Title

Regulation of Alternative Pre-mRNA Splicing by a Heteromeric Complex of RNA Binding Proteins

Permalink

https://escholarship.org/uc/item/409992xd

Author

Peyda, Parham

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Regulation of Alternative Pre-mRNA Splicing by a Heteromeric Complex of RNA Binding

Proteins

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Molecular Biology

by

Parham Peyda

2024

ABSTRACT OF THE DISSERTATION

Regulation of Alternative Pre-mRNA Splicing by a Heteromeric Complex of RNA Binding
Proteins

by

Parham Peyda

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2024

Professor Douglas L. Black, Chair

The Rbfox proteins regulate essential splicing programs in various tissues. These
proteins have a conserved RNA binding domain that binds to the GCAUG element and
a C-terminal domain that binds to a large assembly of splicing regulators (LASR), a
heteromeric complex of RNA-binding proteins. The LASR subunits have varying
affinities for distinct RNA motifs, yet it is unclear how the Rbfox/LASR complex contacts
RNA and regulates specific exons. In chapter 2, we map the transcriptome-wide binding
sites of Rbfox1/LASR via a nuclease-protection assay. These sites contain
combinations of motifs for Rbfox and LASR subunits hnRNP M, hnRNP H/F, hnRNP C,
and Matrin3. These regions of RNA are adjacent to many cassette exons and individual
motifs within them contribute additively to exon activation. LASR influences the target
recognition of Rbfox by enabling it to activate exons through binding not only to GCAUG
elements but also to lower-affinity secondary motifs adjacent to LASR binding sites.

LASR bound to an RNA binding mutant Rbfox1 regulates additional exons. These results demonstrate that the Rbfox/LASR complex regulates splicing through multi-subunit recognition of cis-regulatory RNA modules, illustrating how splicing signals are decoded by combinatorial interactions between RNA-binding proteins.

In addition to contacting RNA, the Rbfox/LASR complex self-assembles into higher-order structures. This process is mediated by homo-oligomerization of a low complexity, tyrosine-rich region, called C2, in Rbfox's C-terminal domain. Self-assembly of Rbfox is essential for splicing activation of a subset of its targets. However, it remains unclear what parts or motifs within C2, aside from the tyrosine residues, promote oligomerization. In chapter 3, we develop two *in vitro* assays to investigate how the C2 region of Rbfox2 oligomerizes. We find that different parts of C2 can promote or inhibit oligomerization. Furthermore, clusters of differently spaced tyrosines in C2 have distinct effects on its self-assembly. These assays and findings can be useful for future explorations of Rbfox oligomerization.

The dissertation of Parham Peyda is approved.

Feng Guo

Kathrin Plath

James Akira Wohlschlegel

Xinshu Xiao

Douglas L. Black, Committee Chair

University of California, Los Angeles

2024

Dedication

To my grandmothers, Monavvar and Jamileh

For their perseverance, diligence, and love

Table of contents

List of Figures

List of Tables

Acknowledgements

This dissertation would have been impossible without the direct and indirect work and guidance of so many individuals that I cannot mention them all in this section; but I will try my best to give acknowledgements and thank the following individuals:

- Dr. Douglas Black: Throughout graduate school, the door to Doug's office was almost always open and whenever I walked in with a question – whether about a new immunoblot, a vague idea, a fellowship application, or any other issue – Doug listened patiently and provided thoughtful and insightful responses. He has been an ideal mentor, providing guidance but also giving me the freedom to explore ideas, experiment, fail, and grow as a scientist.

- The Black Lab: The members of the Black lab are distinguished by their kindness and each person helped me in various forms. Dr. Andrey Damianov taught me how to purify the LASR complex and the IP-seq protocol. Dr. Xiao-Jun Wang helped me get started in the lab and taught me protein expression and purification. Chia-Ho Lin helped with analysis of the RNA-seq data in chapter 2. Kelechi Onwuzurike worked tirelessly alongside me, assisting with cloning of constructs in chapter 2 and performing most of the biochemistry in chapter 3. Julia Nikolic, our lab manager, ensured the lab ran smoothly. Finally, Dr. Nazim Mohammed, Dr. Wen Xiao, and Daniel Arce have been wonderful lab mates and gave valuable advice.

- Thesis committee: Drs. Kathrin Plath, Grace Xiao, Feng Guo and James Wohlschlegel provided insightful and encouraging questions and discussions in our annual committee meetings.

Chapter 2 is modified from a manuscript in preparation:

Parham Peyda, Chia-Ho Lin, Kelechi Onwuzurike, Douglas Black. (2024). The Rbfox1/LASR complex controls alternative splicing by multi-subunit recognition of cis-regulatory RNA modules. *In preparation*.

Dr. Black and I conceptualized the project. Chia-Ho Lin and I analyzed the data. Kelechi Onwuzurike and I performed the experiments.

Chapter 3: Kelechi Onwuzurike purified the recombinant proteins and performed the co-immunoprecipitation assays. Aubrey Emmi, a rotation student in the lab, did the size-exclusion chromatography assays. Dr. Black and I conceptualized and supervised the work.

Education

University of California, Los Angeles   2012– 2016
Bachelor of Science (B.S.) in Biochemistry

Fellowships

Ruth L. Kirschstein Individual Predoctoral NRSA Fellowship (NIMH F30)   2022

Leon D and Leah E Rivenburg Fellowship   2018-2019

NIH Medical Scientist Training Program Fellowship (NIH T32)   2018

Publications

Damianov, A., Lin, C. H., Huang, J., Zhou, L., Jami-Alahmadi, Y., **Peyda, P**., Wohlschlegel, J., and Black, D. L. (2024). The splicing regulators RBM5 and RBM10 are subunits of the U2 snRNP engaged with intron branch sites on chromatin. Molecular Cell, 84(8), 1496–1511.e7.

Guo, J. K., Blanco, M. R., Walkup, W. G., 4th, Bonesteele, G., Urbinati, C. R., Banerjee, A. K., Chow, A., Ettlin, O., Strehle, M., **Peyda, P**., Amaya, E., Trinh, V., and Guttman, M. (2024). Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA in vivo. Molecular Cell, 84(7), 1271–1289.e12.

Choi, J.-S., Zhu, Y., Li, H., **Peyda, P**., Nguyen, T. T., Shen, M. Y., Yang, Y. M., Zhu, J., Liu, M., Lee, M. M., Sun, S.-S., Yang, Y., Yu, H.-H., Chen, K., Chuang, G. S., and Tseng, H.-R. (2017) Cross-Linked Fluorescent Supramolecular Nanoparticles as Finite Tattoo Pigments with Controllable Intradermal Retention Times. ACS Nano, 11.1, 153–162.

Hou, S., Choi, J.-s., Chen, K.-J., Zhang, Y., Peng, J., Garcia, M. A., Yu, J.-H., Thakore-Shah, K., Ro, T., Chen, J.-F., **Peyda, P.**, Fan, G., Pyle, A. D., Wang, H. and Tseng, H.-R. (2015), "Supramolecular Nanosubstrate-Mediated Delivery for Reprogramming and Transdifferentiation of Mammalian Cells. Small, 11: 2499–2504.

Peng, J., Garcia, M. A., Choi, J. S., Zhao, L., Chen, K. J., Bernstein, J. R., **Peyda, P**., Hsiao, Y. S., Liu, K. W., Lin, W. Y., Pyle, A. D., Wang, H., Hou, S., Tseng, H. R. (2014). "Molecular Recognition Enables Nanosubstrate-Mediated Delivery of Gene-Encapsulated Nanoparticles ACS Nano, 8, 4621– 4629

Talks

Peyda, P., Lin, C.-H., Onwuzurike, K., Damianov, A., & Black, D. (2023). Combinatorial Recognition of Splicing Regulatory Motifs by a Heteromeric Assembly of RNA Binding Proteins. Cold Spring Harbor meeting on Eukaryotic mRNA Processing, Cold Spring Harbor, NY.

Posters

Peyda, P., Onwuzurike, K., and Black, D. (2023). Splicing Regulation Through Combinatorial Recognition of Cis-Regulatory RNA Modules. RNA Annual Symposium of UC Riverside, Riverside, CA.
   - Best poster award

Peyda, P., Onwuzurike, K., and Black, D. (2023). Splicing Regulation Through Combinatorial Recognition of Cis-Regulatory RNA Modules. UCLA MBI Annual Retreat and Research Conference, Los Angeles, CA.

Peyda, P., Onwuzurike, K., and Black, D. (2023). Splicing Regulation Through Combinatorial Recognition of Cis-Regulatory RNA Modules. UCLA-Caltech Medical Scientist Training Program Annual Conference. Los Angeles, CA, Los Angeles, CA.

Peyda, P., Onwuzurike, K., and Black, D. (2023). Splicing Regulation Through Combinatorial Recognition of Cis-Regulatory RNA Modules. UCLA MBI Annual Retreat and Research Conference, Los Angeles, CA.

Chapter 1: Introduction to Alternative pre-mRNA Splicing, Rbfox, and a Large

Assembly of Splicing Regulators

Catalysis of RNA splicing by the spliceosome

      Most organisms use DNA as their genetic material; however, there is great variation and complexity as to how this information gets expressed in different cells. In prokaryotes, DNA is transcribed to messenger RNA (mRNA) and translated to proteins in the same cellular compartment. In contrast, eukaryotic transcription occurs in the nucleus and is uncoupled from translation which occurs in the cytoplasm. This spatial segregation enables pre-processing steps that eukaryotic pre-mRNA molecules undergo, including the addition of a 5' 7-methylguanosine CAP and a 3' poly-A tail, prior to being translated as mRNAs (Hocine et al., 2010). In addition, most eukaryotic protein-coding genes are split into expressed regions (exons), which contain coding sequences, and intergenic regions (introns), which are mostly non-coding (Gilbert, 1978). Thus, prior to being exported to the cytoplasm, most eukaryotic pre-mRNAs undergo splicing, a process by which introns get excised out and exons ligated together (Sharp, 1994).

      RNA splicing of pre-mRNA in eukaryotes occurs as two trans-esterification reactions where first a 2'-OH group from the branch point adenosine attacks the phosphate at the 5' splice site, forming a lariat structure (Padgett et al., 1984; Ruskin et al., 1984). This is followed by the 3'-OH group of the upstream exon attacking the phosphate at the 3' splice site, leading to the exon-exon ligation and release of the intron lariat. In higher eukaryotes, this reaction is mostly catalyzed by the major spliceosome, a dynamic RNA-protein complex consisting of five ribonucleoprotein subunits, U1, U2, U4, U5, U6 and other protein co-factors (Wilkinson et al., 2020). The catalytic core of this spliceosome consists of the U2 and U6 snRNAs that coordinate two magnesium cations (Steitz & Steitz, 1993). In addition, splicing can also occur in a

self-catalytic manner as seen with Group I and II introns. Group I introns use a guanosine cofactor for the initial nucleophilic attack on the phosphate at the 5' splice site and thus do not generate an intron lariat (Cech, 1990). In group II introns, the adenosine in a branch point attacks the 5' splice site to form a lariat, and allows the 3'OH of the exon to 5' exon to attack the phosphate at the 3' splice site and therefore join the exons (Pyle, 2016). RNA splicing either self-catalyzed or by the spliceosome, therefore, is an example of RNA processing by RNA.

The spliceosome recognizes splice sites and assembles in a stepwise fashion (Wan et al., 2020; Wilkinson et al., 2020). The 5' end of U1 snRNA recognizes the 5' splice site by base pairing with a consensus sequence: 5'-GUAUGU in yeast and 5'-GURAG in humans. The 3' splice site contains a 5'-YAG sequence followed by an upstream branch site that contains the nucleophilic adenosine: UACUAAC in yeast and YNYURAY in humans. In metazoans, there is an additional polypyrimidine tract between the branchpoint sequence and the 3' splice site. In these organisms, the 3' splice site is recognized by the U2AF65–U2AF35 heterodimer and the branchpoint sequence is recognized by the SF1/mBBP. Binding of U1 to the 5' splice site, SF1/mBBP to the branchpoint, and U2AF65–U2AF35 to the 3' splice site constitutes the E complex (Wan et al., 2020; Wilkinson et al., 2020).

After the formation of the E complex, SF1 and U2AF are displaced through the action of the DEAD-box helicases Prp5 and Sub2 and the U2 snRNP is recruited to the BP sequence, where the U2 snRNA basepairs with the branchpoint (Wilkinson et al., 2020). This complex containing U1 and U2 is the A complex. In higher eukaryotes, the formation of the E and A complexes are the major determinants for choice of splice

sites; after these steps an intron is usually committed to being removed. After the formation of the A complex, the U4/U6·U5 tri-snRNP are recruited to form the pre-B complex. The spliceosome then undergoes a series of rearrangements, largely driven by action of helicases, to catalyze the splicing reaction and eventually disassembles.

The 5' splice site, branchpoint sequence, and 3' splice site comprise the core splicing signals. These splicing signals are highly conserved in yeast. But these sequences are mostly degenerate in metazoans, making the recognition of bona fide splice site challenging for the spliceosome (Y. Lee & Rio, 2015). So how are exons recognized in higher eukaryotes?

Recognition of exons in lower versus higher eukaryotes

Recognition of exons in higher eukaryotes is not as straightforward as their recognition in lower eukaryotes. It is estimated that only half the information required to define exon-intron boundaries are contained within human splice site signals (Wang & Burge, 2008). Furthermore, in higher eukaryotes exons are often much shorter than introns, with the median human exon length being 131 bp and the median intron length being 1747 bp (Piovesan et al., 2019). These large introns often contain decoy splice sites which can occur in pairs to constitute pseudo-exons that appear like bona fide exons in length and strength but are not spliced (Wang & Burge, 2008). One factor that influences the recognition of exons in these organisms is their genetic architecture.

In lower eukaryotes introns are often short. Therefore, exons can be recognized when the spliceosome assembles by U1 binding to the 5' splice site and U2AF binding to the downstream 3' splice site. This process is called intron definition (De Conti et al., 2013). However, in higher eukaryotes the downstream 3' splice site can lie hundreds to

thousands of nucleotides away from the 5' splice site of the exon. In these organisms, it is thought that the exon is recognized by U1 binding to the 5' splice site and U2AF binding to the upstream 3' splice site across the exon through the process of exon definition (De Conti et al., 2013; Robberson et al., 1990). The distance between the 3' splice site and 5' splice site across an exon can affect splicing where shortening the exon can lead to steric hindrance and prevention of formation of the exon complex and lengthening an exon above 300 nucleotides can lead to activation of cryptic splice sites within the exon (Black, 1991; De Conti et al., 2013). Therefore, the genetic architecture of eukaryotes plays an important role in distinguishing exons from pseudo-exons.

In addition to genetic architecture, cis-acting splicing regulators elements (SREs) play an important role in control of splice site choice in higher eukaryotes. SREs are comprised of exonic splicing enhancers (ESE), exonic splicing silencers (ESS), intronic splicing enhancers (ISE), and intronic splicing silencers (ISS) (Black, 2003). Trans-acting RNA binding proteins (RBP) bind to these elements and influence the choice of splice site. These proteins can generally be divided into three families of SR proteins, hnRNP proteins, and tissue-specific splicing regulators such as PTBP, NOVA, and Rbfox (Fu & Ares, 2014; Y. Lee & Rio, 2015).

The SR family of proteins play an important role in recognition of splice sites. These proteins bind to purine rich ESEs in exons and use their RS domains to recruit U1 snRNP to the 5' splice site and U2AF to the 3' splice site (Chen & Manley, 2009). In contrast, hnRNP proteins generally inhibit the exclusion of exons. For example, hnRNP C competes with U2AF65 for binding to splice sites at Alu exons and suppresses the exonization of many of these elements in the transcriptome (Zarnack et al., 2013).

5

Another example among the hnRNP proteins is the polypyrimidine tract binding (PTB) family. PTB binds to polypyrimidine sequences in intronic regions, inhibiting spliceosome assembly by interfering with 3′ splice site recognition in addition to an interaction with U1 snRNA (Fu & Ares, 2014). However, it is important to note that these general rules for SR proteins and hnRNP P proteins are not always true; these proteins can have different effects depending on the context of where they bind.

Regulation of alternative splicing

The high degree of conservation of splice site sequences in yeast is correlated with the fact that most exons in yeast are constitutive, meaning the same set of exons will be recognized by the spliceosome and spliced together (Y. Lee & Rio, 2015). In contrast, higher eukaryotes display considerable variability in the degeneracy of core splicing signals. This variability results in exons with a range of splice site strengths: some adhere closely to the consensus sequence, indicating high strength, while others diverge, indicating lower strength. Such variations in splice site strength contribute to alternative splicing in higher eukaryotes, where exons may be included or excluded from the final transcript in a context-dependent manner (Fu & Ares, 2014). Alternative exons generally have splice sites that are weaker compared to the splice sites of constitutive exons, allowing for control of their splicing by other factors.

In humans, alternative splicing is the rule rather than the exception for most genes, with an estimate 95% of genes being able to undergo alternative splicing (Barbosa-Morais et al., 2012; Pan et al., 2008). This process is now widely recognized to be essential in many physiological and diseases processes. The function of alternative splicing is likely multifaceted. Through inclusion of different exons, multiple

proteins can be made from the same transcript, thereby increasing the complexity of the genetic information in higher eukaryotes. In accordance, humans do not possess markedly more genes than simpler organisms such as worms or fruit flies. However, humans exhibit a higher rate of alternative splicing, correlating with their increased complexity (Y. Lee & Rio, 2015). Besides altering the amino acid content of the final translated protein, alternative splicing can also modulate the level of gene expression through nonsense mediated decay by inclusion or exclusion of exons with premature stop codons (Maquat, 2004).

Alternative splicing occurs in many different forms (Black, 2003). The most common form is inclusion or exclusion of cassette exons. In addition, alternative splicing can also result in a pair of mutually exclusive exons where two pairs of nearby exons are spliced so that only one can end up in the final transcript. Furthermore, there can be usage of alternative 5' or 3' splice sites that can lengthen or shorten an exon. In addition, through usage of alternative promoters or poly(A) sites, transcripts can alter their 5' and 3' ends, respectively. There is also intron retention, where an intron gets retained in the final transcript.

The regulation of alternative splicing is multifaceted. An important form of regulation occurs through the interaction of RBPs with SREs, as discussed previously for the choice of splice site. In alternative splicing, the SR, hnRNP, and tissue-specific splicing regulators once again become the major players. In addition to regulation by SREs and RBPs, the choice of splice site can be affected by RNA secondary structure. For example, in the DSCAM transcript a constitutive exon contains a docking site that can hybridize to selector sequences that are upstream of a group of alternatively spliced

exons in a mutually exclusive manner (Graveley, 2005). This mechanism ensures only one exon from the group of alternatively spliced exons is included in the final transcript. In addition to RNA structure, transcription can affect alternative splicing through at least two modes: 1) CTD of Pol II can recruit splicing factors and influence the inclusion of nearby exons and 2) kinetic coupling of transcription and splicing can lead to slower elongation promoting inclusion of some exons (Kornblihtt et al., 2013). Regulation of alternative splicing, therefore, can occur in a highly context-dependent manner and affected by many different processes in the cell.

The RNA-binding Fox family of proteins

The RNA Binding Fox family of proteins (Rbfox) are conserved from *c. elegans* to humans (Conboy, 2017). All Rbfox proteins share a conserved RNA-recognition motif (RRM) that recognizes a UGCAUG motif and its close variants (Conboy, 2017; Jin et al., 2003; Underwood et al., 2005). This family of proteins, consisting of Rbfox1, Rbfox2, and Rbfox3 in mammals have specific patterns of expression. Rbfox1 is mostly expressed in the brain, heart, and muscle. Rbfox2 is ubiquitously expressed. Rbfox3, also known as NeuN, is exclusively expressed in the brain (Conboy, 2017; Kim et al., 2009; Kuroyanagi, 2009).

The Rbfox genes also undergo complex patterns of splicing and have different isoforms. Nuclear isoforms of Rbfox are generally involved in regulating splicing whereas cytoplasmic isoforms can bind in 3' UTR sequences to regulate translation (Carreira-Rosario et al., 2016; J.-A. Lee et al., 2016) . In addition, an exon in Rbfox's RRM is repressed through an autoregulatory mechanism and leading to an isoform that lacks the ability to bind to RNA and exerts a dominant negative effect on the wildtype

protein (Damianov & Black, 2010). Rbfox also has two mutually exclusive exons in its C-terminal domain. The B40 exon is expressed ubiquitously and the M43 exon is expressed exclusively in muscle (Nakahata & Kawamoto, 2005).

The Rbfox proteins play a crucial role in the nervous system. Rbfox1 deletion in the central nervous system of mice heightens their seizure susceptibility, while deletion of Rbfox2 impairs cerebellum development (Gehman et al., 2011, 2012). Triple knockout of the Rbfox proteins in ventral spinal neurons leads to immature electrophysiological activity and diminishment of the axon initial segment (Jacko et al., 2018). Human mutations in RBFOX1 and RBFOX3 have been associated with familial epileptic disorders (Lal et al., 2015; Lal, Reinthaler, et al., 2013; Lal, Trucks, et al., 2013). Copy number variations at the RBFOX1 locus are associated with autism spectrum disorder (ASD) and transcriptomic analysis of ASD patient brains showed changes in Rbfox regulated alternative splicing events (Sebat et al., 2007; Voineagu et al., 2011). There is also an association between variants in the RBFOX1 locus and an increased amyloid burden in the pre-clinical phase of Alzheimer's disease (Raghavan et al., 2020).

In addition, the Rbfox proteins are also important for cardiac development and function. Multiple mutations in Rbfox2 have been linked to hypoplastic left heart syndrome (HLHS) (Homsy et al., 2015). These mutations can create a truncated Rbfox2, which alters its subcellular distribution and results in dysregulation of a set of transcripts by binding to their 3' UTR sites (Verma et al., 2016). Furthermore, knockout of Rbfox2 in mouse and zebrafish hearts can recapitulate many features of HLHS (M. Huang et al., 2022; Verma et al., 2022). In addition, transverse aortic constriction can

lead to a reduction in Rbfox2 expression and subsequent splicing changes in many genes involved in cardiac function and pathology (Wei et al., 2015). Furthermore, Rbfox1 expression is lowered in failing human and mouse hearts and induction of its expression in mouse pressure overload models can alleviate some pathologies of the disease (Gao et al., 2016).

Rbfox also plays an important role in metabolism. In the liver, Rbfox2 is dysregulated under diet-induced obesity (Paterson et al., 2022). This dysregulation leads to changes in many physiologically relevant targets including the Scarb1 gene which then leads to an alteration in lipid homeostasis. Rbfox2 is also dysregulated in diabetes. In the pancreas, Rbfox2 regulates several genes involved in insulin secretion and Rbfox2 knockout mice have insulin secretion defects (Moss et al., 2023). Furthermore, dysregulation of Rbfox2 in diabetes leads to expression changes in the heart linked to cardiac pathology (Nutter et al., 2016).

