	C1orf149	0.10	0.17	0.16	0.14	0.028952	0.036722	2
454	FAM126B	0.18	1.67	1.49	0.21	0.046581	0.047979	2
407	UTRN	0.06	0.37	0.29	0.08	0.010153	0.023240	2
377	ECT2	0.25	1.51	1.14	0.27	0.033714	0.039687	2
373	ZRANB2	0.03	0.06	0.06	0.03	0.001319	0.010543	2
361	PDLIM7	0.09	0.22	0.23	0.07	0.019485	0.031879	2
329	HNRNPD	0.03	0.08	0.09	0.04	0.001254	0.010543	2
23	FMNL2	1.42	3.36	3.22	2.00	0.001955	0.011185	2
6605	ERCC2	14.55	8.70	12.51	11.52	0.019972	0.031879	1
6263	SRRM2	1.26	0.52	0.68	1.83	0.012609	0.026359	1
6185	DBF4B	2.97	1.67	1.75	2.02	0.000109	0.004378	1
5988	QTRTD1	4.27	1.42	2.01	4.82	0.003811	0.015394	1
5905	6-Sep	8.31	3.74	3.08	7.79	0.030452	0.037789	1
5622	CDC42SE2	9.91	2.78	1.23	7.68	0.012073	0.026179	1
5455	MASTL	0.18	0.09	0.09	0.15	0.014374	0.026910	1
5378	ING3	0.13	0.06	0.13	0.09	0.048149	0.048621	1
5356	KIAA1245	49.35	32.38	38.86	40.67	0.000045	0.002408	1
5261	NBPF10	49.35	32.40	39.01	40.67	0.000020	0.002408	1
5258	KIAA0723	0.10	0.05	0.11	0.13	0.001303	0.010543	1
5204	NBPF8	0.02	0.01	0.01	0.01	0.032078	0.038889	1
3716	LOC728855	21.68	3.55	3.69	18.30	0.040330	0.044666	1
3213	NUMB	1.13	0.36	0.83	1.15	0.001863	0.010965	1
2996	JAKMIP2	20.58	7.94	13.73	17.66	0.001638	0.010543	1
2917	MCL1	129.64	55.17	78.58	120.28	0.023946	0.034182	1

2697	ATP2B4	151.56	30.78	85.06	104.67	0.030308	0.037789	1
2379	RICTOR	0.54	0.33	0.30	0.45	0.000332	0.008560	1
2363	ADD3	62.36	13.15	20.57	49.04	0.027922	0.036636	1
2327	RAI14	1.47	0.71	0.94	1.50	0.005871	0.018606	1
2321	FAM184A	1.83	0.99	0.87	2.21	0.021837	0.032638	1
2174	BAZ2B	14.79	4.14	8.58	14.31	0.007453	0.020056	1
1933	STX2	15.97	8.83	8.86	16.98	0.019508	0.031879	1
1149	ING4	35.28	21.61	28.80	42.02	0.047610	0.048314	1
1120	PBRM1	6.82	1.55	1.49	5.87	0.041657	0.045403	1
652	ERBB2IP	2.19	1.05	0.93	2.03	0.000561	0.008626	1
472	EP400	3.73	2.48	3.38	3.89	0.028535	0.036722	1
254	BNIP1	0.06	0.03	0.05	0.05	0.003426	0.015343	1
10	ANK2	0.61	0.32	0.26	0.45	0.014841	0.026950	1

Table S4.1C. PCR Primers used in *in vivo* splicing assays.