Several studies have analyzed Rbfox's RNA binding properties. An *in vitro* selection assay initially found that Rbfox binds the GCAUG pentamer (Jin et al., 2003). Since this study many more comprehensive *in vitro* binding assays of Rbfox have been carried out and show that Rbfox binds to (U)GCAUG with extremely high affinity and can also bind to a few secondary motifs with lower affinity (Begg et al., 2020; Ye et al., 2023). Rbfox is unusual in its high specificity for GCAUG compared to other RNA binding proteins that generally bind to degenerate motifs. An NMR solution structure of Rbfox bound to UGCAUGU shows that this protein's RNA binding domain has the canonical RNA Recognition Motif fold of $\beta_1\alpha_2\beta_3\beta_4\alpha_5\beta_6$ (Auweter et al., 2006). The last three UGU nucleotides are recognized by the β sheet in this structure, a pattern also

observed in structures of RRMs of other RBPs. However, the first three UGC

nucleotides contact a single phenylalanine in the β1α1-loop that is outside the β sheet.

In addition, the G2 and A4 nucleotides form a base-pair. This unusual mode of binding

to RNA by Rbfox is thought to define its high specificity for the GCAUG element.

Rbfox's RNA binding has also been studied *in vivo* in several crosslinking-

immunoprecipitation (CLIP) studies. In CLIP, cells are subjected to ultraviolet radiation

that creates covalent bonds between proteins that contact RNA. After this crosslinking,

cells are lysed and a protein of interest is purified via denaturing conditions; its binding

sites are then determined by sequencing co-purified RNA (Hafner et al., 2021). The first

CLIP of Rbfox was done on Rbfox2 in mouse embryonic stem cells (mESCs) (Yeo et

al., 2009). This study showed that, in general, Rbfox2 activates splicing when bound to

GCAUGs downstream of a cassette exon and represses a cassette exon when bound

to GCAUGs in its upstream intron. Another CLIP study of Rbfox2 in mESCs revealed

that this protein controls the splicing of exons with premature stop codons in many

RBPs and changes their expression levels through this mechanism, contributing to the

establishment of an autoregulatory network of splicing (Jangi et al., 2014). CLIP of

Rbfox2 in HEK293 revealed that many of Rbfox's binding sites can be more than 500

nucleotides away from its target exons and exert their effect by forming RNA-RNA

basepairs with sites more proximal to the target exon through formation of an RNA

bridge (Lovci et al., 2013). The binding sites of Rbfox1/2/3 were also mapped out in the

mouse brain (Damianov et al., 2016; Weyn-Vanhentenryck et al., 2014). These

genome-wide studies identified *in vivo* targets of Rbfox in different biological contexts

and demonstrated general principles of Rbfox's regulatory functions. Yet, there is a

more limited understanding of how Rbfox's binding to a specific regulatory element leads to its downstream functional consequences.

There are reports of splicing repression by Rbfox through competition for binding sites of snRNP subunits. Rbfox1 and Rbfox2 can repress the inclusion of exon 4 of CGRP by binding to a UGCAUG elements in the upstream intron to inhibit binding of SF1 and block formation of the E' complex. Additionally, these proteins bind to a UGCAUG in the cassette exon and prevent the transition from E' to E complex for transcripts that failed to be regulated at the first step (Zhou et al., 2007; Zhou & Lou, 2008). Rbfox proteins also prevent the formation of minor spliceosome E complex by binding to a UGCAUG element upstream of exon 9 of the hF1γ transcript and repressing the inclusion of this exon (Fukumura et al., 2007, 2009). These studies demonstrated that Rbfox utilizes various strategies for splicing repression, including direct competition for binding to splice sites.

But besides binding to RNA through its RRM, Rbfox contains additional domains which can engage in protein-protein interactions and affect Rbfox's binding and activity. Rbfox has a highly negatively charged N-terminal domain (NTD) and a low complexity tyrosine-rich CTD (Conboy, 2017). MS2 tethering assays indicate that the CTD is sufficient for splicing activation when tethered downstream of an exon whereas both the CTD and the RRM are needed to repress splicing when tethered to upstream of an exon (Sun et al., 2012). This CTD can interact with hnRNP H1 which can affect Rbfox's splicing activity. In addition, Rbfox2 also enhances splicing repression by hnRNP H/F on exon IIIc of fibroblast growth factor receptor 2 (Mauger et al., 2008). Rbfox2 also activates the exon 16 of protein 4.1R by binding to a UGCAUG element downstream of

the exon and recruiting the U1 snRNP through direct interaction of Rbfox's CTD with the zinc finger region of U1C (S.-C. Huang et al., 2012). This study is the only report of a direct interaction of Rbfox with a subunit of a snRNP.

Together these studies demonstrated that the Rbfox proteins bind to GCAUGs and employ a diverse set of strategies through protein-protein interactions or competition for splice sites of other RNA binding proteins to regulate their targets. However, these studies were complicated by the observation that only a small fraction of expressed GCAUGs crosslinked to Rbfox (Begg et al., 2020). Furthermore, Rbfox crosslinked to sites that lacked this motif or its related secondary motifs. In addition, Rbfox's function was difficult to assess *in vitro* with splicing extracts, a strategy that is used to study the activity of many other regulatory RNA binding proteins (Ying, 2016). These observations and previously reported interactions of Rbfox with protein co-factors led the Black lab to investigate whether there are additional co-factors of Rbfox that can affect its RNA binding and splicing activity.

A large assembly of splicing regulators

To investigate the protein partners of Rbfox, the Black lab utilized a new biochemical approach. They found that after nuclear lysis many RNA binding proteins remain associated with an insoluble pellet that contains chromatin. To examine the proteins in this fraction, they treated this pellet with nucleases that digest DNA and RNA to solubilize proteins associated with chromatin. Using this method, the Black lab found that most of Rbfox in the nucleus is this insoluble pellet and bound to a large assembly of (LASR) (Damianov et al., 2016). The members of this protein complex were hnRNP M, hnRNP H/F, hnRNP C, Matrin3, NF110, NF45, and DDX5. The LASR subunits were

13

present in near stoichiometric amounts to Rbfox and Rbfox bound to them through its CTD (Ying et al., 2017). A more subsequent study found that MeCP2 can also be part of LASR (Jiang et al., 2021).

The presence of Rbfox in LASR partly explained its selectivity for particular GCAUG elements and its ability to regulate targets without this motif. GCAUG motifs that crosslinked to Rbfox were enriched for adjacent sequences rich in GU-rich elements, motifs predicted to bind to the LASR subunit hnRNP M, suggesting cooperative binding to these sites (Damianov et al., 2016). Furthermore, splicing repression by hnRNP M was stimulated by Rbfox and depletion of hnRNP M affected the regulation of a subset of Rbfox's targets. In addition, some of the sites that Rbfox crosslinks to that do not contain GCAUG or related secondary motifs, contain motifs for LASR subunit hnRNP C and hnRNP M, suggesting that perhaps Rbfox crosslinks to these sites by virtue of being proximal to them through being recruited there by these LASR subunits.

The association of Rbfox with LASR raises several questions: 1) do segments of RNA get contacted by Rbfox and other LASR subunits? 2) how does this complex of multiple RNA binding proteins that have different specifies for different RNA motifs cooperate to recognize their targets in the transcriptome? 3) what are the regulatory targets of the Rbfox/LASR complex and how does binding of the complex to these sites lead to their regulation? In chapter 2 of this dissertation, we show that there are nuclease-protected RNAs present in our purifications containing Rbfox1/LASR complexes. These RNAs helped us identify sites that are contacted by Rbfox and multiple LASR subunits and understand how this complex gets recruited to its targets.

We also analyzed splicing regulation of a wildtype and an RNA binding mutant Rbfox. This analysis showed that besides regulating targets that contain nearby GCAUG elements, Rbfox also gets recruited to sites lacking this motif via LASR. Moreover, Rbfox restricts LASR's regulatory effects; LASR that is bound to an RNA binding mutant Rbfox binds to an additional set of sites and regulates nearby cassette exons.

The Rbfox/LASR complex sedimented around 45S, a size which is larger than expected for a unit complex. It was determined that besides binding to LASR, Rbfox's CTD can also oligomerize and form higher-order Rbfox/LASR complexes (Ying et al., 2017). Within the CTD, the C2 region contained 10 tyrosine residues that are essential for the oligomerization of Rbfox. Furthermore, these tyrosine residues were important for Rbfox's ability to activate a subset of its targets. However, it is not clear if there are other specific regions or motifs within C2 that drive Rbfox's oligomerization and contribute to its splicing activity. In chapter 3, we use size-exclusion assay and co-immunoprecipitation experiments and find that the first potion of C2 drives its oligomerization whereas the second portion of the C2 might have an inhibitory effect on the oligomerization. Furthermore, we find that different clusters of tyrosines within the C2 have distinct effects on the self-assembly of this region.

References

Auweter, S. D., Fasan, R., Reymond, L., Underwood, J. G., Black, D. L., Pitsch, S., & Allain, F. H. (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *The EMBO Journal*, *25*(1), 163–173. https://doi.org/10.1038/sj.emboj.7600918

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., Slobodeniuc, V., Kutter, C., Watt, S., Çolak, R., Kim, T., Misquitta-Ali, C. M., Wilson, M. D., Kim, P. M., Odom, D. T., Frey, B. J., & Blencowe, B. J. (2012). The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. *Science*, *338*(6114), 1587–1593. https://doi.org/10.1126/science.1230612

Begg, B. E., Jens, M., Wang, P. Y., Minor, C. M., & Burge, C. B. (2020). Concentration-dependent splicing is enabled by Rbfox motifs of intermediate affinity. *Nature Structural & Molecular Biology*, *27*(10), 901–912. https://doi.org/10.1038/s41594-020-0475-8

Black, D. L. (1991). Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? *Genes & Development*, *5*(3), 389–402. https://doi.org/10.1101/gad.5.3.389

Black, D. L. (2003). Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry*, *72*(1), 291–336. https://doi.org/10.1146/annurev.biochem.72.121801.161720

Carreira-Rosario, A., Bhargava, V., Hillebrand, J., Kollipara, R. K., Ramaswami, M., & Buszczak, M. (2016). Repression of Pumilio protein expression by Rbfox1 promotes germ cell differentiation. *Developmental Cell*, *36*(5), 562–571. https://doi.org/10.1016/j.devcel.2016.02.010

Cech, T. R. (n.d.). *SELF-SPLICING OF GROUP I INTR.ONS*.

Chen, M., & Manley, J. L. (2009). Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. *Nature Reviews Molecular Cell Biology*, *10*(11), 741–754. https://doi.org/10.1038/nrm2777

Conboy, J. G. (2017). Developmental regulation of RNA processing by Rbfox proteins. *WIREs RNA*, *8*(2), e1398. https://doi.org/10.1002/wrna.1398

Damianov, A., & Black, D. L. (2010). Autoregulation of Fox protein expression to produce dominant negative splicing factors. *RNA*, *16*(2), 405–416. https://doi.org/10.1261/rna.1838210

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., & Black, D. L. (2016). Rbfox Proteins

Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell*, *165*(3), 606–619. https://doi.org/10.1016/j.cell.2016.03.040

De Conti, L., Baralle, M., & Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *WIREs RNA*, *4*(1), 49–60. https://doi.org/10.1002/wrna.1140

Fu, X.-D., & Ares, M. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews. Genetics*, *15*(10), 689–701. https://doi.org/10.1038/nrg3778

Fukumura, K., Kato, A., Jin, Y., Ideue, T., Hirose, T., Kataoka, N., Fujiwara, T., Sakamoto, H., & 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*(16), 5303–5311. https://doi.org/10.1093/nar/gkm569

Fukumura, K., Taniguchi, I., Sakamoto, H., Ohno, M., & Inoue, K. (2009). U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. *Nucleic Acids Research*, *37*(6), 1907–1914. https://doi.org/10.1093/nar/gkp050

Gao, C., Ren, S., Lee, J.-H., Qiu, J., Chapski, D. J., Rau, C. D., Zhou, Y., Abdellatif, M., Nakano, A., Vondriska, T. M., Xiao, X., Fu, X.-D., Chen, J.-N., & Wang, Y. (2016). RBFox1-mediated RNA splicing regulates cardiac hypertrophy and heart failure. *The Journal of Clinical Investigation*, *126*(1), 195–206. https://doi.org/10.1172/JCI84015

Gehman, L. T., Meera, P., Stoilov, P., Shiue, L., O'Brien, J. E., Meisler, M. H., Ares, M., Otis, T. S., & Black, D. L. (2012). The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. *Genes & Development*, *26*(5), 445–460. https://doi.org/10.1101/gad.182477.111

Gehman, L. T., Stoilov, P., Maguire, J., Damianov, A., Lin, C.-H., Shiue, L., Ares, M., Mody, I., & Black, D. L. (2011). The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nature Genetics*, *43*(7), Article 7. https://doi.org/10.1038/ng.841

Gilbert, W. (1978). Why genes in pieces? *Nature*, *271*(5645), 501–501. https://doi.org/10.1038/271501a0

Graveley, B. R. (2005). Mutually Exclusive Splicing of the Insect Dscam Pre-mRNA Directed by Competing Intronic RNA Secondary Structures. *Cell*, *123*(1), 65–73. https://doi.org/10.1016/j.cell.2005.07.028

Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., & Zavolan, M. (2021). CLIP and complementary methods. *Nature Reviews Methods Primers*, *1*(1), Article 1. https://doi.org/10.1038/s43586-021-00018-1

Hocine, S., Singer, R. H., & Grünwald, D. (2010). RNA Processing and Export. *Cold Spring Harbor Perspectives in Biology*, *2*(12), a000752. https://doi.org/10.1101/cshperspect.a000752

Homsy, J., Zaidi, S., Shen, Y., Ware, J. S., Samocha, K. E., Karczewski, K. J., DePalma, S. R., McKean, D., Wakimoto, H., Gorham, J., Jin, S. C., Deanfield, J., Giardini, A., Porter, G. A., Kim, R., Bilguvar, K., López-Giráldez, F., Tikhonova, I., Mane, S., … Chung, W. K. (2015). De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science*, *350*(6265), 1262–1266. https://doi.org/10.1126/science.aac9396

Huang, M., Akerberg, A. A., Zhang, X., Yoon, H., Joshi, S., Hallinan, C., Nguyen, C., Pu, W. T., Haigis, M. C., Burns, C. G., & Burns, C. E. (2022). Intrinsic myocardial defects underlie an Rbfox-deficient zebrafish model of hypoplastic left heart syndrome. *Nature Communications*, *13*, 5877. https://doi.org/10.1038/s41467-022-32982-x

Huang, S.-C., Ou, A. C., Park, J., Yu, F., Yu, B., Lee, A., Yang, G., Zhou, A., & Benz, E. J. (2012). RBFOX2 Promotes Protein 4.1R Exon 16 Selection via U1 snRNP Recruitment. *Molecular and Cellular Biology*, *32*(2), 513–526. https://doi.org/10.1128/MCB.06423-11

Jacko, M., Weyn-Vanhentenryck, S. M., Smerdon, J. W., Yan, R., Feng, H., Williams, D. J., Pai, J., Xu, K., Wichterle, H., & Zhang, C. (2018). Rbfox Splicing Factors Promote Neuronal Maturation and Axon Initial Segment Assembly. *Neuron*, *97*(4), 853-868.e6. https://doi.org/10.1016/j.neuron.2018.01.020

Jangi, M., Boutz, P. L., Paul, P., & Sharp, P. A. (2014). Rbfox2 controls autoregulation in RNA-binding protein networks. *Genes & Development*, *28*(6), 637–651. https://doi.org/10.1101/gad.235770.113

Jiang, Y., Fu, X., Zhang, Y., Wang, S.-F., Zhu, H., Wang, W.-K., Zhang, L., Wu, P., Wong, C. C. L., Li, J., Ma, J., Guan, J.-S., Huang, Y., & Hui, J. (2021). Rett syndrome linked to defects in forming the MeCP2/Rbfox/LASR complex in mouse models. *Nature Communications*, *12*(1), 5767. https://doi.org/10.1038/s41467-021-26084-3

Jin, Y., Suzuki, H., Maegawa, S., Endo, H., Sugano, S., Hashimoto, K., Yasuda, K., & Inoue, K. (2003). A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *The EMBO Journal*, *22*(4), 905–912. https://doi.org/10.1093/emboj/cdg089

Kim, K. K., Adelstein, R. S., & 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*(45), 31052–31061. https://doi.org/10.1074/jbc.M109.052969

Kornblihtt, A. R., Schor, I. E., Alló, M., Dujardin, G., Petrillo, E., & Muñoz, M. J. (2013). Alternative splicing: A pivotal step between eukaryotic transcription and translation.

Nature Reviews Molecular Cell Biology, 14(3), 153–165.
https://doi.org/10.1038/nrm3525

Kuroyanagi, H. (2009). Fox-1 family of RNA-binding proteins. Cellular and Molecular Life Sciences: CMLS, 66(24), 3895–3907. https://doi.org/10.1007/s00018-009-0120-5

Lal, D., Pernhorst, K., Klein, K. M., Reif, P., Tozzi, R., Toliat, M. R., Winterer, G., Neubauer, B., Nürnberg, P., Rosenow, F., Becker, F., Lerche, H., Kunz, W. S., Kurki, M. I., Hoffmann, P., Becker, A. J., Perucca, E., Zara, F., Sander, T., & Weber, Y. G. (2015). Extending the phenotypic spectrum of RBFOX1 deletions: Sporadic focal epilepsy. *Epilepsia*, *56*(9), e129–e133. https://doi.org/10.1111/epi.13076

Lal, D., Reinthaler, E. M., Altmüller, J., Toliat, M. R., Thiele, H., Nürnberg, P., Lerche, H., Hahn, A., Møller, R. S., Muhle, H., Sander, T., Zimprich, F., & Neubauer, B. A. (2013). RBFOX1 and RBFOX3 Mutations in Rolandic Epilepsy. *PLoS ONE*, *8*(9), e73323. https://doi.org/10.1371/journal.pone.0073323

Lal, D., Trucks, H., Møller, R. S., Hjalgrim, H., Koeleman, B. P. C., de Kovel, C. G. F., Visscher, F., Weber, Y. G., Lerche, H., Becker, F., Schankin, C. J., Neubauer, B. A., Surges, R., Kunz, W. S., Zimprich, F., Franke, A., Illig, T., Ried, J. S., Leu, C., … EPICURE Consortium. (2013). Rare exonic deletions of the RBFOX1 gene increase risk of idiopathic generalized epilepsy. *Epilepsia*, *54*(2), 265–271. https://doi.org/10.1111/epi.12084

Lee, J.-A., Damianov, A., Lin, C.-H., Fontes, M., Parikshak, N. N., Anderson, E. S., Geschwind, D. H., Black, D. L., & Martin, K. C. (2016). Cytoplasmic Rbfox1 Regulates the Expression of Synaptic and Autism-Related Genes. *Neuron*, *89*(1), 113–128. https://doi.org/10.1016/j.neuron.2015.11.025

Lee, Y., & Rio, D. C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annual Review of Biochemistry*, *84*, 291–323. https://doi.org/10.1146/annurev-biochem-060614-034316

Lovci, M. T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T. Y., Stark, T. J., Gehman, L. T., Hoon, S., Massirer, K. B., Pratt, G. A., Black, D. L., Gray, J. W., Conboy, J. G., & Yeo, G. W. (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nature Structural & Molecular Biology*, *20*(12), 1434–1442. https://doi.org/10.1038/nsmb.2699

Maquat, L. E. (2004). Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics. *Nature Reviews Molecular Cell Biology*, *5*(2), 89–99. https://doi.org/10.1038/nrm1310

Mauger, D. M., Lin, C., & 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*(17), 5403–5419. https://doi.org/10.1128/MCB.00739-08

Moss, N. D., Wells, K. L., Theis, A., Kim, Y.-K., Spigelman, A. F., Liu, X., MacDonald, P. E., & Sussel, L. (2023). Modulation of insulin secretion by RBFOX2-mediated alternative splicing. *Nature Communications*, *14*(1), 7732. https://doi.org/10.1038/s41467-023-43605-4

Nakahata, S., & Kawamoto, S. (2005). Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. *Nucleic Acids Research*, *33*(7), 2078–2089. https://doi.org/10.1093/nar/gki338

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., Garg, N. J., Wehrens, X. H. T., Yeo, G. W., & Kuyumcu-Martinez, M. N. (2016). Dysregulation of RBFOX2 is an early event in cardiac pathogenesis of diabetes. *Cell Reports*, *15*(10), 2200–2213. https://doi.org/10.1016/j.celrep.2016.05.002

Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F., & Sharp, P. A. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science (New York, N.Y.)*, *225*(4665), 898–903. https://doi.org/10.1126/science.6206566

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), 1413–1415. https://doi.org/10.1038/ng.259

Paterson, H. A. B., Yu, S., Artigas, N., Prado, M. A., Haberman, N., Wang, Y.-F., Jobbins, A. M., Pahita, E., Mokochinski, J., Hall, Z., Guerin, M., Paulo, J. A., Ng, S. S., Villarroya, F., Rashid, S. T., Le Goff, W., Lenhard, B., Cebola, I., Finley, D., … Vernia, S. (2022). Liver RBFOX2 regulates cholesterol homeostasis via Scarb1 alternative splicing in mice. *Nature Metabolism*, *4*(12), 1812–1829. https://doi.org/10.1038/s42255-022-00681-y

Piovesan, A., Antonaros, F., Vitale, L., Strippoli, P., Pelleri, M. C., & Caracausi, M. (2019). Human protein-coding genes and gene feature statistics in 2019. *BMC Research Notes*, *12*(1), Article 1. https://doi.org/10.1186/s13104-019-4343-8

Pyle, A. M. (2016). Group II Intron Self-Splicing. *Annual Review of Biophysics*, *45*(Volume 45, 2016), 183–205. https://doi.org/10.1146/annurev-biophys-062215-011149

Raghavan, N. S., Dumitrescu, L., Mormino, E., Mahoney, E. R., Lee, A. J., Gao, Y., Bilgel, M., Goldstein, D., Harrison, T., Engelman, C. D., Saykin, A. J., Whelan, C. D., Liu, J. Z., Jagust, W., Albert, M., Johnson, S. C., Yang, H.-S., Johnson, K., Aisen, P., … Alzheimer's Disease Neuroimaging Initiative. (2020). Association Between Common Variants in RBFOX1, an RNA-Binding Protein, and Brain Amyloidosis in Early and Preclinical Alzheimer Disease. *JAMA Neurology*, *77*(10), 1288–1298. https://doi.org/10.1001/jamaneurol.2020.1760

Robberson, B. L., Cote, G. J., & Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Molecular and Cellular Biology*, *10*(1), 84–94.

Ruskin, B., Krainer, A. R., Maniatis, T., & Green, M. R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell*, *38*(1), 317–331. https://doi.org/10.1016/0092-8674(84)90553-1

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y.-H., Hicks, J., Spence, S. J., Lee, A. T., Puura, K., Lehtimäki, T., … Wigler, M. (2007). Strong association of de novo copy number mutations with autism. *Science (New York, N.Y.)*, *316*(5823), 445–449. https://doi.org/10.1126/science.1138659

Sharp, P. A. (1994). Split genes and RNA splicing. *Cell*, *77*(6), 805–815. https://doi.org/10.1016/0092-8674(94)90130-9

Steitz, T. A., & Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences*, *90*(14), 6498–6502. https://doi.org/10.1073/pnas.90.14.6498

Sun, S., Zhang, Z., Fregoso, O., & Krainer, A. R. (2012). Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. *RNA*, *18*(2), 274–283. https://doi.org/10.1261/rna.030486.111

Underwood, J. G., Boutz, P. L., Dougherty, J. D., Stoilov, P., & Black, D. L. (2005). Homologues of the Caenorhabditis elegans Fox-1 Protein Are Neuronal Splicing Regulators in Mammals. *Molecular and Cellular Biology*, *25*(22), 10005–10016. https://doi.org/10.1128/MCB.25.22.10005-10016.2005

Verma, S. K., Deshmukh, V., Nutter, C. A., Jaworski, E., Jin, W., Wadhwa, L., Abata, J., Ricci, M., Lincoln, J., Martin, J. F., Yeo, G. W., & Kuyumcu-Martinez, M. N. (2016). Rbfox2 function in RNA metabolism is impaired in hypoplastic left heart syndrome patient hearts. *Scientific Reports*, *6*(1), 30896. https://doi.org/10.1038/srep30896

Verma, S. K., Deshmukh, V., Thatcher, K., Belanger, K. K., Rhyner, A. M., Meng, S., Holcomb, R. J., Bressan, M., Martin, J. F., Cooke, J. P., Wythe, J. D., Widen, S. G., Lincoln, J., & Kuyumcu-Martinez, M. N. (2022). RBFOX2 is required for establishing RNA regulatory networks essential for heart development. *Nucleic Acids Research*, *50*(4), 2270–2286. https://doi.org/10.1093/nar/gkac055

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., Mill, J., Cantor, R. M., Blencowe, B. J., & Geschwind, D. H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, *474*(7351), Article 7351. https://doi.org/10.1038/nature10110

Wan, R., Bai, R., Zhan, X., & Shi, Y. (2020). How Is Precursor Messenger RNA Spliced by the Spliceosome? *Annual Review of Biochemistry*, *89*(Volume 89, 2020), 333–358. https://doi.org/10.1146/annurev-biochem-013118-111024

Wang, Z., & Burge, C. B. (2008). Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA*, *14*(5), 802–813. https://doi.org/10.1261/rna.876308

Wei, C., Qiu, J., Zhou, Y., Xue, Y., Hu, J., Ouyang, K., Banerjee, I., Zhang, C., Chen, B., Li, H., Chen, J., Song, L.-S., & Fu, X.-D. (2015). Repression of the Central Splicing Regulator RBFox2 Is Functionally Linked to Pressure Overload-Induced Heart Failure. *Cell Reports*, *10*(9), 1521–1533. https://doi.org/10.1016/j.celrep.2015.02.013

Weyn-Vanhentenryck, S. M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P. A., Zhang, M. Q., Krainer, A. R., Darnell, R. B., & Zhang, C. (2014). HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism. *Cell Reports*, *6*(6), 1139–1152. https://doi.org/10.1016/j.celrep.2014.02.005

Wilkinson, M. E., Charenton, C., & Nagai, K. (2020). RNA Splicing by the Spliceosome. *Annual Review of Biochemistry*, *89*(Volume 89, 2020), 359–388. https://doi.org/10.1146/annurev-biochem-091719-064225

Ye, X., Yang, W., Yi, S., Zhao, Y., Varani, G., Jankowsky, E., & Yang, F. (2023). Two distinct binding modes provide the RNA-binding protein RbFox with extraordinary sequence specificity. *Nature Communications*, *14*(1), 701. https://doi.org/10.1038/s41467-023-36394-3

Yeo, G. W., Coufal, N. G., Liang, T. Y., Peng, G. E., Fu, X.-D., & 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*(2), Article 2. https://doi.org/10.1038/nsmb.1545

Ying, Y. (2016). *Low Complexity Sequences of Rbfox Form Higher-order Complexes with LASR to Regulate Alternative Splicing* [UCLA]. https://escholarship.org/uc/item/2tv8r5f4

Ying, Y., Wang, X.-J., Vuong, C. K., Lin, C.-H., Damianov, A., & Black, D. L. (2017). Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell*, *170*(2), 312-323.e10. https://doi.org/10.1016/j.cell.2017.06.022

Zarnack, K., König, J., Tajnik, M., Martincorena, I., Eustermann, S., Stévant, I., Reyes, A., Anders, S., Luscombe, N. M., & Ule, J. (2013). Direct Competition between hnRNP C and U2AF65 Protects the Transcriptome from the Exonization of Alu Elements. *Cell*, *152*(3), 453–466. https://doi.org/10.1016/j.cell.2012.12.023

Zhou, H.-L., Baraniak, A. P., & 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*(3), 830–841. https://doi.org/10.1128/MCB.01015-06

Zhou, H.-L., & Lou, H. (2008). Repression of prespliceosome complex formation at two distinct steps by Fox-1/Fox-2 proteins. *Molecular and Cellular Biology*, *28*(17), 5507–5516. https://doi.org/10.1128/MCB.00530-08

Chapter 2: The Rbfox1/LASR complex controls alternative splicing by multi-subunit recognition of cis-regulatory RNA modules

Abstract

The proteins Rbfox1, Rbfox2, and Rbfox3 regulate alternative splicing in neurons and other cells by binding to GCAUG and related secondary motifs. However, Rbfox binds only a small fraction of expressed GCAUGs, and how it recognizes its regulatory sites remains unclear. In the nucleus, most of Rbfox is bound to a large assembly of splicing regulators (LASR), a heteromeric complex of RNA-binding proteins. Subunits of LASR have affinities for different RNA elements, and it remains unknown how the Rbfox/LASR complex contacts RNA and regulates splicing. We used a nuclease-protection assay to map the transcriptome-wide footprints of Rbfox1/LASR. In addition to GCAUG, Rbfox1/LASR also binds and protects RNA containing motifs for LASR subunits hnRNP M, H, C, and Matrin3. These elements are often arranged in tandem, forming multi-part modules of RNA motifs. To differentiate contact sites of Rbfox1 from other subunits, we analyzed an RNA-binding mutant Rbfox1(F125A) that remains associated with LASR. Rbfox1/LASR complexes with this mutation lost interaction with GCAUG but retained binding to elements for LASR. Splicing analysis showed that, beyond the canonical mode of activating exons by binding to downstream GCAUGs, Rbfox also stimulates exon inclusion by binding to downstream secondary motifs when LASR subunits bind nearby. Furthermore, Rbfox restricts LASR's regulatory targets; when LASR is in complex with the RNA-binding mutant Rbfox, it binds new sites and activates additional exons. Mini-gene experiments confirmed Rbfox1/LASR activates splicing through binding to diverse elements that are combined to yield additive effects. These findings demonstrate that Rbfox1/LASR recognizes combinations of tandem RNA elements and decodes these combinatorial splicing regulatory motifs.

25

Introduction

Eukaryotes produce multiple mRNAs from the same pre-mRNA by changing the choice of splice sites that define an intron to be excised. This process of alternative splicing is regulated by trans-acting RNA binding proteins (RBPs) that bind to cis-regulatory elements on pre-mRNAs (Black, 2003). RBPs either enhance or repress exon inclusion. Their effects depend on their binding position relative to the target exon and the actions of other factors binding nearby . Alternative exons usually carry binding elements for multiple RBPs that can interact, giving rise to a complex combinatorial code that is difficult to unravel (Smith & Valcárcel, 2000; Ule & Blencowe, 2019).

One family of regulators controlling important splicing programs in the nervous system and over development is the Rbfox proteins (Conboy, 2017). These RBPs, conserved from *c. elegans* to humans, are unusual for having a single RNA binding domain that is highly specific for the element GCAUG (Begg et al., 2020; Jin et al., 2003; Lambert et al., 2014; Ye et al., 2023; Yeo et al., 2009). Mammals have three Rbfox genes: *RBFOX1* is abundant in the brain, heart, and muscle, *RBFOX2* exhibits broad expression across tissues during development, and *RBFOX3* appears exclusive to the brain (Conboy, 2017). Each of these genes produces multiple spliced products, including cytoplasmic isoforms that regulate translation and other processes, and nuclear isoforms that regulate splicing. Besides its conserved RNA binding domain, Rbfox also contains a low complexity, tyrosine-rich C-terminal domain (CTD). This CTD can homo-oligomerize and also binds to the large assembly of splicing regulators (LASR), a heteromeric complex of RBPs consisting of hnRNP M, hnRNP H/F, hnRNP C, DDX5, NF-110, hnRNP UL2, NF-45, Matrin3, and in some preparations MeCP2

(Damianov et al., 2016; Jiang et al., 2021; Ying et al., 2017). The repetitive tyrosines within the CTD are required for assembly of Rbfox/LASR into higher-order complexes and its ability to activate splicing.

Rbfox proteins play critical roles in the development and function of multiple organs. In the central nervous system, mutation or aberrant expression of these proteins can lead to electrophysiological abnormalities, seizures, and defects in cerebellar development in mice, and epileptic and/or autism spectrum disorders in human patients (Gehman et al., 2011, 2012; Jacko et al., 2018; Lal, Reinthaler, et al., 2013; Lal, Trucks, et al., 2013; Vuong et al., 2018). In the heart, dysregulation of Rbfox2 can lead to hypoplastic left heart syndrome, congenital heart disease, and conduction defects in Myotonic Dystrophy 1 (Homsy et al., 2015; Huang et al., 2022; Misra et al., 2020; Verma et al., 2016). Rbfox2 is also involved in cholesterol homeostasis in the liver and controlling insulin secretion from the pancreas (Moss et al., 2023; Paterson et al., 2022). Moreover, Rbfox2 contributes to epithelial-mesenchymal transition and is important for pancreatic cancer metastasis and breast cancer development (Jbara et al., 2023; Li et al., 2018; Maurin et al., 2023; Shapiro et al., 2011; Venables et al., 2013).

The *in vivo* binding sites for RBPs can be identified using cross-linking and immunoprecipitation (CLIP) (Hafner et al., 2021). Multiple studies have correlated Rbfox cross-linked sites with exons affects by its expression (Begg et al., 2020; Damianov et al., 2016; Jangi et al., 2014; Lovci et al., 2013; Weyn-Vanhentenryck et al., 2014; Yeo et al., 2009). These analyses confirmed that all three Rbfox proteins bind GCAUG, with a preference for UGCAUG, with close variants of this motif also crosslinked. Binding to

GCAUGs downstream of an exon usually enhances its inclusion while binding upstream generally represses it.

However, GCAUG is a common pentamer in genomic RNA, and not all GCAUGs in the transcriptome crosslink to Rbfox. Conversely, some Rbfox crosslinked sites do not contain GCAUG or its related secondary motifs (Begg et al., 2020). These sites might contain undefined motifs with an affinity for Rbfox or could arise from Rbfox crosslinking by virtue of its proximity to a sequence without a high affinity interaction. Although binding sites can be inferred by enriched motifs at the crosslink site, positions of crosslinking do not necessarily indicate high-affinity interactions or provide information on the extent of RNA surrounding the crosslinked nucleotide that contacts a protein.

Rbfox's association with LASR helps explain its selectivity for particular GCAUG elements (Damianov et al., 2016; Ying et al., 2017). GCAUGs that crosslink to Rbfox are enriched for adjacent elements predicted to bind the LASR subunit hnRNP M, suggesting cooperative binding. Furthermore, many Rbfox crosslinked sites lacking GCAUG contain predicted binding motifs for the LASR subunit hnRNP M and hnRNP C. Integrative analysis of CLIP for Rbfox, hnRNP M, hnRNP C, and SRSF1 indicates that Rbfox may regulate targets through direct binding, cooperative binding with a partner, or indirect binding via a partner (Zhou et al., 2021). However, it is unclear if these effects result from Rbfox co-binding with LASR and how these RNA binding proteins, each with affinities for different motifs, work together to recognize their targets.

In this study, we use a subcellular fractionation and nuclease-protection assay to map the transcriptome-wide RNA footprints of Rbfox1/LASR in HEK293 cells. These footprints provide information on the extent of RNA contact by the proteins at different

28

regulatory sites. We find that these footprints consist of multi-part RNA elements that can bind Rbfox and the LASR subunits hnRNP M, hnRNP H/F, hnRNP C, and Matrin3. Comparing footprints of LASR containing wildtype Rbfox1 with an RNA-binding mutant F125A allowed us to distinguish the binding sites of Rbfox1 from those of LASR subunits. By comparing splicing changes in Rbfox1 wildtype and F125A expressing cells, we identify exons dependent on binding by LASR subunits, in addition to exons requiring Rbfox binding to GCAUG. Finally, in mini-gene experiments, we find that both GCAUG and other elements within the tandem sites exert a positive additive effect on exon inclusion.

Results

Mapping Rbfox1/LASR binding sites on chromatin-associated RNA by nuclease-protection

We recently described IP-seq as a method to map the sites of U2 binding across the transcriptome (Damianov et al., 2023). This protocol involves isolation of epitope tagged factors from the chromatin fraction of cells after extraction with nuclease. We found that the isolated U2 snRNP in this fraction was bound to protected RNA fragments corresponding to branchpoints basepaired to the U2 snRNA. This method can potentially map the contact sites of other proteins bound to nascent RNA, assuming their binding affinity is sufficient to withstand the nuclease degradation. Using the same chromatin extraction, we previously isolated FLAG-tagged Rbfox proteins from HEK293 cells and examined their co-interacting protein partners. We found that nuclear Rbfox proteins were almost entirely bound by the large assembly of splicing regulators, LASR, a complex of other RBPs. In this study, we asked whether there were nuclease-protected RNAs found in our Rbfox/LASR preparations.

We first engineered a HEK293 line to express FLAG-tagged Rbfox1. Rbfox1 and Rbfox3 are not expressed in HEK293 cells, and we previously created a HEK293 cell line with the endogenous Rbfox2 knocked out (Damianov et al., 2016). In this study, we used this Rbfox deficient line as a recipient for a doxycycline inducible FLAG-Rbfox1 construct and titrated its expression to levels found in the brain. From these cells, we isolated nuclei, lysed them with Triton X-100, and centrifuged the lysate to obtain a soluble nucleoplasmic supernatant and a pellet containing chromatin and other high molecular weight material (Fig. 2.1A). The chromatin pellet was treated with the

Benzonase nuclease, degrading both RNA and DNA, to solubilize material within the pellet. Anti-FLAG immunoprecipitates were isolated from both the nucleoplasm and the chromatin extract, eluted with FLAG peptide, and analyzed by SDS-PAGE stained for total protein (Figure 2.1B). As seen previously, Rbfox1 was more abundant in the chromatin extract than the soluble nucleoplasm. This FLAG-Rbfox1 copurified with additional bands of roughly equal intensity corresponding to the LASR subunits. Thus, as observed previously, the majority of nuclear Rbfox1 is associated stoichiometrically with LASR on chromatin (Damianov et al., 2016; Jiang et al., 2021; Ying et al., 2017).

To examine if any RNA co-purified with the complex, the immunoprecipitated material was deproteinized, extracted with phenol-chloroform, DNAse treated, and 5' end-labeled with [γ-^{32}P]ATP. This material was then analyzed by Urea-PAGE and autoradiography. As shown in Figure 2.1C, small RNA fragments less than 50 nucleotides long copurified with Rbfox1/LASR, presumably protected from nuclease degradation by their interaction with the complex. To identify these RNAs, we converted fragments of 20-50 nucleotides into cDNA using a modified iCLIP library protocol and sequenced them.

The FLAG-Rbfox1 IP-seq yielded 16.5 to 32.7 million reads per replicate with an average length of 32 nucleotides (Fig. S2.1A). Across the three replicates, greater than 60% of the reads mapped to unique locations in the genome, forming clusters of aligned reads. We defined 561,273 clusters that contained at least 10 reads in merged replicates using the peak caller Yodel (Palmer et al., 2017). Control IP-seq from cells that do not express FLAG-Rbfox1 yielded 14 million reads with a 68% mapping rate. Interestingly, many of the reads from the control cells aligned at intron branchpoints

31

similar to what we observed with isolated U2 snRNPs (Fig. S2.1B). These reads from

the control cells are attributed to weak cross-reactivity of the anti-FLAG antibody with a

subunit of the U2 snRNP. To distinguish bona fide binding sites from background, we

filtered for clusters that were enriched in the experimental set compared to the

background set, leaving us with 472,757 clusters. These clusters were mostly

intragenic, and the majority mapped to introns or 3' UTRs, similar to Rbfox binding

patterns measured by CLIP (Fig. S2.1C-D).

Figure 2.1D shows a genome browser view of fragments that map in introns

adjacent to an alternative exon of *MARK3*. Two peaks of protected fragments are

adjacent to the 5' splice site of the upstream intron (Fig. 2.1D). These regions are more

defined than the broader clusters of crosslink sites from a previous eCLIP study of

Rbfox2 also in HEK293 cells (Van Nostrand et al., 2016). The two IP-seq clusters in this

region contain distinct sets of motifs. The first cluster contains two GCAUG elements

along with a downstream GU-rich element. The second cluster lacks GCAUGs and

instead contains G-rich and GU-rich elements. These non-GCAUG motifs within the

protected fragments are potential binding sites for the LASR subunits hnRNP M, which

binds GU-rich motifs (Feng et al., 2019; Ho et al., 2021; Zhu et al., 2022), and hnRNP

H/F, which binds G-rich elements (Matunis et al., 1994; Uren et al., 2016; Van Nostrand

et al., 2020). The eCLIP tags in this region are distributed across all these elements. A

similar pattern is observed in the IP-seq protected fragments adjacent to the 5' splice

site of the downstream intron. In this region the protected fragments align in two peaks,

and each contains a GCAUG element plus other motifs. The eCLIP tags are again more

broadly distributed across this region. Rbfox crosslinking to regions without GCAUG can

occur due to its proximity as part of the LASR complex, whose subunits can bind to these sites.

We then assessed the overlap between Rbfox1/LASR IP-seq clusters and Rbfox2 eCLIP clusters. Both the IP-seq and eCLIP clusters were each segregated into two groups based on the presence or absence of GCAUG. 50% of GCAUG-containing Rbfox2 eCLIP clusters overlapped with 5% of GCAUG-containing IP-seq clusters (Fig. S2.2A). The higher fraction of eCLIP clusters overlapping with a small subset of IP-seq clusters can be due to a variety of reasons: 1) some Rbfox binding sites might not crosslink and therefore be undetected by CLIP, 2) some sites might not have a high enough affinity to withstand the nuclease treatment in IP-seq and therefore be undetected by this method, and 3) eCLIP and IP-seq clusters are defined using different computational approaches, which can result in differences in sensitivity and specificity of cluster detection. In addition to the clusters containing GCAUG, 18% of Rbfox2 eCLIP clusters without GCAUG overlapped with 3% of IP-seq clusters lacking this motif (Fig. S2.2B). The overlap of clusters without GCAUG is presumably lower than those containing GCAUG since CLIP will mostly detect Rbfox bound sites whereas IP-seq can also detect sites bound by LASR.
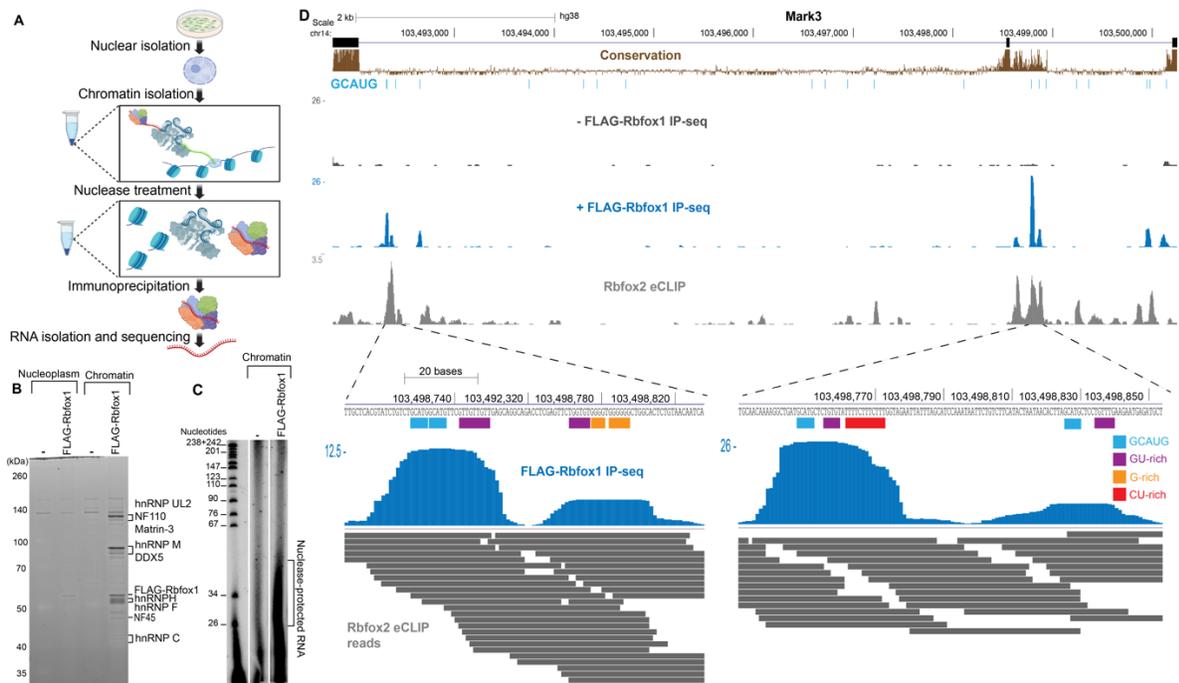
Figure 2.1. Mapping Rbfox1/LASR binding sites on chromatin-associated RNA by nuclease-protection.

A) Experimental design for purifying nuclease-protected RNA associated with chromatin-enriched protein complexes through the IP-seq method. B) SYPRO Ruby stained SDS-PAGE of immunoprecipitations from parental (-) and FLAG-Rbfox1 expressing HEK293 cells. C) Urea-PAGE of P32 labeled nuclease-protected RNAs purified from immunoprecipitations. D) UCSC Genome Browser view of nuclease-protected fragments in adjacent introns of Mark3 exon 16. Light blue bars indicate position of GCAUGs in transcript and the phyloP conservation across 100 vertebrae species is displayed in brown. The IP-seq browser track in black is from parental (-) and the blue is from FLAG tagged Rbfox1 expressing cells. Rbfox2 eCLIP from HEK293T cells from Van Nostrand et al. is shown in gray. Two regions containing protected sites

and CLIP reads are shown in more detail below the plot and motifs are delineated by colored bars as indicated in the figure.