minigene reporter	**Primer Name**	**Sequence**	Comment
Forward primer	DUP8	GACACCATGCATGGTGCACC	FAM labelled
Reverse primer	DUP3	AACAGCATCAGGAGTGGACAG	

endogenous exons	**Gene Name**	**Sequence**	**Goup**
Forward primer	PBRM1	TGTGATTAAGGCCCAACACC	1
Reverse primer		CTACCATAGGGGCCACTCCT	
Forward primer	FAM184A	AAAGGGCCCAAGACATTTTT	1
Reverse primer		ATGTGAAGTACCGGGCAAAC	
Forward primer	QTRTD1	CAGAACATCATGAAGTCTTGACAG	1
Reverse primer		ACTGGAACCAGTCTGGCTGA	
Forward primer	RAI14	GCAGGAATTCAAAGCCTTCT	1
Reverse primer		GAAGGGTGGTTCAGCAAAAA	
Forward primer	FAM126B	CGGACTGCAATTACAACAGC	2
Reverse primer		AGCCCCTGATGAAAATCCTT	
Forward primer	ST7	GCTACACAGCTGCTTTGCTC	2
Reverse primer		TTTTGGCACATGAGGATTGA	
Forward primer	RNF216L	TCACCAGAAACCAGTGGAAA	2
Reverse primer		CCTGGTGGTAATCGAGCAGT	
Forward primer	MARK3	CATGAAGCCACACCATTGTC	3
Reverse primer		CCCTCATATCTCCCGTTCCT	
Forward primer	USP47	AACCAACTGGTCCCGAAAG	3
Reverse primer		TCCGTTCATTCACTGTCTTTG	
Forward primer	KIF21A	TGGAAGGTCGACTCAAACAA	3
Reverse primer		TGGGCTGTTTAAAGGAGCAT	
Forward primer	RREB1	GATCACCTGTCCCCACTGTC	3
Reverse primer		GTCCCGTGAGGTGAGGTCTA	

CHAPTER 5

CONCLUDING REMARKS AND FUTURE DIRECTIONS

Concluding Remarks

Rbfox protein family is one of the important splicing regulators, which has been associated with many human diseases. Remarkable progress has been made towards understanding the role of Rbfox proteins since being discovered more than a decade ago. The RNA-binding properties have been well-characterized by SELEX experiments that identified the RNA sequence (U)GCAUG recognized by Rbfox RRM motif (Jin et al., 2003), and the structure of this motif bound to short RNA sequence was resolved by NMR (Auweter et al., 2006). Earlier work on a limited number of model exons and minigene reporters suggest that the Rbfox proteins could act as both splicing activators and repressors depending on their binding position relative to the alternative exon (Baraniak et al., 2006; Kuroyanagi et al., 2007; Mauger et al., 2008; Ponthier et al., 2006; Sun et al., 2012; Tang et al., 2009; Zhou et al., 2007; Zhou and Lou, 2008). Although much effort has been made to uncover the mechanism of Rbfox regulation, little is known about the underlying molecular mechanisms. Thanks to the recent modern high-throughput sequencing technology, numerous Rbfox-regulatory targets were identified in different tissues and cell lines (Weyn-Vanhentenryck et al., 2014; Yeo et al., 2009a). Rbfox proteins regulate many targets that are critical for proper development. The

transcriptome-wide binding profiles of Rbfox proteins revealed by CLIP-seq provided tremendous amount of information on the spatiotemporal binding of Rbfox proteins on their targets and RNA sequence specificities of Rbfox proteins. These new findings corroborate and extend our understanding of Rbfox proteins from earlier work. However, the molecular basis of how Rbfox proteins promote or suppress splicing in a position-dependent manner is poorly understood. This is the initial big question we asked. For long time, our lab and many others could not recapitulate the Rbfox-dependent splicing regulation in *in vitro* system using recombinant purified Rbfox protein and minigenes containing known Rbfox-dependent exons, a strategy that has been utilized successfully for many other splicing regulators. We hypothesized that some components critical for Rbfox-dependent splicing may be missing in the *in vitro* splicing reaction carried out with nuclear extract. Therefore, identifying protein partners of Rbfox proteins was the first step towards answering the big question of this dissertation.