RNA protected by Rbfox1/LASR is enriched in binding motifs for Rbfox and other LASR subunits

We next searched for RNA motifs that were enriched in nuclease-protected fragments relative to their frequency in their total intron sequence. HOMER analysis revealed the canonical Rbfox UGCAUG element to be the most enriched motif, found in 25% of the reads (Fig. 2.2A) (Heinz et al., 2010). Additionally, 30% of the reads contain the GCAUG motif, while secondary Rbfox motifs (e.g. GCACG) were less prevalent though still significant (Fig. S2.2A). Other motifs were also enriched in the protected fragments and these match defined motifs for individual LASR subunits. GU-rich elements (GUGUGU, GUUGUU) that are known to bind hnRNP M were present in 21% of reads, nearly as enriched as Rbfox motifs. The GUGUGU motif has also been reported to be a secondary Rbfox motif (Begg et al., 2020). However, its abundance in the protected fragments compared to other secondary Rbfox motifs suggests its isolation is likely due to a LASR subunit interaction rather than Rbfox. G-rich, U-rich, and CU-rich elements suggesting hnRNP H/F, hnRNP C, and Matrin3 binding, respectively, were each present in at least 10% of reads (Ramesh et al., 2020; Zarnack et al., 2013). Also among the most enriched sequences was a motif containing UAG whose cognate binding factor is not yet clear.

In an alternative approach, we compared the frequency of pentamers within the intronic IP-seq clusters with their frequencies in randomly sampled regions from the same introns. Z-scores for these pentamers were calculated, and their distribution is

illustrated in figure 2.2B. Pentamers scoring two standard deviations above the mean are highlighted and their sequences are shown in figure 2.2C. All these pentamers can be classified into motif categories identified in the HOMER analysis. The motifs clustered together with the highest z-scores are GCAUG (684), UUGUU (688), and UGUGU (681). A population of GU-rich, G-rich, CU-rich, and U-rich elements have z-scores between 300-600 and another population have z-scores below 300. The most enriched UAG containing element, UUAGU, has a z-score of 307. Z-score analysis of hexamers gave similar results (Table 2.3.).

Figure 2.2. RNA protected by Rbfox1/LASR is enriched in binding motifs for Rbfox and other LASR subunits.

A) HOMER analysis of enriched motifs of 4-6 nucleotides in protected fragments within clusters in introns of protein-coding genes. B-C) Z-score analysis of pentamers within clusters in introns of protein-coding genes (B). Pentamers with a z-score of 2 standard deviations above the mean are categorized and displayed (C).
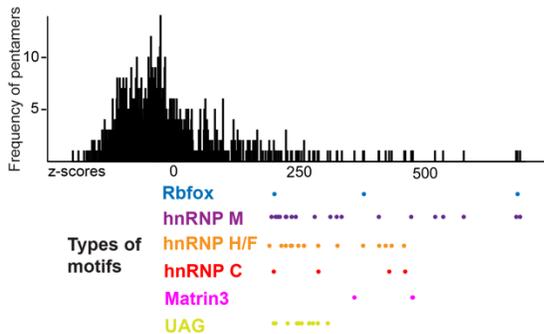
RNA bound to Rbfox1/LASR contains modules of tandem multi-part RNA motifs

We found that 13% of the GCAUG motifs within expressed transcripts were recovered in protected RNA fragments. For the full UGCAUG hexamer, 27% were found in RNA fragments (Fig. 2.3A; Fig. S2.5A). This observation is consistent with CLIP studies where not all GCAUGs crosslink to Rbfox (Damianov et al., 2016; Jangi et al., 2014; Lovci et al., 2013; Weyn-Vanhentenryck et al., 2014). Previous findings suggest that neighboring motifs may affect Rbfox binding (Damianov et al., 2016). To examine this hypothesis, we analyzed motif co-occurrence near protected versus unprotected GCAUGs (Fig. 2.3B). GU-rich and U-rich elements, known to be enriched near crosslinked GCAUGs, were also more common around protected GCAUGs. More than 50% of regions containing a GCAUG also contained a GU-rich element. GCAUG and other Rbfox motifs were also enriched as a second motif in protected RNAs already containing one GCAUG. Furthermore, G-rich, CU-rich, and UAG-containing motifs were more frequent near protected GCAUGs. In contrast, C-rich motifs, not enriched in our IP-seq, showed similar prevalence around both protected and unprotected GCAUGs. Thus, presence of motifs bound by LASR can influence what GCAUG elements Rbfox binds to. We also analyzed co-occurrence of these motifs relative to GCAUGs with and without a 5' terminal U (Fig. S2.5B-C). There is a slightly higher occurrence of motifs adjacent to GCAUG without a 5' U compared to UGCAUG, suggesting LASR plays a bigger role for Rbfox's target selection when it binds to a lower affinity motif.

RBPs within ribonucleoprotein complexes can make direct contact with contiguous motifs along an RNA strand (Hennig et al., 2014; Kuwasako et al., 2014; Wysoczański et al., 2014). Recognition of such elements can increase the affinity of

binding and might confer greater resistance to nuclease digestion. Therefore, we assess for co-occurrence of motifs within the same protected fragment. Approximately 90% of protected fragments with GCAUG also contained at least one other binding motif, with U-rich elements being the most common, appearing in 66% of GCAUG-containing fragments. GU-rich elements were also common, found in 52% of GCAUG-protected fragments (Fig. 2.3C). Binding of multiple proteins to the same RNA fragment might also extend the length of the nuclease-protected region. Accordingly, protected fragments containing GCAUG alongside another enriched motif exhibited a mean length of 34 nucleotides, compared to 28 nucleotides for fragments containing only GCAUG and no other enriched motifs (Fig. 2.3D). The presence of different types of motifs in reads is also correlated with different distributions of read lengths (Fig. S2.4). These results suggest that Rbfox and LASR can have multiple contact points within the same region of RNA.

To examine the placement of additional motifs relative to the GCAUG, we plotted the frequency of other motifs along sequences adjacent to protected GCAUG motifs (Fig. 2.2E). All motifs showed a preference to be within 1-10 nucleotides of the protected GCAUG and were enriched both upstream and downstream. GU-rich, and CU-rich motifs were particularly common directly upstream of the GCAUG, presumably a reflection of the enrichment for U as the initial nucleotide of UGCAUG. Interestingly, G-rich motifs are usually slightly further away from the Rbfox motif, indicating a structural constraint on the co-binding of these two proteins. None of the motifs showed an enrichment proximal to unprotected GCAUGs. Notably, U-rich and CU-rich motifs showed enrichment approximately 25 nucleotides downstream of unprotected

GCAUGs, suggesting that a motif at this distance disfavors binding of Rbfox to the

GCAUG. In contrast to the LASR elements, C-rich motifs, which are not enriched in the

Rbfox1/LASR protected RNA, are depleted in sequences adjacent to protected

GCAUGs. We also analyzed co-occurrence of motifs in protected fragments as a

separate group from the above analysis of regions containing a protected GCAUG. This

analysis shows similar patterns of positional enrichment of other motifs relative to

GCAUG (Fig. S2.5C). Examples showing the presence of multiple motifs in protected

fragments are shown for the *MAD1L1* and *LAMP2* transcripts in figure 2.2F.
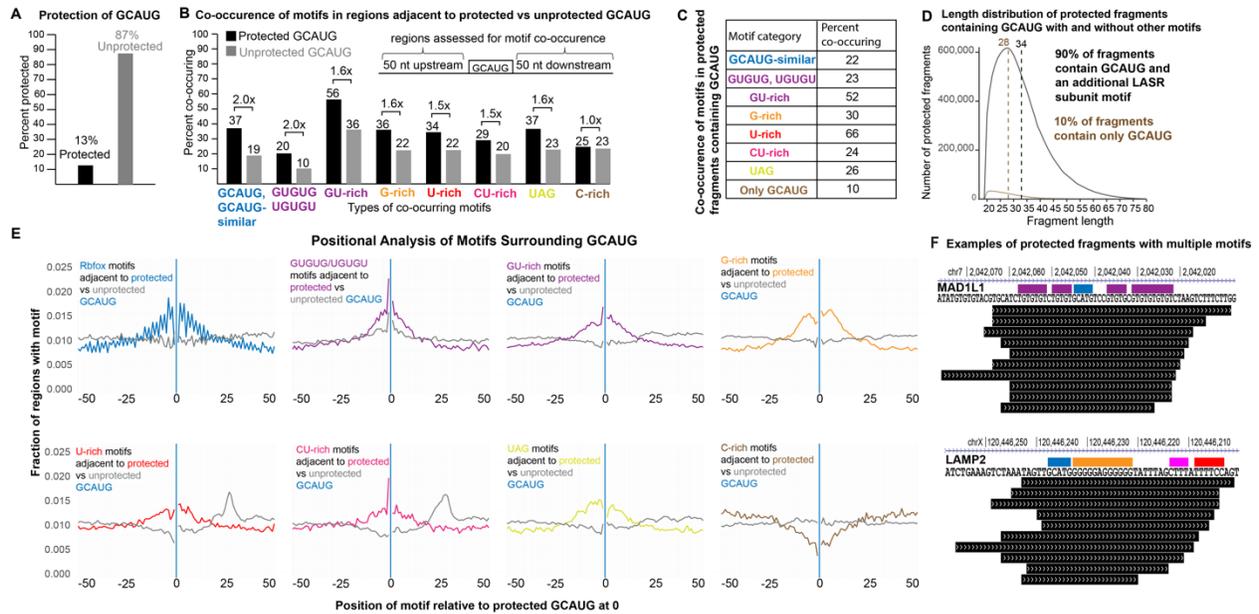
Figure 2.3. RNA bound to Rbfox1/LASR contains modules of tandem multi-part RNA motifs

A) Fraction of expressed GCAUGs protected by Rbfox/LASR. To be considered in the analysis, the GCAUG element must be within in an intron that contains at least one IP-seq cluster. B) Co-occurrence of motifs in regions surrounding protected (black) and unprotected GCAUG (grey). A region extending 50 nucleotides upstream and downstream of a GCAUG was defined as the surrounding region. GCAUG-like motifs include all Rbfox secondary motifs shown in supplementary figure 2.2. GU-rich motifs consist of all pentamers under hnRNP M in figure 2.2, except for GUGUG and UGUGU, which form their own category. U-rich motifs cover all pentamers associated with hnRNP C, CU-rich includes all motifs under Matrin3, and UAG, encompasses all motifs labeled unknown in figure 2.2. C-rich motifs include ten pentamers with the lowest z-scores: AAACA, CCAGG, GCCAC, CCUCA, GCCCA, CAGCC, CUCCC, CCAGC, UCCCA, and CCCAG. C) Percent of protected fragments that contain GCAUG and other specified

motifs. D) Distribution of length of protected fragments that contain GCAUG and motifs of LASR subunits compared to fragments with GCAUG that do not contain these motifs. Dotted lines are at the mean of the distribution. E) Frequency of motif-occurrences surrounding protected (colored) and unprotected (grey) GCAUGs. F) Examples of protected fragments containing combinations of motifs in the MAD1L1 and LAMP2 transcripts. Colors follow the same pattern indicated in panels C-E.

Rbfox1/LASR containing the F125A RNA binding mutant Rbfox1 loses binding of GCAUG but not LASR binding elements

To examine the contribution of the Rbfox RNA binding domain to the isolation of the protected fragments, we analyzed mutant Rbfox1 protein (Rbfox1(F125A)). This mutation eliminates a critical phenylalanine on the RNA binding surface of the domain and yields protein with a K_d for GCAUG that is approximately 1500 times higher than wildtype Rbfox1 (Auweter et al., 2006). FLAG-Rbfox1(F125A) was integrated into the Rbfox2$^{-/-}$ cells and isolated via its epitope tag. Total protein staining of the immunoprecipitates yielded a banding pattern matching that of wildtype FLAG-Rbfox1, indicating the mutant and wildtype Rbfox1 both interact with LASR (Fig. 2.4A). RNA isolated from the Rbfox1(F125A)/LASR complex had a similar Urea-PAGE profile to the RNA protected by Rbfox1(WT)/LASR (Fig. 2.4B). This protected RNA was sequenced and had similar average lengths and mapping rates to wildtype (Fig. S2.1A).

Rbfox1(F125A)/LASR protected RNA fragments with different motif frequencies than Rbfox1(WT) (Fig. 2.4C). The GCAUG motif was no longer enriched, while all the other motifs enriched with Rbfox1(WT)/LASR complex remained, although their frequencies shifted (Fig. 2.4C, Fig. S2.6). G-rich (hnRNP H/F) and CU-rich elements (Matrin3) were more frequent compared to wildtype. GU-rich elements (hnRNP M sites) decreased in frequency when Rbfox binding was lost. DESeq2 was used to analyze differential binding between the LASR complexes containing wildtype and F125A Rbfox1 (Fig. 2.4D). We classified 24,616 sites as WT enriched, characterized by a log2(WT/FA) > 1 and an FDR < 0.05. A total of 26,337 sites were defined as unchanged and therefore common between WT and the FA mutant, with log2(WT/FA) values

43

ranging between -0.1 and 0.1. Fewer sites (1,578) were identified as FA enriched, with log2(WT/FA) < -1 and an FDR < 0.05.

We then analyzed motif prevalence in differentially or commonly bound sites (Fig. 2.4E). GCAUG appeared in 92% of WT enriched sites, but only 3% of FA enriched and 2% of commonly bound sites. Rbfox secondary motifs (GCUUG, GAAUG, GCACG, GUAUG, GUUUG) were moderately more prevalent in WT enriched sites (32%) compared to FA enriched (21%) and commonly bound (24%) sites. Conversely, G-rich elements were more frequent in the common sites (41%) and FA enriched sites (48%) than in WT enriched sites (25%). GU-rich motifs were similarly prevalent across all sites (62-70%). U-rich, CU-rich, or UAG containing elements all had the highest prevalence in commonly bound sites, followed by WT enriched, and the lowest in FA enriched sites.

An example of these differentially protected sites is downstream of an Rbfox-regulated exon in the *TSC2* transcript, a gene essential for cell growth (Huang et al. 2008) (Fig. 2.4F). eCLIP of Rbfox2 exhibited broadly distributed crosslinking across the entire region downstream of the exon. In contrast, IP-seq produced two clear peaks of protected fragments. The more proximal peak, containing several conserved G-rich elements, was bound by both wildtype Rbfox1 (WT) and the F125A mutant, indicating that these elements are protected independently of Rbfox's RNA binding domain, most likely by hnRNP H/F. A second downstream cluster containing conserved GCAUG motifs was isolated with the wildtype Rbfox1, but disappeared in the F125A mutant. Thus, Rbfox's selective affinity for GCAUG is essential for binding to this region and protecting it from nuclease cleavage.

In another example in the *SPHK2* transcript, differential protection by Rbfox versus LASR was observed within a single protected region (Fig. 2.4G). In the Rbfox1(WT) IP-seq, the predominant peak aligns with a single UGCAUG, but extends upstream to include GU-rich, CU-rich, and G-rich elements. In the Rbfox1(F125A) IP-seq, protection at the UGCAUG site and the upstream CU-rich site decreases significantly, while protection of the G-rich and GU-rich regions is maintained. Comparison of IP-seq between wildtype and F125A Rbfox can therefore identify protected GCAUGs even when they are directly adjacent to other elements.

Figure 2.4. Rbfox1/LASR containing the F125A RNA binding mutant Rbfox1 loses binding of GCAUG but not LASR binding elements.

A) SYPRO Ruby stained SDS-PAGE of immunoprecipitation from FLAG-Rbfox1(WT) and FLAG-Rbfox1(F125A) expressing cells. B) Urea-PAGE of P32 labeled nuclease-protected RNAs purified from immunoprecipitations of FLAG-Rbfox1(WT) and FLAG-Rbfox1(F125A). C) Enriched motifs in protected fragments of Rbfox1(F125A)/LASR

from clusters in introns of protein-coding genes. D) Volcano plot illustrating differential binding in Rbfox1(WT)/LASR versus Rbfox1(F125A)/LASR as analyzed by DESeq. E) Analysis of occurrence of motifs in differentially bound sites. F-G) UCSC Genome Browser view of nuclease-protected fragments in the downstream intron of the 129 nucleotide exon 25 of Tsc2 (F) and an intronic region in SPHK2 (G). Positions of enriched motifs in protected sites are indicated with colored bars.

Rbfox1(WT)/LASR and Rbfox1(F125A)/LASR each regulate splicing through distinct binding sites adjacent to cassette exons.

Given that LASR can bind to numerous sites in complex with Rbfox(F125A), sometimes binding to new sites, we assessed its ability to regulate splicing. We compared splicing in cells lacking Rbfox with cells expressing either Rbfox1(WT) or Rbfox(F125A) at comparable levels (Fig. 2.5A). PolyA(+) RNA was isolated and subjected to standard short-read RNAseq. rMATS was used to identify splicing changes in each condition and we discerned three regulation patterns (Fig. 2.5B, Fig. S2.7). 462 exons were uniquely regulated by the wildtype Rbfox1 (WT-regulated) whereas 542 exons were affected by both the wildtype and F125A mutant (WT-FA regulated). A smaller third set of exons, 288, were regulated by the F125A mutant but not by the wildtype (FA-regulated).

RBPs often have position-dependent effects on their regulatory targets, with Rbfox typically activating exons when bound downstream and inhibiting them when bound upstream. To examine how protected regions containing different motifs affected splicing, we analyzed their frequencies within 500 nucleotides up- and downstream of activated and repressed exons of each regulatory class (Fig. 2.5C, Fig. S2.8). Clusters from both Rbfox(WT) and Rbfox1(F125A) IP-seq were categorized based on the presence of GCAUG, Rbfox-secondary (GCUUG, GAAUG, GCACG, GUAUG, GUUUG), GUGU, and G-rich motifs. For exons activated by WT only (first column), Rbfox1(WT)/LASR clusters with GCAUG were enriched downstream, consistent with previous CLIP results of downstream binding leading to splicing activation. In contrast, Rbfox1(F125A)/LASR clusters with GCAUG were seldom found near these exons.

Rbfox1(WT)/LASR clusters containing Rbfox secondary motifs, GUGU, and G-rich elements were also found more frequently downstream of WT-activated exons compared to Rbfox1(F125A)/LASR clusters with these motifs. These data indicate that the activation of exons by the wildtype Rbfox1 is highly dependent on binding to downstream GCAUG, Rbfox secondary, and LASR RNA elements.

For the exons activated by both WT and FA (Figure 2.5C, second column), there was only limited downstream enrichment of Rbfox1(W)/LASR clusters with GCAUG. Rbfox1(WT)/LASR and Rbfox1(F125A)/LASR clusters with Rbfox secondary, GUGU, and G-rich motifs exhibited similar enrichment downstream of these exons. These exons can apparently be activated by either the wildtype or F125A Rbfox1 binding to lower affinity secondary motifs aided by LASR binding to nearby GUGU and G-rich motifs. The coincident distributions of wildtype and F125A clusters with GUGU and G-rich motifs downstream of these exons supports the idea that these exons are regulated by both Rbfox and LASR contacts.

The number of exons activated only by FA was smaller than the other two groups making enrichments more difficult to discern. Rbfox1(WT)/LASR and Rbfox1(F125A)/LASR clusters containing GCAUG show no clear positional bias. However, there were notable enrichments of Rbfox1(F125A)/LASR clusters with Rbfox secondary, GUGU, and G-rich elements directly adjacent to the 5' splice site of some of these exons (bottom three graphs in third column of Fig. 2.5C), possibly indicating activation by LASR subunits rather than Rbfox1. Interestingly, there were also enrichments of Rbfox1(WT)/LASR and possibly Rbfox1(F125A)/LASR clusters containing secondary Rbfox and G-rich motifs upstream of some of these exons. These

binding patterns raise the possibility that the F125A mutant may sometimes act as a dominant negative factor to reverse splicing repression by upstream binding of Rbfox/LASR. Such a dominant negative effect has been observed by a splice variant of Rbfox that lacks a portion of the RNA-binding domain.

We also assessed binding of Rbfox1/LASR around repressed exons (Fig. S2.8). For the exons repressed only by WT, Rbfox1(WT)/LASR clusters with GCAUG showed modest upstream preferential compared to Rbfox1(F125A)/LASR clusters. This less pronounced position-dependent effect of splicing repression compared to activation is similar to patterns observed in CLIP (Jangi et al., 2014). Occurrence of protected sites containing Rbfox secondary, GUGU, and G-rich elements did not seem to differ between the wildtype and F125A IP-seq for this set of exons. In the WT-FA repressed set, the prevenance of protected sites containing GCAUG, Rbfox secondary, GUGU, and G-rich motifs were low in the IP-seq for both the wildtype and the F125A mutant. The same pattern was also observed for the FA repressed exons with no clear discernable enrichment patterns present. These results suggest that the mechanism of exon repression by Rbfox/LASR is likely more complex compared to exon inclusion.
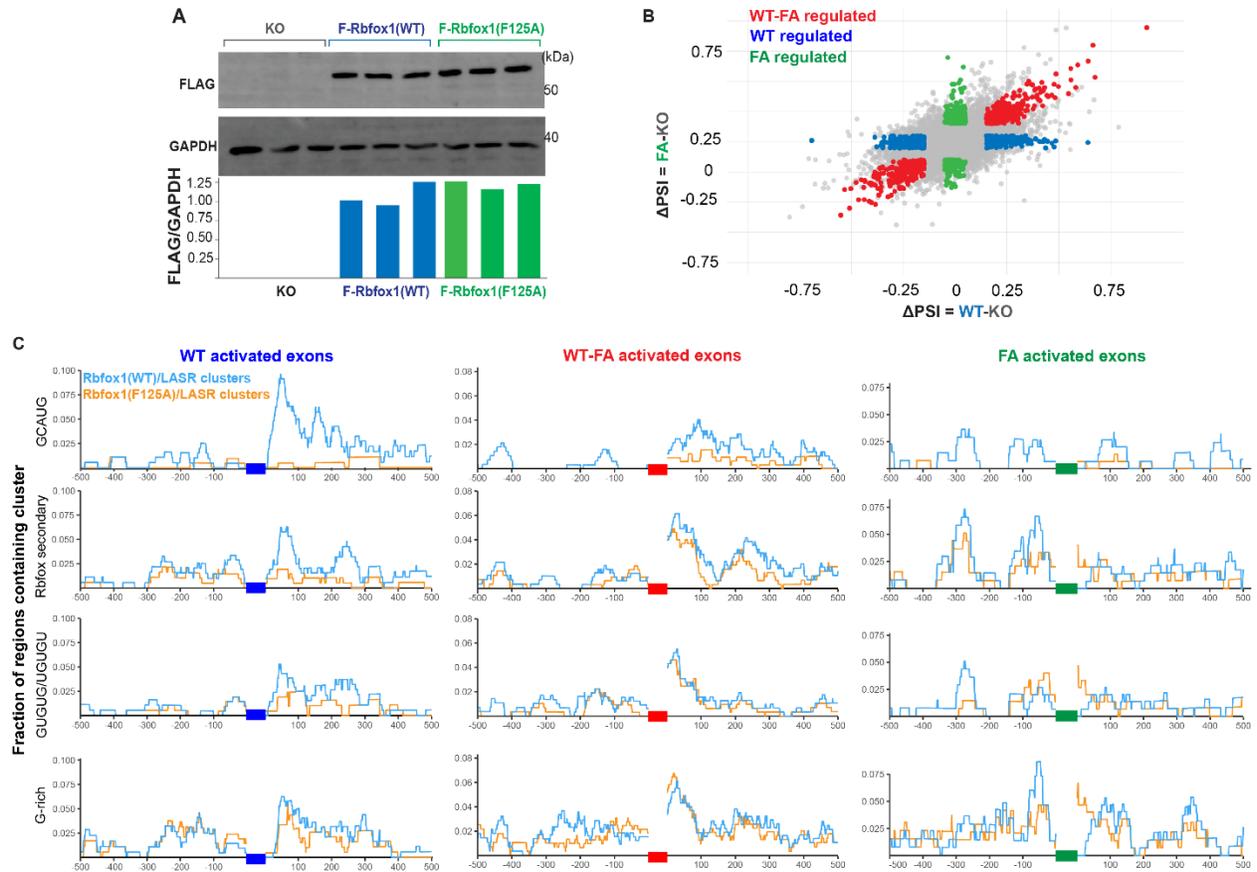
Figure 2.5. Rbfox1(WT)/LASR and Rbfox1(F125A)/LASR each regulate splicing through distinct binding sites adjacent to cassette exons.

A) Immunoblot of FLAG and GAPDH in HEK293 Rbfox deficient, FLAG-Rbfox1(WT), and FLAG-Rbfox1(F125A) expressing cells, B) Comparison of regulated exons in Rbfox1(WT) and Rbfox1(F125A) expressing cells. The x-axis is the ΔPSI calculated from subtracting PSI of the exon in Rbfox1(WT) cells from PSI of the exon in Rbfox KO cells. The y-axis is the ΔPSI from subtracting PSI of the exon in Rbfox1(F125A) cells from PSI of exons in Rbfox KO cells. Exons that have a ΔPSI in both Rbfox1(WT) and Rbfox(F125A) that is greater than 0.15 and are not different from each other by more than 0.15 are considered regulated by both WT and FA and are colored red; exons that

have a ΔPSI >= 0.15 in WT and <= 0.05 in FA cells are considered WT-regulated and

are colored blue; exons that have a ΔPSI >= 0.15 in FA and <= 0.05 in WT cells are

considered WT-regulated and are colored green. All other exons are colored grey. C)

RNA binding map of Rbfox1(WT)/LASR and Rbfox1(F125A) protected sites in 500 bp

upstream and downstream of WT-activated, WT-FA activated, and FA-activated exons.

Clusters are categorized based on the motifs they contain: those with the GCAUG motif,

those with Rbfox secondary motifs (including GCUUG, GAAUG, GCACG, GUAUG,

GUUUG), those with GUGU (GUGUG, UGUGU) motifs, and those with G-rich motifs.

Multipart elements within Rbfox/LASR binding sites have additive effects on exon inclusion

We next examined the regulatory effects of individual motifs within protected sites via mini-gene assays. Exon 16 of the CAMKK2 transcript was previously reported to be regulated by PKA and involved in neurite branching (Cao et al., 2011). This exon is strongly activated by wildtype Rbfox and not the F125A mutant (Fig. 2.6A). The intron downstream of exon 16 contains three clusters with G-rich and GCAUG elements. The most proximal cluster, region 1, contains G-rich elements, and a more distal region 2 contains two GCAUG motifs surrounding a G-rich element. A third protected region, region 3, contained a GCAUG motif. All three protected regions were isolated with Rbfox1(WT); but binding to regions 1 and 2 was greatly reduced, and binding to region 3 eliminated with the F125A mutant.

To assess the regulatory effects of these regions, we leveraged a mini-gene construct containing CAMKK2 exon 16, along with 353 nucleotides of upstream and 324 nucleotides of downstream intronic sequences (Cao et al., 2011). This fragment contained the first two protected regions isolated with Rbfox1/LASR and lacks the third region. We integrated this fragment into a DUP backbone that any GCAUG elements to avoid their possible confounding regulatory effects (Fig. 2.6B). This mini-gene was transiently expressed in the Rbfox$^{-/-}$, Rbfox1(WT), and Rbfox1(F125A) cell lines, with its spliced products assayed by RT-PCR. The splicing of this mini-gene recapitulated that of the endogenous transcript with exon 16 completely excluded from the mRNA in the Rbfox2$^{-/-}$ and Rbfox1(F125A) cells, but strongly included in the Rbfox1(WT) cells (Fraction Spliced In, FSI = 0.6, Fig. 2.6C).

Given that hnRNP H and F have the potential to bind to the G-rich elements in the protected regions, we assessed splicing after siRNA-mediated depletion of these proteins. The siRNA treatment of the Rbfox1(WT) cells reduced hnRNP H/F protein expression by about 70% (Fig. 2.6D). The exon FSI in the control siRNA treatment was 0.33 compared to 0.17 after the hnRNPH/F knockdown. Thus, CamKK2 exon 16 is highly sensitive to hnRNP/F, which stimulate its splicing.

To assess the regulatory contributions of individual motifs present in the two protected regions, we constructed a series of deletion mutants (Fig. 2.6E). G-rich elements were removed from the first or second protected regions, or from both. Loss of the first G-rich element reduced exon 16 splicing from 0.6 to 0.18, while loss of the second G-rich element reduced splicing to 0.39. Double mutation of both G-rich elements further reduced the FSI to 0.06. These two G-rich elements are positive regulators of exon 16 and have additive effects on its splicing.

The GCAUG elements also had effects on exon inclusion. Deletion of the first GCAUG motif decreased FSI from 0.6 to 0.13, while excising the second reduced FSI to 0.43. Consistent with the first GCAUG's stronger regulatory effect, RNA fragments encompassing the first GCAUG are isolated in higher yield than the second, indicating a higher affinity for the Rbfox1/LASR complex. Deleting both motifs together brought the FSI down to 0.03, indicating that although the first GCAUG exerts a stronger regulatory effect, they contribute additively to exon regulation. We also deleted the first G-rich region along with both GCAUG motifs. This led to complete exon skipping, mirroring the splicing observed in Rbfox deficient and Rbfox(F125A) expressing cells. Thus, hnRNP

H/F and Rbfox1, together, provide the splicing enhancement needed to include this CamKK2 exon.

We also assessed the regulatory effects of motifs in protected sites adjacent to an activated exon in *MARK3* (Fig. S2.9). This exon was activated in both the wildtype and F125A expressing cells, with a slightly higher PSI in the F125A (80%) compared to wildtype (66%). The downstream intron of this exon contains three protected sites: the most proximal region 1 contains G-rich and CU-rich elements, the middle region 2 contains GCAUG and CU-rich elements, and the distal region 3 contains GCAUG and GU-rich elements (Fig. S2.9A). The region containing this exon, along with 500 nucleotides upstream and downstream intronic sequences, was cloned into the same DUP mini-gene backbone used for the CAMKK2 exon 16 (Fig. S2.9B).

The splicing pattern of the *MARK3* mini-gene largely recapitulates the splicing of the endogenous gene. The exon is included in both the wildtype and F125A expressing cells and the FSI is higher in the F125A expressing cells, 0.31, compared to the wildtype, 0.16 (Fig. S2.9C). We then deleted potential regulatory motifs in protected sites and tested their effects on splicing in cells expressing wildtype Rbfox1. Some of the elements such as the G-rich element in region 1 and the CU-rich element in region 2 had almost no effect on the inclusion of this exon, yielding an FSI of 0.17 and 0.14 respectively. In contrast, deleting the GCAUG elements in either region 2 or 3 resulted in less inclusion of this exon, 0.08 and 0.09 respectively, demonstrating that interaction of Rbfox/LASR with these elements contributes to the regulation of this exon. Interestingly, the CU-rich element in region 1 had the strongest effect on the inclusion of this exon, with an FSI of 0.02. Thus, in this exon that is regulated by both the wildtype

and the F125A Rbfox1, the element with the strongest regulatory effect is not a Rbfox

binding site and rather an element for another LASR subunit.
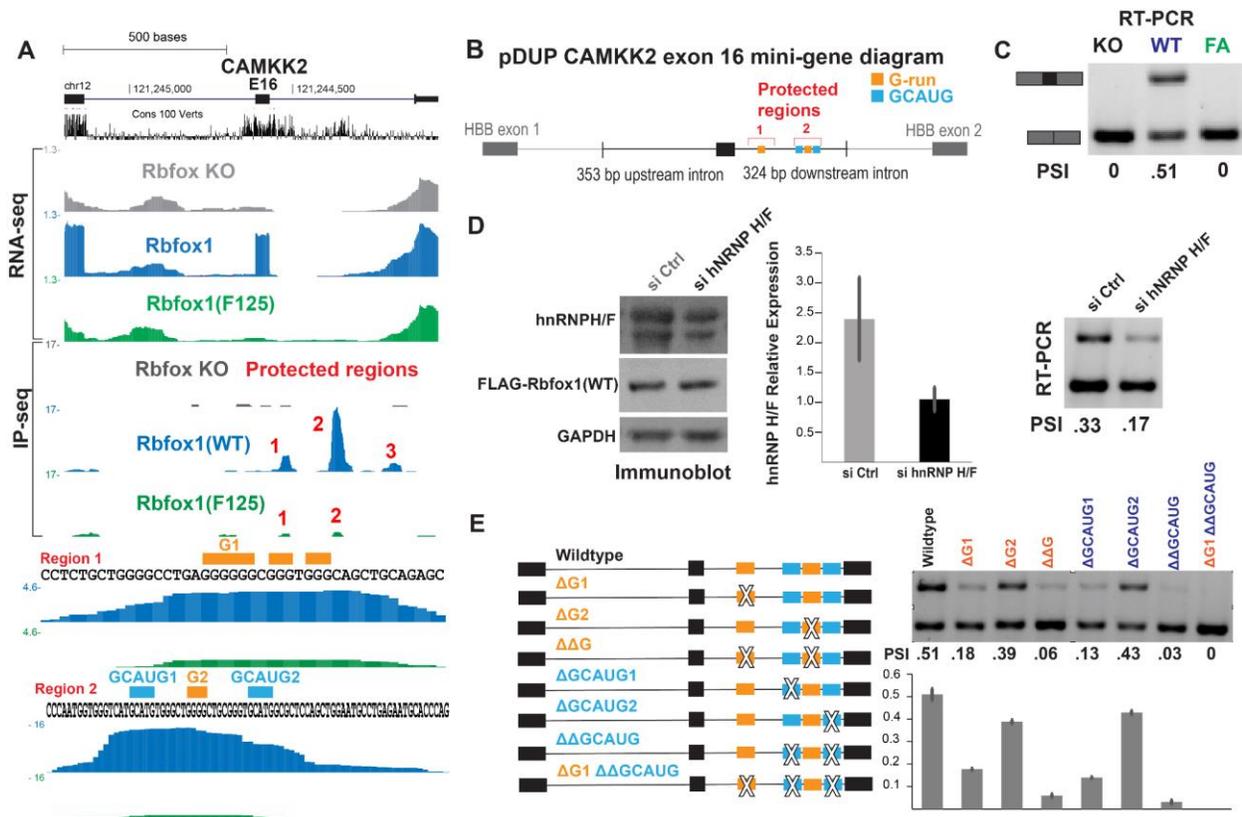
Figure 2.6. Multipart elements within Rbfox/LASR binding sites have additive effects on exon inclusion.

A) Genome browser view of RNA-seq tracks from Rbfox deficient, Rbfox1(WT), and Rbfox1(F125A) expressing cells in the CAMKK2 transcript. IP-seq tracks, displayed below the RNA-seq, show three protected regions downstream of a Rbfox1(WT) activated exons. The sequence features of two of the protected regions are displayed in more detail at the bottom. B) Diagram of CAMKK2 exon 16 DUP mini-gene, C) Agarose gel electrophoresis of RT-PCR of CAMKK2 exon 16 DUP mini-gene, D) Immunoblot of hnRNPH/F siRNA knockdown and agarose gel electrophoresis of CAMKK2 exon 16 DUP-mini gene in these conditions, E) Diagram of deletion mutants of CAMKK2 exon 16 DUP mini-gene along with RT-PCR analysis of mutants.

Discussion

Recognition of multi-element RNA modules by the Rbfox/LASR complex

We previously found that most of Rbfox in the nucleus is associated with nascent RNA and bound to LASR, a protein complex containing hnRNP M, hnRNP H/F, hnRNP C, Matrin3, NF110/NFAR-2, NF45, and DDX5. Furthermore, Rbfox's binding to GCAUG elements and its splicing activity were affected by the LASR subunit hnRNP M. However, at the time it was not clear how RNA gets contacted by Rbfox and LASR. In this study, we found that our preparations of Rbfox1/LASR complexes contained small fragments of nuclease-protected RNA. This RNA was enriched in GCAUG, in addition to motifs for LASR subunits hnRNP M, hnRNP H/F, hnRNP C, and Matrin3. RNA fragments with UAG sequences were also enriched. A well-known binder to this motif, hnRNP A1, is not present in LASR. hnRNP A1 might be lost during the purification. Alternatively, this motif may interact with LASR subunits whose RNA binding properties are less characterized. *In vitro* binding assays, such as RNA Bind N Seq, or *in vivo* approaches, such as CLIP, can unravel binding preferences of yet-to-be characterized LASR subunits.

Complexes of transcription factors often bind to elements of DNA arranged in a particular order (Whitington et al. 2011). The overall structural arrangements in these complexes put constraints on how their DNA binding proteins contact their motifs. We asked if there is a similar optimal arrangement of motifs in RNA that facilitates recruitment of Rbfox/LASR. Analysis of positional distributions of different elements relative to GCAUG shows that motifs of LASR subunits are often proximal to GCAUG, residing within 1-10 nucleotides. GUGUG/UGUGU motifs are often directly upstream of

GCAUG while G-rich elements have a slight preference to be downstream. This preference hints at a spatial arrangement of LASR subunits relative to Rbfox. However, these preferences are smaller than those observed with transcriptional factors. RNA is a more flexible molecule than DNA. This flexibility might allow for accommodation of a wider range of binding modes across a given sequence. Furthermore, hnRNP M and hnRNP H/F all have multiple RNA binding domains. RNA binding domains from the same protein might be positioned in different positions relative to Rbfox, allowing for binding to motifs positioned in various places across a transcript.

In *c. elegans*, the FGFR gene egl-15 is controlled by a downstream UGCAUGGUGUGC element. The GCAUG in this sequence is bound by Rbfox and the GUGUGC element is recognized by the SUP-12 protein (Kuwasako et al., 2014). The structure of a ternary complex of the RNA binding domains of Rbfox, SUP-12, and the UGCAUGGUGUGC sequence shows that Rbfox and SUP-12 can simultaneously contact this RNA and sandwich the G7 nucleotide between their binding domains. SUP-12 does not have any direct paralogues in mammals. However, hnRNP M is a well-known binder to motifs with repeating GU nucleotides. The presence of GCAUG elements directly adjacent to GU-rich elements in both *c. elegans* and humans suggests an evolutionary conserved splicing regulatory module. It would be interesting to assess if exons that have combinations of these motifs in their adjacent introns show any distinct patterns of conservation or tissue-specific expression. Furthermore, a structure of Rbfox and hnRNP M co-bound to a segment of RNA containing GCAUG and a GU-rich element would be interesting to obtain and compare to the Rbfox/SUP-12 structure.

We found that LASR in complex with the RNA binding mutant Rbfox1(F125A) lost binding to almost all binding sites that contained GCAUG. Although it is expected that the F125A mutation would reduce Rbfox's affinity for these sites, resulting in loss of protection, many of these sites also contained motifs for LASR. Therefore, affinity of LASR subunits for these sites is not sufficient for tight binding and protection from the nuclease cleavage. Rbfox likely drives LASR to bind to motifs in these regions through its highly specific and strong interaction with GCAUG. LASR, in turn, specifies Rbfox's binding to GCAUGs that are adjacent to its own motifs. Rbfox/LASR's target recognition, therefore, is shaped by a balance between the highly specific interaction of Rbfox with GCAUG and the more versatile recognition of RNA by LASR. Many Rbfox1/LASR protected sites also entirely lack GCAUGs. We found that for regions that lack GCAUG but have either GU-rich or G-rich elements, multiple copies of the same motif are often placed in tandem. The multiple RNA binding domains present in hnRNP M and hnRNPH/F likely enable them to bind stably to sites with repeated binding elements.

Regulation of alternative splicing by combinatorial interactions between Rbfox and the LASR co-factors

Splicing regulators often form networks. Changes in expression of one regulator can lead to direct splicing changes by binding to cis-regulatory RNA elements and indirect splicing changes by altering the expression of other splicing regulators. The Rbfox proteins coordinate essential splicing programs in neurons, muscle, and other tissues. Their regulatory effects are thought to arise from binding to GCAUG elements near cassette exons. It is also known that Rbfox binds to other protein co-factors.

However, it is not clear how Rbfox's interactions with its co-factors shape splicing programs within cells.

In addition to exon activation by binding downstream GCAUG elements, we find that Rbfox also activates splicing by binding to secondary motifs adjacent to LASR elements. Previous studies have shown that increased Rbfox levels during neuronal development enhance its binding to these secondary motifs (Begg et al., 2020). We show that beyond increased site occupancy due to higher concentrations, Rbfox's association with LASR co-factors can also drive its binding to these secondary motifs. LASR's association with Rbfox, therefore, broadens its regulatory repertoire and allows Rbfox to activate an additional set of exons. It will be important to investigate what physiological processes involve splicing events dependent on regulatory regions containing Rbfox secondary motifs and LASR motifs.

RBFOX genes can express protein isoforms that lack the second half of their RNA Recognition Motif (RRM). The expression of this Rbfox(ΔRRM) is auto-regulated by full-length Rbfox binding to a GCAUG element upstream of an exon that is part of the RRM and repressing it (Damianov & Black, 2010). This isoform produces a stable protein and is thought to act in a dominant negative manner to repress splicing regulation by the full-length isoform. Here, we find that when Rbfox's affinity for GCAUG is reduced, it also influences binding of the LASR complex in the transcriptome. LASR in complex with the F125A Rbfox1 mutant binds to new regulatory sites and results in alternative splicing changes. IP-seq of LASR in complex with Rbfox(ΔRRM) could reveal whether this complex can also bind to novel regulatory sites. By changing the

targets of LASR, the ΔRRM mutant might lead to splicing changes that are beyond its role as a dominant negative actor.

Mini-gene experiments show that regulatory elements adjacent to a cassette exon exert an additive effect. Many of these elements are in discrete protected sites that occur spaced apart from each other. The Rbfox/LASR complex can multimerize into higher-order structures through Rbfox's low complexity tyrosine-rich C-terminal domain. It was previously shown that this multimerization is important for Rbfox's ability to activate splicing. The discrete nature of the protected sites may be due to each site binding a single Rbfox/LASR particle. The larger Rbfox/LASR complex, which is multimerized, could bridge these individual particles across the sites, enhancing the affinity of the RNA binding proteins (RBPs) at various elements. It would be interesting to examine the nuclease-protected RNAs within LASR complexes containing Rbfox variants that lack multimerization capabilities.

Methods

Table 2.1. Antibodies

FLAG	MilliporeSigma	F3165
GAPDH	Invitrogen	PA1-987
hnRNPH/F	Santa Cruz Biotechnology	sc-32310

Cell culture conditions

The growth medium used was 90% DMEM ([+] 4.5 g/L glucose, L-glutamine[-] sodium pyruvate, Corning) and 10% (v/v) fetal bovine serum (Omega scientific). Cells were cultured at 37 °C with 5% CO_2 and monitored for mycoplasma contamination using the PCR-based VenorGeM® Mycoplasma Detection Kit.

Cell lines

Flp-In™ T-REx™ 293 Cell Line (ThermoFisher Scientific) is the parental line for all derived cell lines. As previously described, a Rbfox-deficient line was derived from the parental line by CRISPR/Cas9 deletion of the first constitutive Rbfox2 exon. FLAG-Rbfox1(WT) and FLAG-Rbfox1(F125A) were integrated into the FRT site and a mixed population of each respective line was selected via hygromycin treatment. Cells were treated for 48 hours with doxycycline to induce expression of transgenes.

Plasmid construction

FLAG-Rbfox1(WT): Rbfox1 with a n-terminal 1x FLAG tag was cut out of pcDNA3.1 and ligated into the pcDNA™5/FRT/TO vector via restriction cloning using HindIII-HF and XhoI sites.

FLAG-Rbfox1(F125A): Based on the study by Auweter et al. a critical residue in Rbfox's RBD is phenylalanine 126 in the Swissprot Q9NWB1 entry. This phenylalanine (underline here VSNIP_F_RFRD) is at position 125 in our Rbfox1 construct. We mutated

this phenylalanine to an alanine in the pcDNA5/FRT/TO FLAG-Rbfox1(WT) vector via site-directed mutagenesis PCR using primers Rbfox1(F125A)_F and Rbfox1(F125A)_R.

CAMKK2 exon 16 splicing reporter: the CAMKK2 exon 16 splicing reporter construct used by Cao et al. was obtained from Jiuyong Xie. The region of this construct containing CAMKK2 was subcloned into the DUP-51M1 backbone by restriction cloning via BglII and ApaI sites. This backbone, developed in Damianov et al, has all GCAUG motifs and potential hnRNP M sites mutated. This DUP-51M1 CAMKK2 exon 16 construct was used as the splicing reporter in this study. Deletions of regulatory elements in Figure 4 were done on this mini-gene by site-directed mutagenesis via the following primers: ΔG1 = CAMKK2_ΔG1_F + CAMKK2_ΔG1_R, ΔG2 = CAMKK2_ΔG2_F + CAMKK2_ΔG2_R, ΔGCAUG1 = CAMKK2_ΔGCAUG1_F + CAMKK2_ΔGCAUG1_R, ΔGCAUG2 = CAMKK2_ΔGCAUG2_F + CAMKK2_ΔGCAUG2_R.

Mark3 exon 16 splicing reporter: a region comprising exon 16 along with 522 nucleotides in the upstream and 539 nucleotides in the downstream intron was PCR amplified with the following primers containing ApaI and BglII cut sites: MARK3_F_ApaI, MARK3_R_BglII. This amplified region was cloned into the The DUP-51M1 backbone via restriction cloning using the BglII and ApaI sites. Regulatory elements shown in supplementary figure 9 were deleted via site-directed mutagenesis PCR primers: ΔGU-rich1 = MARK3_ΔGU_rich1_F + MARK3_ΔGU_rich1_R, ΔCU-rich1 = MARK3_ΔCU_rich1_F + MARK3_ΔCU_rich1_R, ΔCU-rich2 = MARK3_ΔCU_rich2_F + MARK3_ΔCU_rich2_R ΔGCAUG2 = MARK3_ΔGCAUG2_F, MARK3_ΔGCAUG2_R, ΔGCAUG3 = MARK3_ΔGCAUG3_F, MARK3_ΔGCAUG3_R.