To identify potential protein partners of Rbfox proteins, we started by carefully assessing the protein profiles of Rbfox proteins in the cell. Rbfox proteins have diverse variants showing tissue-specific expression and different localization. Particularly, the inclusion or exclusion of exon 19 of Rbfox1 results in two isoforms with different C-termini: FAPY and TALVP. The variant ending in FAPY is predominantly nuclear, while the TALVP isoform is localized to the cytoplasm (Lee et al., 2016; Lee et al., 2009). Since splicing occurs in the nucleus, we focused on the FAPY variant. Splicing could happen while the transcripts are still attached with chromatin, or after their release from chromatin.

The splicing regulatory complexes might be different in these two processes, thus we isolated and fractionated nuclei, and examined the protein profiles of Rbfox proteins in the obtained fractions. Strikingly, the Rbfox proteins were among the few splicing regulators we found residing predominantly in the high molecular (HMW) nuclear material containing the chromatin. Moreover, Rbfox proteins within the HMW fraction were engaged in a very large multi-protein complex, which we termed a Large Assembly of Splicing Regulators, LASR. The discovery of LASR and its potential role in Rbfox-dependent splicing regulation laid the foundation for this dissertation.

We characterized the protein content of the LASR complex described in Chapter 2, and found that it contains eight RNA/DNA-binding proteins with approximately equal stoichiometry: hnRNP M, hnRNP H, Matrin3, NF110, hnRNPU-like2, hnRNP C, DDX5, NF45. Many of these proteins were also implicated as splicing regulators in many other studies. The LASR complex has three distinct features compared with other known RNA-binding protein complexes. First, LASR is exclusively associated with the high molecular weight fraction material containing chromatin, despite the fact that all the protein subunits except Rbfox could be found in soluble nucleoplasm as well. Since LASR acts as a potential splicing regulatory protein complex, it is understandable that it is primarily associated with nascent transcripts that are attached to the chromatin, yet the actual factor controlling the recruitment of LASR is not defined. Second, LASR is an RNA-independent protein complex, because LASR was extensively treated with nucleases during purification. In contrast, many RNA-binding protein complexes are

dependent on RNA. Nevertheless, we found very short RNA fragments (~30nt) retained within LASR, which were presumably protected from nuclease digestion. Sequencing data from these RNA fragments suggested them to be the binding sites of LASR since they are enriched in sequence motifs recognized by the Rbfox proteins as well as by other RNA-binding members of LASR. Third, LASR sedimented around 55S in glycerol gradients. Among all the subunits of LASR, the Rbfox proteins are almost completely engaged in this complex, whereas all the other proteins are abundant in other gradient fractions, suggesting that these protein subunits also exist as free proteins, or participate in other protein complexes. However, the other subunits could form similar protein complexes sedimenting at ~ 55S in absence of Rbfox expression, indicating that the Rbfox proteins are not the organizers of this large complex. Taken together, the nuclear Rbfox proteins are associated with other RNA binding proteins within LASR and are recruited to intronic sequences on pre-mRNA.

The average weight the 55S LASR complex is expected to be in the megadalton, which is much larger than the sum of one copy of each of the subunits. Understanding this higher-order assembly of LASR, particularly the interactions between Rbfox and the other proteins, will help to understand Rbfox-dependent splicing regulation. We find that the C terminal region of Rbfox1, but not the N terminus or the RRM, is both necessary and required for interacting with LASR and assembling into higher-order complexes. The C terminus of Rbfox1 does not have any defined protein domain structure. Instead it is computationally predicted to have low complexity sequences. Although understanding