64

Table 2.2. Oligos	
Name	Sequence
Rbfox1(F125A)_F	CACGGTTCCGGGATCCAGA
Rbfox1(F125A)_R	CAGGGATGTTGGACACATGC
CAMKK2_ΔG1_F	CTGCAGAGCACGGGGGCATCC
CAMKK2_ΔG1_R	GCCCCAGCAGAGGCTACGGC
CAMKK2_ΔG2_F	GCGGGTGCATGGCGCTCC
CAMKK2_ΔG2_R	ACATGCATGACCCACCATTGGGTCC
CAMKK2_ΔGCAUG1_F	TGGGCTGGGGCTGCGGGT
CAMKK2_ΔGCAUG1_R	ATGACCCACCATTGGGTCCAAGCCC
CAMKK2_ΔGCAUG2_F	GCGCTCCAGCTGGAATGCCTGAG
CAMKK2_ΔGCAUG2_R	ACCCGCAGCCCCAGCCCAC
MARK3_F_ApaI	CTGAGGGCCCGCCGAGGTGGGCTGATCACTTG
MARK3_R_BglII	ATGCAGATCTCTGCTACGCAGCTGTTCAC
MARK3_ΔGU_rich1_F	GCTTTCTGGCCCTGTTTTTTC
MARK3_ΔGU_rich1_R	CTAATGTTAATTGCAGGAGCCATTTGGG
MARK3_ΔCU_rich1_F	CCAAACCTAATGTTAATTGCAGGAG
MARK3_ΔCU_rich1_R	GGCCCTGTTTTTTCCTTATAAACTAAAC
MARK3_ΔCU_rich2_F	TACACAGAGCATGCATCAGC
MARK3_ΔCU_rich2_R	GGTAGAATTATTTAGCATCCAAATAATTCTG
MARK3_ΔGCAUG2_F	ATCAGCCTTTTGTTGCATGG
MARK3_ΔGCAUG2_R	CTCTGTGTATTTTCTTTCTTTGGTAGAATTATTTAG
MARK3_ΔGCAUG3_F	TAAGTGTTATTAGTATGAAGACAGAATTATTTG
MARK3_ΔGCAUG3_R	CTCCTGTTTGAAGAATGAGATGC
Dup8a	CTCAAACAGACACCATGCATGG
Dup10	CAAAGGACTCAAAGAACCTCTG

Purification of ribonucleoprotein complexes

Rbfox/LASR complexes were purified as previously described (Damianov et al. 2016). To isolate nuclei, cells were grown to 80-90% confluency and harvested. Cell pellets were resuspended in nine volumes of ice-cold homogenization buffer (10 mM HEPES-KOH pH 7.6, 15 mM KCl, 1 mM EDTA, 1.8 M sucrose, 5% Glycerol, 0.15 mM Spermine, 0.5 mM Sperimidine) and homogenized with the gentleMACS homogenizer. The homogenate was overlaid onto 10 mL of ice-cold cushion buffer (10 mM HEPES-KOH pH 7.6, 15 mM KCl, 1 mM EDTA, 2.0 M sucrose, 10% Glycerol, 0.15 mM

Spermine, 0.5 mM Spermidine) in a SW32Ti ultracentrifugation tube and centrifuged with at 28,100 rpm (96,970 g) for 1 hr at 4°C.

To lyse nuclei and obtain an extract from the chromatin containing pellet, supernatant and cushion buffer were discarded and the pelleted nuclei were suspended in 1 mL of nuclear resuspension buffer (10 mM HEPES-KOH pH 7.6, 15 mM KCl, 1 mM EDTA, 0.15 mM spermine, 0.5 mM spermidine) and aliquoted into three separated tubes of equal volume. The samples were centrifuged for 5 mins at 1k rcf, the supernatant was discarded, and the nuclei were resuspended in at least 10x volume of nuclear lysis buffer (20 mM HEPES-KOH pH 7.6, 150 mM NaCl, 1.5 mM MgCl2, 0.5 mM DTT, 1x protease inhibitors, and 0.6% Triton X-100) and kept on ice for 5 mins for nuclear lysis. This lysate was then centrifuged at 20k rcf for 10 mins and the supernatant was kept as the nucleoplasm portion. The same volume as the nucleoplasm portion of the nuclear lysis buffer was added to the pellet and Benzonase nuclease was added to all nucleoplasm and pellet-containing samples to a final concentration of 5 units/ul. The nuclease digestion was done until the pellet could be resuspended by a P200 ul tip. The nuclease treated samples were then centrifuged at 20k rcf for 10 mins and the supernatant was kept and the pellet was discarded.

To perform immunoprecipitations, the supernatant was added to 7.5 ul of packed M2 FLAG agarose beads (Sigma) and this mixture was rotated overnight at 4°C. The beads were washed 5x with each wash containing 1 mL of wash buffer (20 mM HEPES-KOH pH 7.6, 150 mM NaCl, and 0.05% Triton X-100). 50 ul of elution buffer (20 mM HEPES-KOH pH 7.9, 150 mM NaCl, and 150 μg/ml of 3xFLAG peptide) was then added

to the beads and this mixture was agitated intermittently (15 sec on, 4:45 min off) at 1100 rpm at 4°C for 1 hr. Supernatant was saved as eluate for further processing.

SDS-PAGE analysis of purified complexes

For visualization of immunoprecipitated Rbfox/LASR complexes, 20% of eluate of immunoprecipitated material was denatured and run on a NuPAGE® Novex 4-12% Bis-Tris gel. The gel was then stained with SYPRO Ruby (ThermoFisher Scientific) overnight following the manufacturer's instructions. Stained gels were imaged with the Amersham Typhoon.

Urea-PAGE analysis and sequencing of nuclease-protected RNA

For all phenol-chloroform extractions, the aqueous phase was separated from the organic phase using Phase Lock Gel Heavy tubes (QuantaBio). All ethanol precipitations were done overnight at -20°C and glycoblue was used. 40% of the immunoprecipitated material was deproteinized with Proteinase K and an acid-phenol chloroform extraction followed by ethanol precipitation was done. This material was then DNAse treated and dephosphorylated in a one-pot reaction containing TurboDNAse and FastAP Thermosensitive Alkaline Phosphatase. This material was then treated with Proteinase K followed by a phenol-chloroform extraction and ethanol precipitation. 5% of this material was used for end-labeling and the rest was used to make sequencing libraries. T4 PNK was used to ligate γ-[32P] ATP (Perkin Elmer) to 5' end of the RNA. This material was denatured with formamide and run on a 10% Urea-PAGE. The gel was dried and used to generate an autoradiograph which was imaged with the Amersham Typhoon. The other 95% of the material was used to prepare sequencing

libraries using a modified iCLIP library preparation protocol (Damianov et al. 2024). This library was sequenced with the NovaSeq 6000 SP 2x100bp.

Sequencing RNA for splicing analysis

Total RNA was extracted using Trizol. This RNA was purified via Zymo's RNA Clean & Concentrator-5 Kit, including a DNAseI treatment (R1013). cDNA was prepared from poly-A selected RNA using the Illumina TruSeq kit. This cDNA was sequenced in one lane of the NovaSeq X.

For assessing expression of Rbfox proteins in the RNA-seq experiment, 50% of harvested cells were lysed with RIPA. Total protein concentrations were measured with Pierce BCA kit and the same amount of total protein was denatured and analyzed per sample with 10% SDS-PAGE. Immunoblotting was done by transferring the samples to a PVDF membrane and probing with FLAG and GAPDH primary antibodies and fluorescent secondary antibodies.

Mini-gene analysis

All transfections and the subsequent steps were done in triplicate for all the constructs tested. A mixture of lipofectamine 2000 and the construct of interest was added to cells in complete growth medium and incubated for 6 hrs at 37 °C. The media was then changed to fresh complete growth medium and cells were incubated and harvested the next day after 24 hrs of being transfected. Total RNA was extracted as described in the "RNA-sequencing for splicing analysis" section. 4.5 ug of this RNA was reverse transcribed via SuperScript™ (Invitrogen) and an oligo with 20 dT nucleotides. PCR amplifications were done on this cDNA using Dup8a and Dup10 primers via the GoTaq Green Master Mix (Promega). Amplified material was run on a 2% agarose gel,

stained with SYBR Gold, and imaged via the iBright Imaging System (ThermoFisher Scientific).

siRNA knockdown of hnRNPH/F

For hnRNPH/F siRNA knockdowns, siHNRNPH/F or siCONTROL were used to transfect cells with Lipofectamine RNAiMAX (Invitrogen) per manufacturer's protocol. After 5 hours, medium was change to complete growth medium and cells were incubated for 24 hours. Another transfection was then done with siHNRNPH/F or siCONTROL using lipofectamine 2000 with both conditions including the DUP-51M1 CAMKK2 exon 16 construct. After 5 hours, medium was change to complete growth medium and cells were incubated for 48 hours. Cells were subsequently harvested and split into two equal portions to be used for RNA extraction and RIPA lysis followed by immunoblotting.

Mapping IP-seq sequencing data and defining clusters

Samples were demultiplexed, PCR duplicates were removed, and the reads were mapped to hg38 using STAR (Dobin et al., 2013). Yodel was used to define clusters in regions containing at least 10 reads in merged replicates. RPKM values of clusters were calculated with the SeqMonk software (v1.45.4, Babraham Institute). A chi-squared test was performed comparing RPKM of size-matched clusters from the experimental to the control IP-seq clusters. All clusters with an FDR < 0.05 and $\log_2$(sample/control) > 0 were called as significant and further processed.

Analysis of clusters in different gene types and genomic regions

Clusters were annotated based on Ensembl Canonical genes annotations. Based on the annotations, the percent of clusters that fell into different gene types and regions

were determined. To define clusters in introns of protein-coding genes, intronic sites that also resided in regions annotated as snRNA, miRNA, scaRNA, snoRNA, ncRNA, and lncRNA in Ensembl canonical genes or NCBI Refseq databases were discarded.

Motif enrichment analysis

Reads that fall into clusters in introns of protein-coding introns were extracted. A background set was generated by sampling random regions from introns that contained the clusters. The background set of reads were 4x higher than the experimental. HOMER was used for motif analysis via the following command: findMotifsGenome.pl <experimental set> hg38 <output file> -bg <background set> -rna -len 4,5,6 -S 10 -size given.

Enrichment of pentamers and hexamers was determined by comparing their frequency in clusters within introns of protein-coding genes to randomly sampled regions from introns that contain clusters. Z-scores were determined for these pentamers and hexamers. Pentamers having a z-score greater than 2 standard deviations above the mean were considered significant.

Determining protected vs unprotected GCAUGs

The SeqKit software was used to calculate coordinates for all (U)GCAUG, UGCAUG, and GCAUG (without a 5' U) elements (Shen et al., 2016). Expressed elements were defined as those present in introns that contained at least one IP-seq cluster. We then determined the protection status of these elements by assessing their overlap with IP-seq clusters. The percent of protected versus unprotected elements was calculated accordingly.

Motif co-occurrence analysis

To determine co-occurrence of GCAUGs with other motifs, a region of 50 nucleotides upstream and 50 nucleotides downstream of each GCAUGs was defined. The number of these GCAUG adjacent regions that contained at least one motif falling into a category was then determined. The percent co-occurrence was determined by dividing the total number of GCAUGs with a motif in their adjacent region vs total number of GCAUGs.

For quantifying how many protected fragments containing GCAUGs also contained another motif, protected fragments from clusters in introns of protein-coding genes were annotated with HOMER. The co-occurrence of GCAUG with other motif types was then determined based on this annotation. Furthermore, these annotated protected fragments were used to generate plots of the distribution of length of reads associated with different types of motifs.

To generate plots displaying the positional frequency of motifs around GCAUGs, GCAUG was considered to be at position 0. If a motif was found starting at 1 nucleotide downstream of a GCAUG it was considered to be at position 1. Conversely, if a motif was found starting at 1 nucleotide upstream of a GCAUG, it was considered to be at position -1. The number of times a motif occurred at each position was then determined relative to GCAUGs for all motifs in each motif category. Then the total number of times a motif occurred at a given position was divided by the total number of GCAUGs to determine the fraction of regions with a given motif type. A plot was then generated to display this fraction for different motif types surrounding unprotected and protected GCAUGs separately.

Differential binding analysis via DESeq

All significant clusters of Rbfox1(WT)/LASR and Rbfox1(F125A)/LASR were merged. To avoid merging overlapping clusters into a single bigger cluster, the clusters were shortened to only include the region that is covered by 50% of the total reads within clusters. DEseq analysis was performed on this merged set of clusters comparing wildtype to F125A IP-seq. Differentially bound sites were defined as the following: WT enriched sites have a log2FoldChange (WT/FA) >=1 and FDR < 0.05, FA enriched sites have a log2FoldChange (WT/FA) <= -1, common sites have a log2FoldChange (WT/FA) > -0.1 and < 0.1.

Splicing analysis of sequenced RNA via rMATS

Alternative splicing was analyzed by rMATS. Pairwise comparisons were done for WT vs KO, FA vs KO, and WT vs FA. Exons that had a $|\Delta PSI| >= 0.15$, FDR <= 0.05, and at least 5 reads in each replicate were considered regulated. Exons considered co-regulated by WT or FA also had a $|\Delta PSI(WT-KO) - \Delta PSI(FA-KO)| <= 0.15$.

Mapping clusters around regulated exons

Regulated exons as defined above that had a $\Delta PSI >= 0.15$ were defined as activated and with a $\Delta PSI =< -0.15$ were defined as repressed. Upstream and downstream intronic regions were defined by the region spanning the cassette exon and its adjacent exons as determined by rMATS. Clusters that occurred in these regions were extracted and split according to the elements they contained. The occurrence of clusters containing elements of interest were assessed at every single nucleotide in both the upstream and downstream introns. The total number of clusters that occurred at a given nucleotide were summed and divided by the total number of activated exons to give the fraction of regions containing clusters with an element of interest.

Supplementary data

A   **Summary of IP-seq sequencing information**

Barcode	barcode	reads	unique reads	unique mapped reads	% mapped	average read length
Rbfox1(WT)_rep1	CAAT	16,506,680	11,281,535	7,038,416	62.39%	29
Rbfox1(WT)_rep2	TTAG	26,142,640	18,138,879	11,747,732	64.77%	32
Rbfox1(WT)_rep3	AATG	32,725,995	22,771,342	14,808,411	65.03%	32
Control	ACCT	14,049,657	6,851,926	4,646,956	67.82%	38
Rbfox1(F125A)_rep1	GGTT	39,147,666	24,409,329	15,741,834	64.49%	34
Rbfox1(F125A)_rep2	TTGT	17,956,165	11,415,104	6,879,980	60.27%	32

Supplementary Figure 2.1. Sequencing, mapping, and read length statistics of RNA fragments isolated from control and Rbfox1/LASR preparations.

A) Table summarizing read information. The barcodes are for sample demultiplexing. Reads are the total number of sequencing reads associated with each sample. Unique reads are how many reads are left after removing PCR duplicates. Unique mapped reads are the reads after mapping with STAR and the % mapped is also included. The average read length for each sample is also included. B) Genome browser view of control and FLAG-Rbfox1 IP-seq tracks across the Mark3 gene. C-D) Distribution of IP-seq clusters across different types of genes (C) and genomic regions (D).

A   **Clusters with GCAUG**   **B**   **Clusters without GCAUG**

eCLIP
4792

101946
IP-seq

Common
5444

eCLIP
52765

353512
IP-seq

Common
11855
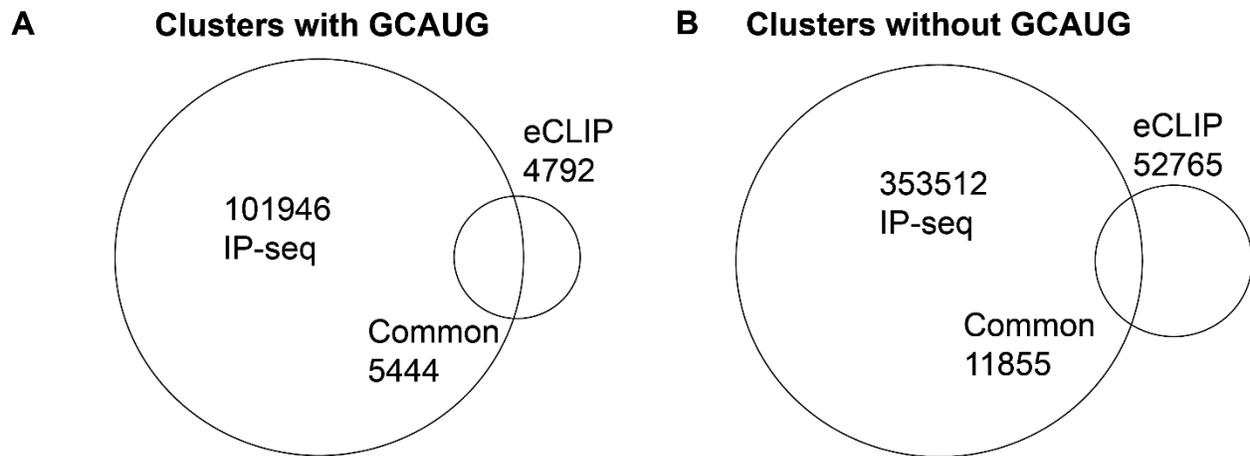
Supplementary Figure 2.2. Comparison of Rbfox1 IPseq to Rbfox2 eCLIP in

HEK293 cells.

A-B) Venn diagram showing overlap of IP-seq and eCLIP clusters with GCAUG (A) and

without GCAUG.

Top motifs

Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background	STD(Bg STD)
1	UGCAUG	1e-2902725	-6.684e+06	24.58%	1.41%	9.6bp (17.4bp)
2	GUGUGU	1e-135923	-3.130e+05	21.14%	13.27%	11.9bp (21.7bp)
3	GGGGG	1e-129272	-2.977e+05	10.97%	5.54%	8.0bp (20.8bp)
4	GUUGUU	1e-104712	-2.411e+05	11.12%	6.09%	10.0bp (19.2bp)
5	GUUAGU	1e-85255	-1.963e+05	16.36%	10.70%	10.1bp (17.5bp)
6	UCUUUU	1e-84654	-1.949e+05	15.87%	10.32%	11.4bp (17.2bp)
7	AGCAUA	1e-53684	-1.236e+05	5.80%	3.15%	9.9bp (15.5bp)

GCAUG similar motifs

Rank	Match Score	Redundant Motif	P-value	log P-value	% of Targets	% of Background
1	0.916	GCAUG	1e-2324508	-5352412.743078	29.92%	3.73%
2	0.917	UGCAUG	1e-1838744	-4233892.981247	29.82%	5.06%
3	0.820	GCAU	1e-900306	-2073044.876245	33.06%	11.84%
4	0.781	AGCAUG	1e-139101	-320295.029771	3.46%	0.81%
5	0.672	GCUUG	1e-49667	-114364.227062	7.34%	4.40%
6	0.672	GAAUG	1e-22751	-52388.575541	5.73%	3.90%
7	0.672	GCACG	1e-17376	-40011.010945	1.11%	0.48%

GUGUGU similar motifs

Rank	Match Score	Redundant Motif	P-value	log P-value	% of Targets	% of Background
1	0.901	UGUGU	1e-19051	-43867.799182	19.88%	16.78%
2	0.804	GUGU	1e-200	-460.611932	30.39%	30.01%

GGGGG similar motifs

Rank	Match Score	Redundant Motif	P-value	log P-value	% of Targets	% of Background
1	0.888	GGGGGG	1e-114179	-262908.819142	23.88%	16.17%
2	0.791	GGUGG	1e-93792	-215965.706949	19.30%	12.90%
3	0.806	GAGGG	1e-78460	-180663.541507	15.97%	10.58%
4	0.823	GGGG	1e-24763	-57020.242126	21.85%	18.20%
5	0.656	GGGU	1e-22157	-51019.631909	27.16%	23.40%

GUUGUU similar motifs

Rank	Match Score	Redundant Motif	P-value	log P-value	% of Targets	% of Background
1	0.892	UUGUU	1e-97294	-224029.712005	39.31%	30.68%
2	0.649	CUUCUU	1e-80604	-185600.547109	13.38%	8.39%
3	0.715	UCUU	1e-44083	-101507.188465	30.76%	25.29%
4	0.800	UUGU	1e-1267	-2918.404351	41.55%	40.52%

UAG similar motifs

Rank	Match Score	Redundant Motif	P-value	log P-value	% of Targets	% of Background
1	0.952	GUUAG	1e-67261	-154876.723780	9.67%	5.79%
2	0.787	GUUA	1e-6545	-15071.166173	14.47%	12.86%

Supplementary Figure 2.3. Sequence motifs enriched in Rbfox1(WT)/LASR bound RNAs identified by Homer.

All motifs of 4-6 nucleotides in protected fragments in clusters within protein-coding introns identified as significant by HOMER.
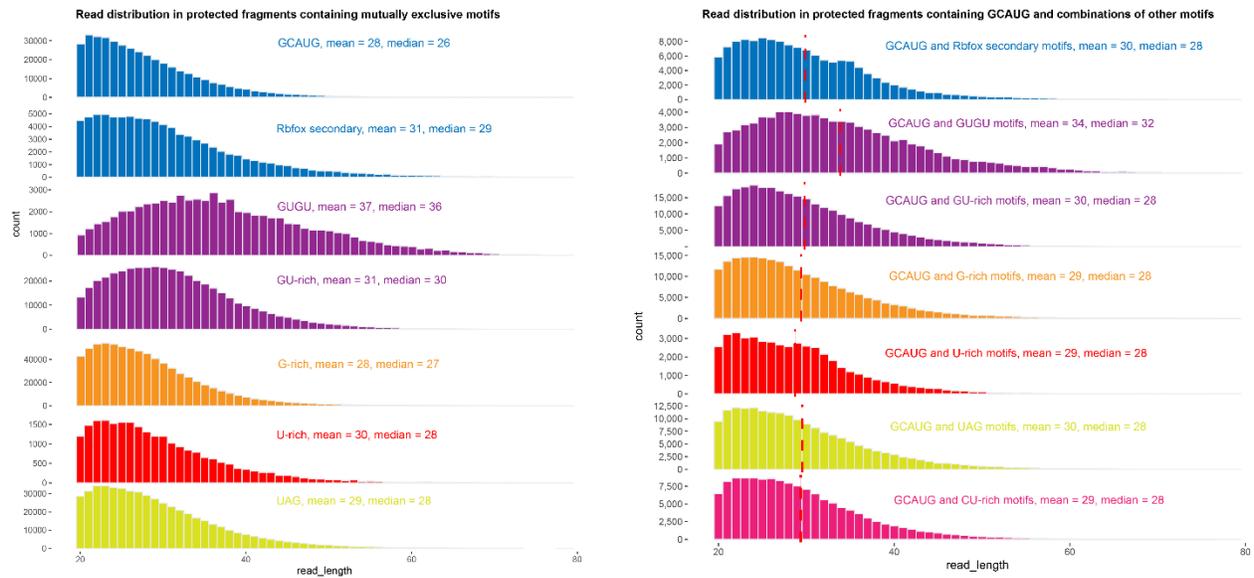
Supplementary Figure 2.4. Length distributions of protected fragments containing different motifs.