the role of low complexity sequences in RNA-binding proteins is still in its infancy, a burst of studies coming out in recent years suggest that these low complexity sequences and intrinsic disordered domains could polymerize to assemble into higher-order structures and play a role in protein functions (Kato et al., 2012; Kwon et al., 2013; Lin et al., 2015; Molliex et al., 2015; Patel et al., 2015). We find that LC sequence of Rbfox1 is required for the higher-order assembly of Rbfox1, but it is dispensable for interacting with LASR. Indeed, multiple interfaces were found in the C terminal part of Rbfox1. Repetitive tyrosine residues are distributed throughout the LC domain of Rbfox1, which are required in higher-order assembly of Rbfox1. These repetitive tyrosine residues match or closely resemble the tripeptide motif [G/S] Y [G/S] reported in other RNA binding proteins such as FUS (Kato et al., 2012). Due to the alternative splicing of Rbfox proteins, two naturally occurring tissue-specific mutually exclusive alternative exons B40 and M43 were located within the LC domain (Nakahata and Kawamoto, 2005). Interestingly, exons B40 and M43 within Rbfox1 and Rbfox2 encoded related but not identical sequences. The Rbfox2 M43 has three tyrosine residues altered, resulting in its incapability of supporting higher-order assembly.

We also examined the LC domain of Rbfox1 *in vitro* to see whether the high-order assembly of Rbfox1 can be related to amyloid-like aggregation like other RNA-binding proteins reported. Hydrogel formation and droplet assembly assays were performed with purified LC domain of Rbfox1 and its ten-tyrosine-to-serine mutant. The wild type LC domain of Rbfox1 formed droplets and fibrous structures that can be stained by Thioflavin

T. Thioflavin T is a dye that stains β-sheet structures (Khurana et al., 2005), suggesting that the structure formed by LC domain is likely to be β-sheets. Over time, the fibrous structures become more stable and form hydrogels. In contrast, the mutant protein stayed in the solution for days under the same experimental conditions. These results suggest that the tyrosine residues in the LC domain of Rbfox1 conferred these properties and established a link between fiber formation *in vitro* and higher-order assembly *in vivo*.

These structural findings motivated us to ask whether the interaction between Rbfox and LASR, as well as the higher-order assembly of the complexes, have a functional role in Rbfox-mediated alternative splicing regulation. As expected, the interaction between Rbfox and LASR is required in both splicing activation and repression, suggesting that Rbfox proteins regulate alternative splicing with the help of LASR. However, strikingly, differential requirements of higher-order assembly were observed in splicing activation and repression. Results from well-characterized Rbfox-regulated minigene reporters and large-scale splicing profiling by RASL-seq indicate that higher-order assembly is largely required in Rbfox-dependent splicing activation, while dispensable in splicing repression.

In summary, the primary goal of this dissertation is to understand the molecular mechanism of how Rbfox proteins regulate alternative splicing. The discovery of LASR led us to study the role of Rbfox proteins under the context of LASR. Indeed, results from this dissertation showed that Rbfox proteins regulate alternative splicing via its interaction with LASR. Differential requirements of higher-order assembly of Rbfox proteins for splicing activation and repression were revealed unexpectedly. In addition, we showed

that both higher-order assembly in vivo and fiber-like aggregation in vitro require repetitive tyrosine residues at the C terminus of Rbfox proteins, suggesting a possible mechanism of higher-order assembly by fiber-like aggregation.

Future directions

These results have provided a clearer picture of how Rbfox proteins work to inference splicing choice, but have left many interesting questions unanswered.

In this dissertation, we figured out the interface of Rbfox proteins with LASR, but the proteins directly interacting with Rbfox proteins and the interfaces on other proteins are unknown. Chemical crosslinking coupled with mass spectrometry allows identifying the interaction and interfaces of proteins, which can be applied on LASR. Information on the interaction between different subunits will also help understand the assembly of the complex and figure out the core components or the organizers of the complex. It will be of interest to determine the structure of LASR, although it will be a tough project to purify large quantity LASR from cells and obtain pure homogenous protein complexes for cryo-EM.

The higher-order assembly of LASR is required in splicing activation by Rbfox proteins. Altering the tyrosine residues in Rbfox proteins will impair the assembly of Rbfox with LASR. But how are the LASR assembly and disassembly regulated *in vivo*? One attractive hypothesis is that phosphorylation and dephosphorylation might regulate this

process. Indeed, we found the higher-order assembled Rbfox is hypophosphorylated (data not shown). However, whether the tyrosine phosphorylation of Rbfox proteins is the driving force of assembly and disassembly needs further investigation. When such regulatory activities are identified, we would be able to modulate the splicing activity of Rbfox proteins via these factors.