Distribution of length of protected fragments from clusters in introns of protein-coding genes. containing different combinations of motifs. The first column contains distribution of length of reads for protected fragments that only contain the motif specified. The second column contains distribution of length of reads for fragments that contain GCAUG and another motif as specified on the plot.
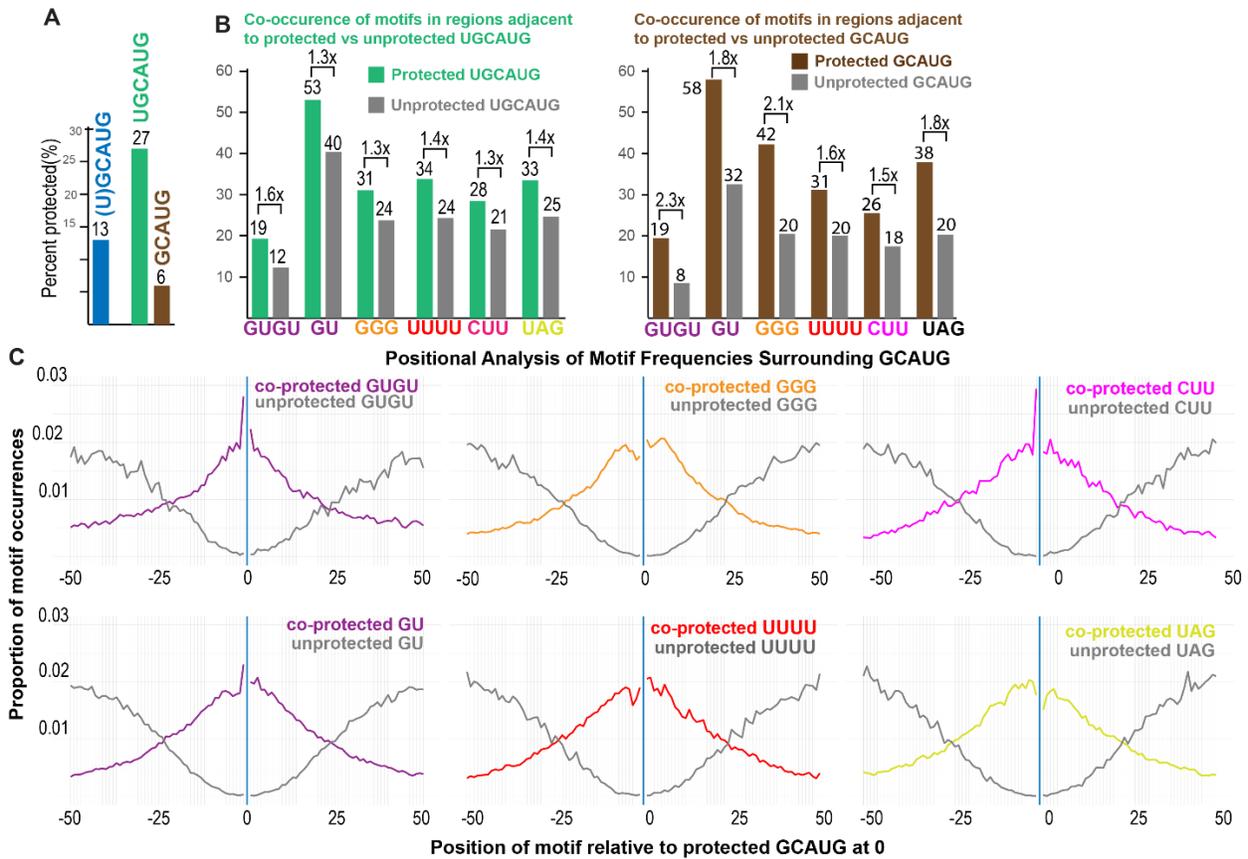
Supplementary Figure 2.5. Co-occurrence of GCAUG with other motifs.

A) Fraction of (U)GCAUG, UGCAUG, and GCAUG protected by Rbfox/LASR. B) Co-occurrence of motifs in regions surrounding protected and unprotected UGCAUG (green) and GCAUG (brown). C) Frequency of motifs in positions that are co-protected and unprotected relative to GCAUG

Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1	[GGGGGG motif logo]	1e-275645	-6.347e+05	22.11%	7.04%	8.0bp (22.4bp)	hsa-miR-3679-3p MIMAT0018105 Homo sapiens miR-3679-3p Targets (miRBase)(0.704) More Information | Similar Motifs Found	motif file (matrix)
2	[UAGUU motif logo]	1e-62025	-1.428e+05	21.18%	12.94%	10.7bp (17.8bp)	hsa-miR-606 MIMAT0003274 Homo sapiens miR-606 Targets (miRBase)(0.696) More Information | Similar Motifs Found	motif file (matrix)
3	[CUUCUU motif logo]	1e-60638	-1.396e+05	17.64%	10.20%	12.0bp (17.4bp)	hsa-miR-3653 MIMAT0018073 Homo sapiens miR-3653 Targets (miRBase)(0.809) More Information | Similar Motifs Found	motif file (matrix)
4	[GUGUGU motif logo]	1e-53501	-1.232e+05	12.70%	6.79%	12.4bp (23.1bp)	hsa-miR-603 MIMAT0003271 Homo sapiens miR-603 Targets (miRBase)(0.798) More Information | Similar Motifs Found	motif file (matrix)
5	[GUUGUU motif logo]	1e-37463	-8.626e+04	7.23%	3.52%	10.4bp (19.6bp)	hsa-miR-548x MIMAT0015081 Homo sapiens miR-548x Targets (miRBase)(0.632) More Information | Similar Motifs Found	motif file (matrix)
6	[GGGA motif logo]	1e-12623	-2.907e+04	27.74%	23.31%	10.3bp (21.0bp)	hsa-miR-4281 MIMAT0016907 Homo sapiens miR-4281 Targets (miRBase)(0.675) More Information | Similar Motifs Found	motif file (matrix)
7	[CUGCUG motif logo]	1e-11233	-2.587e+04	2.43%	1.22%	13.1bp (17.6bp)	hsa-miR-4291 MIMAT0016922 Homo sapiens miR-4291 Targets (miRBase)(0.843) More Information | Similar Motifs Found	motif file (matrix)

Rank	Match Score	Redundant Motif	P-value	log P-value	% of Targets	% of Background	Motif file
1	0.940	[motif logo]	1e-183149	-421720.715910	33.22%	17.06%	motif file (matrix)
2	0.960	[motif logo]	1e-140089	-322569.752498	15.85%	6.13%	motif file (matrix)
3	0.868	[motif logo]	1e-128996	-297026.698933	19.38%	8.77%	motif file (matrix)
4	0.637	[motif logo]	1e-108530	-249902.193629	37.50%	23.96%	motif file (matrix)
5	0.680	[motif logo]	1e-69144	-159211.370543	19.49%	11.21%	motif file (matrix)
6	0.736	[motif logo]	1e-45153	-103970.235658	7.16%	3.20%	motif file (matrix)
7	0.788	[motif logo]	1e-35073	-80761.266804	33.99%	26.24%	motif file (matrix)

Rank	Match Score	Redundant Motif	P-value	log P-value	% of Targets	% of Background	Motif file
1	0.749	[GAUAG motif logo]	1e-35305	-81294.884454	15.88%	10.29%	motif file (matrix)
2	0.766	[UAGG motif logo]	1e-17503	-40302.862793	20.51%	15.93%	motif file (matrix)

Rank	Match Score	Redundant Motif	P-value	log P-value	% of Targets	% of Background	Motif file
1	0.861	[UUGUU motif logo]	1e-47805	-110077.482251	35.87%	26.74%	motif file (matrix)
2	0.811	[UCUU motif logo]	1e-19025	-43808.940053	27.83%	22.43%	motif file (matrix)
3	0.686	[CUCUU motif logo]	1e-10714	-24672.083443	5.64%	3.71%	motif file (matrix)
4	0.682	[CUGCU motif logo]	1e-3291	-7578.802987	3.98%	3.03%	motif file (matrix)

Rank	Match Score	Redundant Motif	P-value	log P-value	% of Targets	% of Background	Motif file
1	0.778	[GUUUGU motif logo]	1e-15731	-36224.398021	7.68%	5.00%	motif file (matrix)
2	0.816	[GUGU motif logo]	1e-226	-520.804888	31.41%	30.77%	motif file (matrix)

Rank	Match Score	Redundant Motif	P-value	log P-value	% of Targets	% of Background	Motif file
1	0.730	[GGUUG motif logo]	1e-11366	-26172.057998	10.10%	7.41%	motif file (matrix)
2	0.816	[GUUG motif logo]	1e-10165	-23406.980504	26.69%	22.75%	motif file (matrix)

Supplementary Figure 2.6. Motifs enriched in Rbfox1(F125A)/LASR IP-seq fragments identified by Homer.

All motifs of 4-6 nucleotides in protected fragments in clusters of Rbfox1(F125A)/LASR within protein-coding introns identified as significant by HOMER.
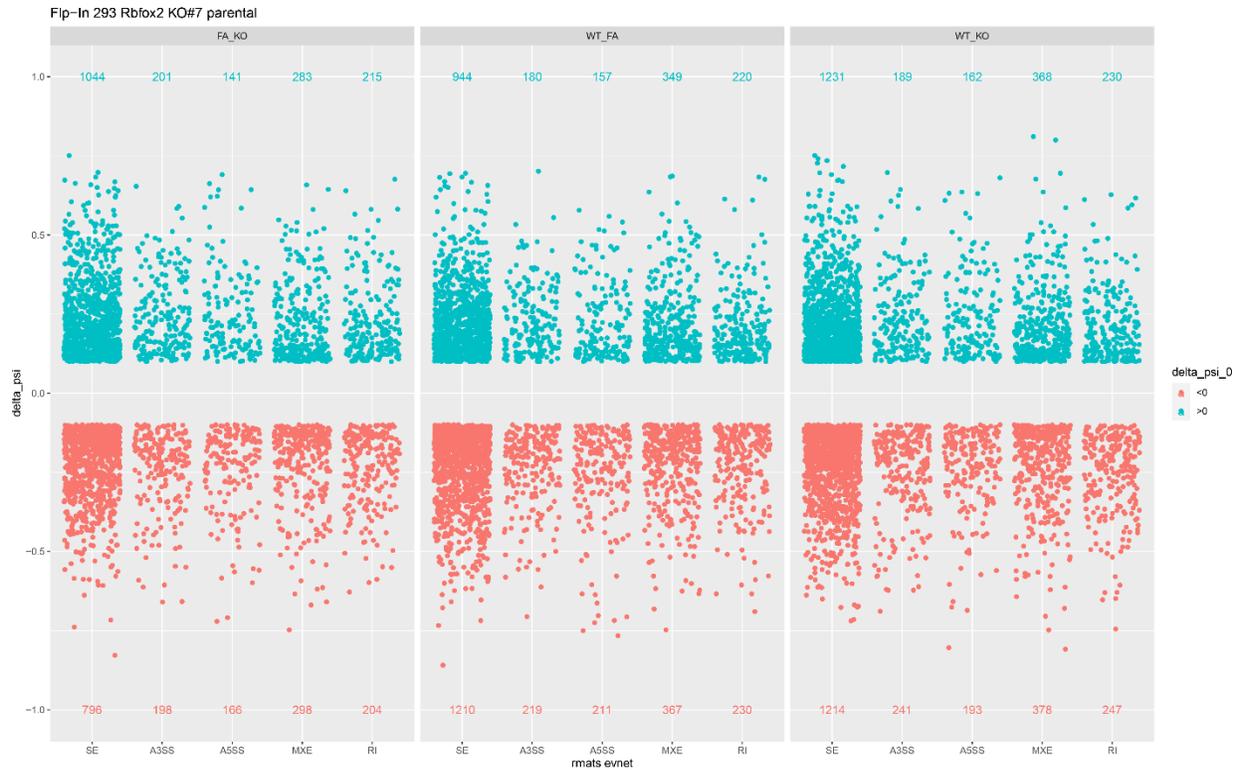
Supplementary Figure 2.7. rMATS analysis of splicing changes induced by Rbfox1 wildtype and Rbfox1(F125A).

A) rMATS analysis of sequenced RNA from Rbfox(WT) expressing, Rbfox1(F125A) expressing, and Rbfox-deficient cells (KO). Events shown all have a ΔPSI >=0.1, FDR <=0.05, total reads across two conditions >=50. Events with ΔPSI <=0.05 or >=0.95 were removed.
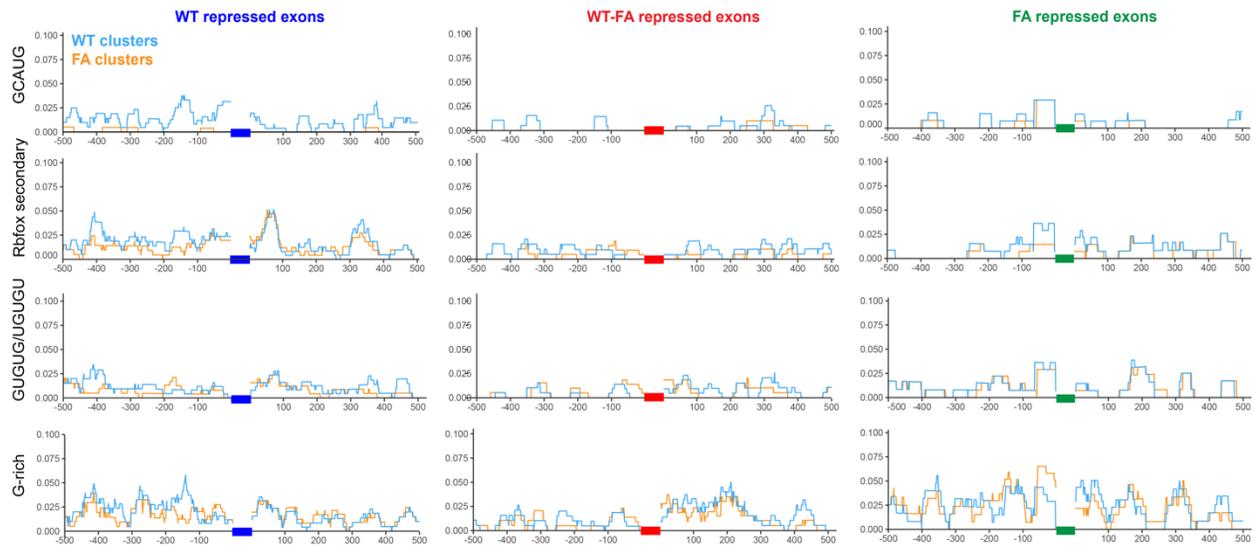
Supplementary Figure 2.8. RNA binding map of Rbfox1(WT)/LASR and Rbfox1(F125A) protected sites in 500 bp upstream and downstream of WT-repressed, WT-FA repressed, and FA-repressed exons

A) Clusters are categorized based on the motifs they contain as described in figure 5. The total number of clusters aligning at each nucleotide was calculated and divided by the total number of exons in their respective categories to obtain the
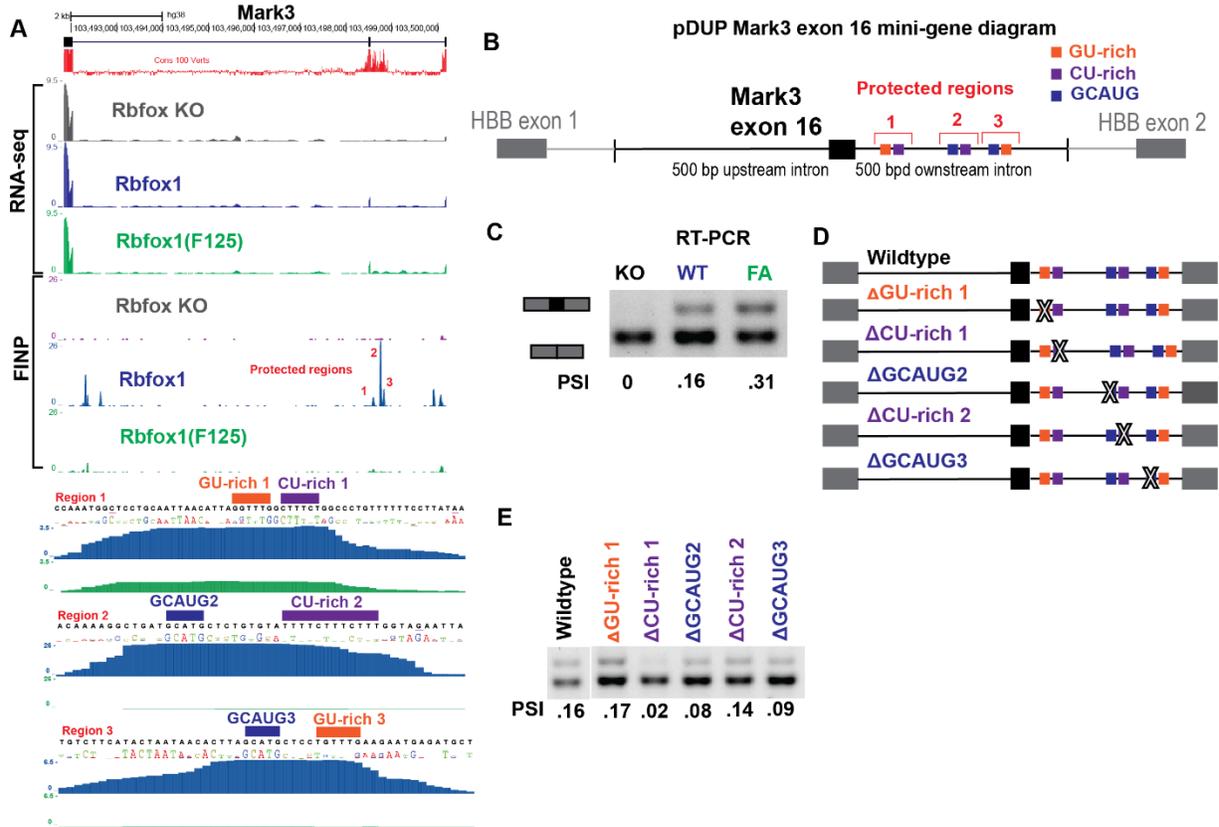
Supplementary Figure 2.9. Mini-gene analysis of Mark3 exon 16 splicing.

A) UCSC genome browser view of Mark3 exon 16 and its upstream and downstream intron and exonic regions including RNA-seq tracks and IP-seq tracks. Protected regions downstream of this exon are delineated in the figure and detailed sequence composition is shown below. B) Diagram of DUP mini-gene containing Mark3 exon 16 and 500 bp upstream and 500 bp downstream intronic regions. C) RT-PCR of Mark3 DUP mini-gene in Rbfox-deficient, Rbfox1(WT) expressing, and Rbfox1(F125A) expressing cells. D) Diagrams of deletion mutagenesis in Mark3 DUP mini-gene. E) RT-

PCR of mini-gene deletion mutants.

Table 2.3. Enriched hexamers

hexamer	Z-score	hexamer	Z-score	hexamer	Z-score
UGCAUG	910.5486	UGGGUG	214.37959	GGUUAG	155.79566
GUGUGU	618.4542	GCAUGA	211.21197	UAUGUG	155.55832
UGUGUG	615.514	CUUUUU	211.00306	UGUCUU	155.07912
UUUGUU	614.4124	GGUGGU	210.37185	UUAGUG	154.77173
UUGUUU	535.0437	UUAGUU	209.17559	UGAGGG	153.88799
GCAUGU	520.3652	GUAGGG	209.05785	GGAUGG	153.87317
GGGGUG	460.3296	GGUAGG	207.24814	GGUUGU	153.59926
GGGUGG	451.1739	GGGAGG	207.2155	UGCUUG	153.58753
UGUUGU	445.35	GUAUGU	205.69331	GUUUAG	153.3041
UGUUUG	431.09	UAGUGU	204.45713	GGUGUG	153.23649
GUUUGU	430.7146	GCAUGG	203.41763	UAGGUA	152.31709
GUUGUU	429.7802	UUCUUC	201.85639	UUUGGG	152.25028
GGUGGG	421.9362	GGUUGG	201.56562	GAGGGA	152.17777
UUCUUU	417.1889	UCUUUC	201.28289	UCUUUG	151.72763
UUUCUU	413.657	AGGUAG	200.6577	UAGCUU	150.20017
UGGGGG	408.8533	UUUGGU	200.05073	UUAGGG	149.915
UUGUUG	408.8463	UUUUUC	197.60083	GUAGUG	149.53153
UGUUUU	402.4877	GUUAGU	196.75718	UUGUUA	149.24837
GUGGGG	387.6117	GAGGGU	195.90361	UCUCUU	148.85082
UUUUGU	378.2952	UGUGUA	194.54951	CUUGUU	148.69779
GUUUUU	374.9262	UAGGUU	190.03351	UGUUCU	148.67597
GGGGGU	358.1141	GUUGGG	188.37356	UUUAGU	148.25585
GUUUUG	337.2124	UUAGGU	188.28427	GUCUUU	147.74085
UCUUUU	328.6412	GUUUGG	187.74257	GUGUUG	147.0623
UUUUUG	317.5502	GGGUAG	185.4495	AGUUAG	146.75235
GCAUGC	315.4275	GUAGUU	185.38916	UGGUGG	145.72813
GGAGGG	314.974	GGUUUG	181.28683	AGGGGU	143.59097
UUUUCU	308.1112	AGGGGG	180.10684	UGUAGU	143.50158
GGGGGG	305.4172	GGGGUU	179.11923	GGGGAU	142.71912
GGGGAG	300.1102	UGGUUU	178.25857	GGGGCG	141.16748
GUGCAU	298.8997	UGGGGA	177.35009	UGGUUG	141.05974

GGGGGA	295.8128	AGGGUG	177.01634	AUUUGU	137.70406
UGUGUU	293.8339	UUUAGG	176.35861	AUGUUU	136.88401
GUGGGU	278.2834	CUUUCU	176.0963	AGGGAG	136.85629
UUGUGU	273.1116	GUGGGA	175.72512	GCGGGG	136.46884
UUGGGG	271.7384	GUUGUG	174.3752	UAGUUG	136.06086
UUGCAU	265.4616	AUGCAU	174.07255	UUGUUC	135.8931
GGGUGU	254.7055	UGUAGG	173.16018	UUGCUU	135.76518
UGUGGG	254.2574	GGGAUG	172.7722	CUUUGU	134.34345
UUGGUU	249.2252	GUUAGG	172.28529	GCUUUU	133.7989
AUGUGU	248.8	UGUUAG	171.73165	CGGGGG	133.70695
UUUGUG	240.383	AGCAUG	170.05667	GGGUUU	133.42023
UGGGGU	239.6743	UCUUGU	169.3378	GGGGGC	133.06297
GUGUUU	238.4204	UUGGGU	169.0445	UAGGGG	132.15111
GUUCUU	238.4045	CAUGUG	168.62104	GCUUGU	131.6767
UAGUUU	234.154	GGUAGU	167.2867	AUGGGU	131.32235
CUUCUU	231.784	UAGGGU	167.16055	UCUUAG	130.82037
UCUUCU	230.0058	GUGUGG	166.98613	GUGUGC	130.34798
UGUAUG	228.1947	UGGGUU	162.74319	GUGGUU	130.0526
GGGUUG	228.1278	GUAGGU	162.74142	GGUGUU	127.55886
GGUUUU	226.8295	UUGUAG	162.13488	UGUUUA	127.08005
GUGUAG	221.8768	UUCUUG	159.76662	UAGGGA	126.49652
GAGGGG	221.2297	AUGGGG	157.96138		
GUGUAU	220.0392	UUUUAG	156.2937		
UUUUGG	216.7935	UUCUUA	155.85392		