Many subunits of LASR act as splicing regulators individually. They recognize different RNA sequences, bind to different pre-mRNA target transcripts and regulate alternative splicing by distinct mechanisms. Thus, how would they coordinate to work as a team when they come together as one protein complex?

LASR is a beautiful representative of combinatorial regulation by many different splicing factors. We found short RNA fragments retained in LASR, which were sequenced to be the binding sites of many subunits in LASR (data not shown). Analyses of these sequences might be a good starting point to understand how the RNA binding and splicing regulation are orchestrated by LASR.

References

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

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

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R.*, et al.* (2012). The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science *338*, 1587-1593.

Barnby, G., Abbott, A., Sykes, N., Morris, A., Weeks, D.E., Mott, R., Lamb, J., Bailey, A.J., Monaco, A.P., and International Molecular Genetics Study of Autism, C. (2005). Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT. American journal of human genetics *76*, 950-966.

Berchowitz, L.E., Kabachinski, G., Walker, M.R., Carlile, T.M., Gilbert, W.V., Schwartz, T.U., and Amon, A. (2015). Regulated Formation of an Amyloid-like Translational Repressor Governs Gametogenesis. Cell.

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

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annual review of biochemistry *72*, 291-336.

Bouvet, P. (2001). Determination of nucleic acid recognition sequences by SELEX. Methods in molecular biology *148*, 603-610.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. Science *324*, 1729-1732.

Calabretta, S., and Richard, S. (2015). Emerging Roles of Disordered Sequences in RNA-Binding Proteins. Trends Biochem Sci *40*, 662-672.

Carreira-Rosario, A., Bhargava, V., Hillebrand, J., Kollipara, R.K., Ramaswami, M., and Buszczak, M. (2016). Repression of Pumilio Protein Expression by Rbfox1 Promotes Germ Cell Differentiation. Developmental cell *36*, 562-571.

Conchillo-Sole, O., de Groot, N.S., Aviles, F.X., Vendrell, J., Daura, X., and Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC bioinformatics *8*, 65.

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

Damianov, A., Ying, Y., Lin, C.H., Lee, J.A., Tran, D., Vashisht, A.A., Bahrami-Samani, E., Xing, Y., Martin, K.C., Wohlschlegel, J.A.*, et al.* (2016). Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. Cell *165*, 606-619.

Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C., Eckmann, C.R., Myong, S., and Brangwynne, C.P. (2015). The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. Proceedings of the National Academy of Sciences of the United States of America *112*, 7189-7194.

Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet *15*, 689-701.

Fukumura, K., Kato, A., Jin, Y., Ideue, T., Hirose, T., Kataoka, N., Fujiwara, T., Sakamoto, H., and Inoue, K. (2007). Tissue-specific splicing regulator Fox-1 induces exon skipping by interfering E complex formation on the downstream intron of human F1gamma gene. Nucleic acids research *35*, 5303-5311.

Fukumura, K., Taniguchi, I., Sakamoto, H., Ohno, M., and Inoue, K. (2009). U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. Nucleic acids research *37*, 1907-1914.

Gao, C., Ren, S., Lee, J.H., Qiu, J., Chapski, D.J., Rau, C.D., Zhou, Y., Abdellatif, M., Nakano, A., Vondriska, T.M.*, et al.* (2016). RBFox1-mediated RNA splicing regulates cardiac hypertrophy and heart failure. The Journal of clinical investigation *126*, 195-206.

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

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

Han, H., Irimia, M., Ross, P.J., Sung, H.K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N.*, et al.* (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature *498*, 241-245.

Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G.*, et al.* (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell *149*, 768-779.

Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., Uversky, V.N., Vidal, M., and Iakoucheva, L.M. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. Plos Comput Biol *2*, 890-901.

Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., and Parker, R. (2016). ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. Cell *164*, 487-498.

Jangi, M., Boutz, P.L., Paul, P., and Sharp, P.A. (2014). Rbfox2 controls autoregulation in RNA-binding protein networks. Genes & development *28*, 637-651.

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

Kalsotra, A., Xiao, X., Ward, A.J., Castle, J.C., Johnson, J.M., Burge, C.B., and Cooper, T.A. (2008). A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. Proceedings of the National Academy of Sciences of the United States of America *105*, 20333-20338.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J.*, et al.* (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753-767.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome research *12*, 656-664.

Khurana, R., Coleman, C., Ionescu-Zanetti, C., Carter, S.A., Krishna, V., Grover, R.K., Roy, R., and Singh, S. (2005). Mechanism of thioflavin T binding to amyloid fibrils. J Struct Biol *151*, 229-238.

Kim, K.K., Adelstein, R.S., and Kawamoto, S. (2009). Identification of neuronal nuclei (NeuN) as Fox-3, a new member of the Fox-1 gene family of splicing factors. The Journal of biological chemistry *284*, 31052-31061.

Kuroyanagi, H. (2009). Fox-1 family of RNA-binding proteins. Cellular and molecular life sciences : CMLS *66*, 3895-3907.

Kuroyanagi, H., Ohno, G., Mitani, S., and Hagiwara, M. (2007). The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo. Molecular and cellular biology *27*, 8612-8621.

Kuwasako, K., Takahashi, M., Unzai, S., Tsuda, K., Yoshikawa, S., He, F., Kobayashi, N., Guntert, P., Shirouzu, M., Ito, T.*, et al.* (2014). RBFOX and SUP-12 sandwich a G base to

cooperatively regulate tissue-specific splicing. Nature structural & molecular biology *21*, 778-786.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell *155*, 1049-1060.

Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S.L. (2014). Poly-dipeptides encoded by the C9ORF72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science.

Lee, J.A., Damianov, A., Lin, C.H., Fontes, M., Parikshak, N.N., Anderson, E.S., Geschwind, D.H., Black, D.L., and Martin, K.C. (2016). Cytoplasmic Rbfox1 Regulates the Expression of Synaptic and Autism-Related Genes. Neuron *89*, 113-128.

Lee, J.A., Tang, Z.Z., and Black, D.L. (2009). An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. Genes & development *23*, 2284-2293.

Li, H., Qiu, J., and Fu, X.D. (2012a). RASL-seq for massively parallel and quantitative analysis of gene expression. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *Chapter 4*, Unit 4 13 11-19.

Li, P.L., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F*., et al.* (2012b). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336-U129.

Lin, Y., Protter, D.S., Rosen, M.K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Molecular cell *60*, 208-219.

Lovci, M.T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T.Y., Stark, T.J., Gehman, L.T., Hoon, S*., et al.* (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. Nature structural & molecular biology *20*, 1434-1442.

Maekawa, S., Leigh, P.N., King, A., Jones, E., Steele, J.C., Bodi, I., Shaw, C.E., Hortobagyi, T., and Al-Sarraj, S. (2009). TDP-43 is consistently co-localized with ubiquitinated inclusions in sporadic and Guam amyotrophic lateral sclerosis but not in familial amyotrophic lateral sclerosis with and without SOD1 mutations. Neuropathology *29*, 672-683.

Martin, C.L., Duvall, J.A., Ilkin, Y., Simon, J.S., Arreaza, M.G., Wilkes, K., Alvarez-Retuerto, A., Whichello, A., Powell, C.M., Rao, K*., et al.* (2007). Cytogenetic and

molecular characterization of A2BP1/FOX1 as a candidate gene for autism. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics *144B*, 869-876.

Mauger, D.M., Lin, C., and Garcia-Blanco, M.A. (2008). hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. Molecular and cellular biology *28*, 5403-5419.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. Science *338*, 1593-1599.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. Cell *163*, 123-133.

Murakami, T., Qamar, S., Lin, J.Q., Schierle, G.S., Rees, E., Miyashita, A., Costa, A.R., Dodd, R.B., Chan, F.T., Michel, C.H.*, et al.* (2015). ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. Neuron *88*, 678-690.

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

Neelamraju, Y., Hashemikhabir, S., and Janga, S.C. (2015). The human RBPome: From genes and proteins to human disease. J Proteomics *127*, 61-70.

Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D.*, et al.* (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Molecular cell *57*, 936-947.

Nutter, C.A., Jaworski, E.A., Verma, S.K., Deshmukh, V., Wang, Q., Botvinnik, O.B., Lozano, M.J., Abass, I.J., Ijaz, T., Brasier, A.R.*, et al.* (2016). Dysregulation of RBFOX2 Is an Early Event in Cardiac Pathogenesis of Diabetes. Cell reports.

Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M.*, et al.* (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell *162*, 1066-1077.

Ponthier, J.L., Schluepen, C., Chen, W., Lersch, R.A., Gee, S.L., Hou, V.C., Lo, A.J., Short, S.A., Chasis, J.A., Winkelmann, J.C.*, et al.* (2006). Fox-2 splicing factor binds to a

conserved intron motif to promote inclusion of protein 4.1R alternative exon 16. The Journal of biological chemistry *281*, 12468-12474.

Schwartz, J.C., Wang, X., Podell, E.R., and Cech, T.R. (2013). RNA seeds higher-order assembly of FUS protein. Cell reports *5*, 918-925.

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

Singh, R., and Valcarcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. Nature structural & molecular biology *12*, 645-653.

Singh, R.K., Xia, Z., Bland, C.S., Kalsotra, A., Scavuzzo, M.A., Curk, T., Ule, J., Li, W., and Cooper, T.A. (2014). Rbfox2-coordinated alternative splicing of Mef2d and Rock2 controls myoblast fusion during myogenesis. Molecular cell *55*, 592-603.

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

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

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212-1215.

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

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P.*, et al.* (2009). Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. Science *323*, 1208-1211.

Venables, J.P., Brosseau, J.P., Gadea, G., Klinck, R., Prinos, P., Beaulieu, J.F., Lapointe, E., Durand, M., Thibault, P., Tremblay, K.*, et al.* (2013a). RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. Molecular and cellular biology *33*, 396-405.

Venables, J.P., Lapasset, L., Gadea, G., Fort, P., Klinck, R., Irimia, M., Vignal, E., Thibault, P., Prinos, P., Chabot, B.*, et al.* (2013b). MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. Nature communications *4*, 2480.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wang, J.T., Smith, J., Chen, B.C., Schmidt, H., Rasoloson, D., Paix, A., Lambrus, B.G., Calidas, D., Betzig, E., and Seydoux, G. (2014). Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically-disordered proteins in C-elegans. eLife *3*.

Wei, C., Qiu, J., Zhou, Y., Xue, Y., Hu, J., Ouyang, K., Banerjee, I., Zhang, C., Chen, B., Li, H.*, et al.* (2015). Repression of the Central Splicing Regulator RBFox2 Is Functionally Linked to Pressure Overload-Induced Heart Failure. Cell reports.

Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q.*, et al.* (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell reports *6*, 1139-1152.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009a). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nature structural & molecular biology *16*, 130-137.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009b). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nature structural & molecular biology *16*, 130-137.

Zhou, H.L., Baraniak, A.P., and Lou, H. (2007). Role for Fox-1/Fox-2 in mediating the neuronal pathway of calcitonin/calcitonin gene-related peptide alternative RNA processing. Molecular and cellular biology *27*, 830-841.

Zhou, H.L., and Lou, H. (2008). Repression of prespliceosome complex formation at two distinct steps by Fox-1/Fox-2 proteins. Molecular and cellular biology *28*, 5507-5516.