References

Auweter, S. D., Fasan, R., Reymond, L., Underwood, J. G., Black, D. L., Pitsch, S., & Allain, F. H. (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *The EMBO Journal*, *25*(1), 163–173. https://doi.org/10.1038/sj.emboj.7600918

Begg, B. E., Jens, M., Wang, P. Y., Minor, C. M., & Burge, C. B. (2020). Concentration-dependent splicing is enabled by Rbfox motifs of intermediate affinity. *Nature Structural & Molecular Biology*, *27*(10), 901–912. https://doi.org/10.1038/s41594-020-0475-8

Black, D. L. (2003). Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry*, *72*(1), 291–336. https://doi.org/10.1146/annurev.biochem.72.121801.161720

Cao, W., Sohail, M., Liu, G., Koumbadinga, G. A., Lobo, V. G., & Xie, J. (2011). Differential effects of PKA-controlled CaMKK2 variants on neuronal differentiation. *RNA Biology*, *8*(6), 1061–1072. https://doi.org/10.4161/rna.8.6.16691

Conboy, J. G. (2017). Developmental regulation of RNA processing by Rbfox proteins. *WIREs RNA*, *8*(2), e1398. https://doi.org/10.1002/wrna.1398

Damianov, A., & Black, D. L. (2010). Autoregulation of Fox protein expression to produce dominant negative splicing factors. RNA, 16(2), 405–416. https://doi.org/10.1261/rna.1838210

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., & Black, D. L. (2016). Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell*, *165*(3), 606–619. https://doi.org/10.1016/j.cell.2016.03.040

Damianov, A., Lin, C.-H., Huang, J., Zhou, L., Jami-Alahmadi, Y., Peyda, P., Wohlschlegel, J., & Black, D. L. (2024). The splicing regulators RBM5 and RBM10 are subunits of the U2 snRNP engaged with intron branch sites on chromatin. Molecular Cell, 84(8), 1496-1511.e7. https://doi.org/10.1016/j.molcel.2024.02.039

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Feng, H., Bao, S., Rahman, M. A., Weyn-Vanhentenryck, S. M., Khan, A., Wong, J., Shah, A., Flynn, E. D., Krainer, A. R., & Zhang, C. (2019). Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein-RNA Crosslink Sites. *Molecular Cell*, *74*(6), 1189-1204.e6. https://doi.org/10.1016/j.molcel.2019.02.002

Gehman, L. T., Meera, P., Stoilov, P., Shiue, L., O'Brien, J. E., Meisler, M. H., Ares, M., Otis, T. S., & Black, D. L. (2012). The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. *Genes & Development*, *26*(5), 445–460. https://doi.org/10.1101/gad.182477.111

Gehman, L. T., Stoilov, P., Maguire, J., Damianov, A., Lin, C.-H., Shiue, L., Ares, M., Mody, I., & Black, D. L. (2011). The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nature Genetics*, *43*(7), Article 7. https://doi.org/10.1038/ng.841

Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., & Zavolan, M. (2021). CLIP and complementary methods. *Nature Reviews Methods Primers*, *1*(1), Article 1. https://doi.org/10.1038/s43586-021-00018-1

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Hennig, J., Militti, C., Popowicz, G. M., Wang, I., Sonntag, M., Geerlof, A., Gabel, F., Gebauer, F., & Sattler, M. (2014). Structural basis for the assembly of the Sxl-Unr translation regulatory complex. *Nature*, *515*(7526), 287–290. https://doi.org/10.1038/nature13693

Ho, J. S., Di Tullio, F., Schwarz, M., Low, D., Incarnato, D., Gay, F., Tabaglio, T., Zhang, J., Wollmann, H., Chen, L., An, O., Chan, T. H. M., Hall Hickman, A., Zheng, S., Roudko, V., Chen, S., Karz, A., Ahmed, M., He, H. H., … Guccione, E. (2021). HNRNPM controls circRNA biogenesis and splicing fidelity to sustain cancer cell fitness. *eLife*, *10*, e59654. https://doi.org/10.7554/eLife.59654

Homsy, J., Zaidi, S., Shen, Y., Ware, J. S., Samocha, K. E., Karczewski, K. J., DePalma, S. R., McKean, D., Wakimoto, H., Gorham, J., Jin, S. C., Deanfield, J., Giardini, A., Porter, G. A., Kim, R., Bilguvar, K., López-Giráldez, F., Tikhonova, I., Mane, S., … Chung, W. K. (2015). De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science*, *350*(6265), 1262–1266. https://doi.org/10.1126/science.aac9396

Huang, J., and Manning, B.D. (2008). The TSC1–TSC2 complex: a molecular switchboard controlling cell growth. Biochemical Journal 412, 179–190. 10.1042/BJ20080281.

Huang, M., Akerberg, A. A., Zhang, X., Yoon, H., Joshi, S., Hallinan, C., Nguyen, C., Pu, W. T., Haigis, M. C., Burns, C. G., & Burns, C. E. (2022). Intrinsic myocardial defects underlie an Rbfox-deficient zebrafish model of hypoplastic left heart syndrome. *Nature Communications*, *13*, 5877. https://doi.org/10.1038/s41467-022-32982-x

Jacko, M., Weyn-Vanhentenryck, S. M., Smerdon, J. W., Yan, R., Feng, H., Williams, D. J., Pai, J., Xu, K., Wichterle, H., & Zhang, C. (2018). Rbfox Splicing Factors Promote Neuronal Maturation and Axon Initial Segment Assembly. *Neuron*, *97*(4), 853-868.e6. https://doi.org/10.1016/j.neuron.2018.01.020

Jangi, M., Boutz, P. L., Paul, P., & Sharp, P. A. (2014). Rbfox2 controls autoregulation in RNA-binding protein networks. *Genes & Development*, *28*(6), 637–651. https://doi.org/10.1101/gad.235770.113
Jbara, A., Lin, K.-T., Stossel, C., Siegfried, Z., Shqerat, H., Amar-Schwartz, A., Elyada, E., Mogilevsky, M., Raitses-Gurevich, M., Johnson, J. L., Yaron, T. M., Ovadia, O., Jang, G. H., Danan-Gotthold, M., Cantley, L. C., Levanon, E. Y., Gallinger, S., Krainer, A. R., Golan, T., & Karni, R. (2023). RBFOX2 modulates a metastatic signature of alternative splicing in pancreatic cancer. *Nature*, *617*(7959), 147–153. https://doi.org/10.1038/s41586-023-05820-3

Jiang, Y., Fu, X., Zhang, Y., Wang, S.-F., Zhu, H., Wang, W.-K., Zhang, L., Wu, P., Wong, C. C. L., Li, J., Ma, J., Guan, J.-S., Huang, Y., & Hui, J. (2021). Rett syndrome linked to defects in forming the MeCP2/Rbfox/LASR complex in mouse models. *Nature Communications*, *12*(1), 5767. https://doi.org/10.1038/s41467-021-26084-3

Jin, Y., Suzuki, H., Maegawa, S., Endo, H., Sugano, S., Hashimoto, K., Yasuda, K., & Inoue, K. (2003). A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *The EMBO Journal*, *22*(4), 905–912. https://doi.org/10.1093/emboj/cdg089

Kuwasako, K., Takahashi, M., Unzai, S., Tsuda, K., Yoshikawa, S., He, F., Kobayashi, N., Güntert, P., Shirouzu, M., Ito, T., Tanaka, A., Yokoyama, S., Hagiwara, M., Kuroyanagi, H., & Muto, Y. (2014). RBFOX and SUP-12 sandwich a G base to cooperatively regulate tissue-specific splicing. *Nature Structural & Molecular Biology*, *21*(9), 778–786. https://doi.org/10.1038/nsmb.2870

Lal, D., Reinthaler, E. M., Altmüller, J., Toliat, M. R., Thiele, H., Nürnberg, P., Lerche, H., Hahn, A., Møller, R. S., Muhle, H., Sander, T., Zimprich, F., & Neubauer, B. A. (2013). RBFOX1 and RBFOX3 Mutations in Rolandic Epilepsy. *PLoS ONE*, *8*(9), e73323. https://doi.org/10.1371/journal.pone.0073323

Lal, D., Trucks, H., Møller, R. S., Hjalgrim, H., Koeleman, B. P. C., de Kovel, C. G. F., Visscher, F., Weber, Y. G., Lerche, H., Becker, F., Schankin, C. J., Neubauer, B. A., Surges, R., Kunz, W. S., Zimprich, F., Franke, A., Illig, T., Ried, J. S., Leu, C., … EPICURE Consortium. (2013). Rare exonic deletions of the RBFOX1 gene increase risk

of idiopathic generalized epilepsy. *Epilepsia*, *54*(2), 265–271. https://doi.org/10.1111/epi.12084

Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P. A., & Burge, C. B. (2014). RNA Bind-n-Seq: Quantitative Assessment of the Sequence and Structural Binding Specificity of RNA Binding Proteins. *Molecular Cell*, *54*(5), 887–900. https://doi.org/10.1016/j.molcel.2014.04.016

Li, J., Choi, P. S., Chaffer, C. L., Labella, K., Hwang, J. H., Giacomelli, A. O., Kim, J. W., Ilic, N., Doench, J. G., Ly, S. H., Dai, C., Hagel, K., Hong, A. L., Gjoerup, O., Goel, S., Ge, J. Y., Root, D. E., Zhao, J. J., Brooks, A. N., … Hahn, W. C. (2018). An alternative splicing switch in FLNB promotes the mesenchymal cell state in human breast cancer. *eLife*, *7*, e37184. https://doi.org/10.7554/eLife.37184

Lovci, M. T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T. Y., Stark, T. J., Gehman, L. T., Hoon, S., Massirer, K. B., Pratt, G. A., Black, D. L., Gray, J. W., Conboy, J. G., & Yeo, G. W. (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nature Structural & Molecular Biology*, *20*(12), 1434–1442. https://doi.org/10.1038/nsmb.2699

Matunis, M. J., Xing, J., & Dreyfuss, G. (1994). The hnRNP F protein: Unique primary structure, nucleic acid-binding properties, and subcellular localization. *Nucleic Acids Research*, *22*(6), 1059–1067. https://doi.org/10.1093/nar/22.6.1059

Maurin, M., Ranjouri, M., Megino-Luque, C., Newberg, J. Y., Du, D., Martin, K., Miner, R. E., Prater, M. S., Wee, D. K. B., Centeno, B., Pruett-Miller, S. M., Stewart, P., Fleming, J. B., Yu, X., Bravo-Cordero, J. J., Guccione, E., Black, M. A., & Mann, K. M. (2023). RBFOX2 deregulation promotes pancreatic cancer progression and metastasis through alternative splicing. *Nature Communications*, *14*(1), 8444. https://doi.org/10.1038/s41467-023-44126-w

Misra, C., Bangru, S., Lin, F., Lam, K., Koenig, S. N., Lubbers, E. R., Hedhli, J., Murphy, N. P., Parker, D. J., Dobrucki, L. W., Cooper, T. A., Tajkhorshid, E., Mohler, P. J., & Kalsotra, A. (2020). Aberrant Expression of a Non-muscle RBFOX2 Isoform Triggers Cardiac Conduction Defects in Myotonic Dystrophy. *Developmental Cell*, *52*(6), 748-763.e6. https://doi.org/10.1016/j.devcel.2020.01.037

Moss, N. D., Wells, K. L., Theis, A., Kim, Y.-K., Spigelman, A. F., Liu, X., MacDonald, P. E., & Sussel, L. (2023). Modulation of insulin secretion by RBFOX2-mediated alternative splicing. *Nature Communications*, *14*(1), 7732. https://doi.org/10.1038/s41467-023-43605-4

Palmer, L. E., Weiss, M. J., & Paralkar, V. R. (2017). *YODEL: Peak calling software for HITS-CLIP data* (6:1138). F1000Research. https://doi.org/10.12688/f1000research.11861.1

Paterson, H. A. B., Yu, S., Artigas, N., Prado, M. A., Haberman, N., Wang, Y.-F., Jobbins, A. M., Pahita, E., Mokochinski, J., Hall, Z., Guerin, M., Paulo, J. A., Ng, S. S., Villarroya, F., Rashid, S. T., Le Goff, W., Lenhard, B., Cebola, I., Finley, D., … Vernia, S. (2022). Liver RBFOX2 regulates cholesterol homeostasis via Scarb1 alternative splicing in mice. *Nature Metabolism*, *4*(12), 1812–1829. https://doi.org/10.1038/s42255-022-00681-y

Ramesh, N., Kour, S., Anderson, E. N., Rajasundaram, D., & Pandey, U. B. (2020). RNA-recognition motif in Matrin-3 mediates neurodegeneration through interaction with hnRNPM. *Acta Neuropathologica Communications*, *8*(1), 138. https://doi.org/10.1186/s40478-020-01021-5

Shapiro, I. M., Cheng, A. W., Flytzanis, N. C., Balsamo, M., Condeelis, J. S., Oktay, M. H., Burge, C. B., & Gertler, F. B. (2011). An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genetics*, *7*(8), e1002218. https://doi.org/10.1371/journal.pgen.1002218

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLOS ONE, 11(10), e0163962. https://doi.org/10.1371/journal.pone.0163962

Smith, C. W. J., & Valcárcel, J. (2000). Alternative pre-mRNA splicing: The logic of combinatorial control. *Trends in Biochemical Sciences*, *25*(8), 381–388. https://doi.org/10.1016/S0968-0004(00)01604-2

Ule, J., & Blencowe, B. J. (2019). Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Molecular Cell*, *76*(2), 329–345. https://doi.org/10.1016/j.molcel.2019.09.017

Uren, P. J., Bahrami-Samani, E., de Araujo, P. R., Vogel, C., Qiao, M., Burns, S. C., Smith, A. D., & Penalva, L. O. F. (2016). High-throughput analyses of hnRNP H1 dissects its multi-functional aspect. *RNA Biology*, *13*(4), 400–411. https://doi.org/10.1080/15476286.2015.1138030

Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J.-Y., Cody, N. A. L., Dominguez, D., Olson, S., Sundararaman, B., Zhan, L., Bazile, C., Bouvrette, L. P. B., Bergalet, J., Duff, M. O., Garcia, K. E., Gelboin-Burkhart, C., … Yeo, G. W. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature*, *583*(7818), 711–719. https://doi.org/10.1038/s41586-020-2077-3

Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., Blue, S. M., Nguyen, T. B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., & Yeo, G. W. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature Methods*, *13*(6), 508–514. https://doi.org/10.1038/nmeth.3810

Venables, J. P., Brosseau, J.-P., Gadea, G., Klinck, R., Prinos, P., Beaulieu, J.-F., Lapointe, E., Durand, M., Thibault, P., Tremblay, K., Rousset, F., Tazi, J., Abou Elela, S., & Chabot, B. (2013). RBFOX2 Is an Important Regulator of Mesenchymal Tissue-Specific Splicing in both Normal and Cancer Tissues. *Molecular and Cellular Biology*, *33*(2), 396–405. https://doi.org/10.1128/MCB.01174-12

Verma, S. K., Deshmukh, V., Nutter, C. A., Jaworski, E., Jin, W., Wadhwa, L., Abata, J., Ricci, M., Lincoln, J., Martin, J. F., Yeo, G. W., & Kuyumcu-Martinez, M. N. (2016). Rbfox2 function in RNA metabolism is impaired in hypoplastic left heart syndrome patient hearts. *Scientific Reports*, *6*(1), 30896. https://doi.org/10.1038/srep30896

Vuong, C. K., Wei, W., Lee, J.-A., Lin, C.-H., Damianov, A., De La Torre-Ubieta, L., Halabi, R., Otis, K. O., Martin, K. C., O'Dell, T. J., & Black, D. L. (2018). Rbfox1 Regulates Synaptic Transmission through the Inhibitory Neuron-Specific vSNARE Vamp1. *Neuron*, *98*(1), 127-141.e7. https://doi.org/10.1016/j.neuron.2018.03.008

Weyn-Vanhentenryck, S. M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P. A., Zhang, M. Q., Krainer, A. R., Darnell, R. B., & Zhang, C. (2014). HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism. *Cell Reports*, *6*(6), 1139–1152. https://doi.org/10.1016/j.celrep.2014.02.005

Whitington, T., Frith, M. C., Johnson, J., & Bailey, T. L. (2011). Inferring transcription factor complexes from ChIP-seq data. Nucleic Acids Research, 39(15), e98. https://doi.org/10.1093/nar/gkr341

Wysoczański, P., Schneider, C., Xiang, S., Munari, F., Trowitzsch, S., Wahl, M. C., Lührmann, R., Becker, S., & Zweckstetter, M. (2014). Cooperative structure of the heterotrimeric pre-mRNA retention and splicing complex. *Nature Structural & Molecular Biology*, *21*(10), 911–918. https://doi.org/10.1038/nsmb.2889

Ye, X., Yang, W., Yi, S., Zhao, Y., Varani, G., Jankowsky, E., & Yang, F. (2023). Two distinct binding modes provide the RNA-binding protein RbFox with extraordinary sequence specificity. *Nature Communications*, *14*(1), 701. https://doi.org/10.1038/s41467-023-36394-3

Yeo, G. W., Coufal, N. G., Liang, T. Y., Peng, G. E., Fu, X.-D., & 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*(2), Article 2. https://doi.org/10.1038/nsmb.1545

Ying, Y., Wang, X.-J., Vuong, C. K., Lin, C.-H., Damianov, A., & Black, D. L. (2017). Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell*, *170*(2), 312-323.e10. https://doi.org/10.1016/j.cell.2017.06.022

Zarnack, K., König, J., Tajnik, M., Martincorena, I., Eustermann, S., Stévant, I., Reyes, A., Anders, S., Luscombe, N. M., & Ule, J. (2013). Direct Competition between hnRNP C and U2AF65 Protects the Transcriptome from the Exonization of Alu Elements. *Cell*, *152*(3), 453–466. https://doi.org/10.1016/j.cell.2012.12.023

Zhou, D., Couture, S., Scott, M. S., & Abou Elela, S. (2021). RBFOX2 alters splicing outcome in distinct binding modes with multiple protein partners. *Nucleic Acids Research*, *49*(14), 8370–8383. https://doi.org/10.1093/nar/gkab595

Zhu, G.-Q., Wang, Y., Wang, B., Liu, W.-R., Dong, S.-S., Chen, E.-B., Cai, J.-L., Wan, J.-L., Du, J.-X., Song, L.-N., Chen, S.-P., Yu, L., Zhou, Z.-J., Wang, Z., Zhou, J., Shi, Y.-H., Fan, J., & Dai, Z. (2022). Targeting HNRNPM Inhibits Cancer Stemness and Enhances Antitumor Immunity in Wnt-activated Hepatocellular Carcinoma. *Cellular and Molecular Gastroenterology and Hepatology*, *13*(5), 1413–1447. https://doi.org/10.1016/j.jcmgh.2022.02.006

Chapter 3: Sequence determinants of Rbfox oligomerization

Abstract

Low complexity domains in RNA-binding proteins can engage in homotypic and heterotypic interactions that drive their phase separation into higher-order structures. However, the molecular determinants of how these associations occur and their physiological roles remain unknown for most of these proteins. The Rbfox family of RNA-binding proteins self-assemble into higher-orders structures through a tyrosine-rich, low complexity C2 region within their C-terminal domain. This property is critical for Rbfox's ability to activate splicing of a subset of its targets. However, the sequence determinants of Rbfox's self-assembly have not been fully investigated. Here we examine what sequences and parts of the Rbfox2 C2 region contribute to its oligomerization via size-exclusion chromatography and co-immunoprecipitation assays. We find that the first half of C2, C2(1-29), which consists of the first twenty-nine amino acids, can oligomerize even more than the full length C2 region. In contrast, the last half of C2, C2(30-59), which consists of the last thirty amino acids, loses the ability to oligomerize. Furthermore, serine mutations at tyrosine residues clustered in different parts of C2 diminish homo-oligomerization but retain binding to a wildtype copy. The findings and approaches we present inform future investigations of how Rbfox self-assembles into higher-order structures.

Introduction

Proteins usually adopt a stable three-dimensional structure that is tightly linked to their function. However, many proteins also contain intrinsically disordered regions (IDRs) that are more flexible and fluctuate rapidly across a series of conformations (Oldfield & Dunker, 2014). These regions often contain amino acid sequences of low complexity that can be biased in their composition (Kato et al., 2022). IDRs are involved in a wide variety of cellular processes such as transcription, cellular signaling, and stress responses (Wright & Dyson, 2015). Many of these regions engage in homotypic or heterotypic interactions to phase separate into membranelles organelles such as the nucleolus, Cajal bodies, and nuclear speckles in the nucleus and stress granules, P-bodies, and germ granules in the cytoplasm (Shin & Brangwynne, 2017). However, the molecular determinants of how these structures are formed are not fully understood.

RNA-binding proteins (RBPs), in particular, are enriched in low complexity domains and mutations in these regions are often linked to neurodegenerative diseases (Zhao et al., 2021). For example, TDP-43 is an RBP implicated in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). This protein undergoes phase separation through forming labile cross-β interactions via a low complexity domain in its c-terminus (Zhou et al., 2022). Similar forms of cross-β associations are also important for homotypic interactions of the RNA binding proteins FUS, also involved in ALS and FTD, and hnRNP A2, implicated in multisystem proteinopathy (Lu et al., 2024; Murray et al., 2017). However, there is also evidence indicating that the interactions that drive the self-assembly of these proteins are more complex than cross-β interactions. An alternative model for these associations is a sticker-and-spacer model taken from

93

polymer theory. This models posits there are "sticky" amino acids that protrude out from a linear side-chain and form associations with other sticky amino acids on another side-chain leading to the phase separation of these proteins (Choi et al., 2020; Murthy & Fawzi, 2020). There are, perhaps, diverse strategies via which IDRs self-assemble and phase separate. Therefore, it is crucial to investigate the underlying mechanism of self-association in IDRs in the context of additional RBPs and understand the functional consequences of these interactions.

As discussed in chapters 1 and 2, the Rbfox proteins have a low complexity tyrosine-rich C-terminal domain that binds to the LASR proteins and also self-assembles into higher-order structures. Ten of these tyrosines are clustered within a fifty-nine amino acid region called C2. Mutation of the tyrosine residues in this region to serines diminishes Rbfox's self-assembly but still retains LASR binding, indicating that the interactions driving Rbfox's self-assembly are distinct from the forces driving its association with LASR (Ying et al., 2017). Furthermore, the tyrosine to serine mutant Rbfox loses ability to activate a subset of its regulatory cassette exons, providing a link between Rbfox's self-assembly and its splicing activity. The C2 region also contains two mutually exclusive exons: the M43 exon expressed only in muscle and the B40 exon expressed in other tissues (Conboy, 2017). Rbfox2's M43 exon lacks three of the tyrosines present in the B40 exon and the M43 isoform is less prone to form higher-order structures. Interestingly, in myotonic dystrophy type 1 the Rbfox2 B40 isoform is aberrantly expressed in the heart and leads to splicing changes of ion channels and conduction defects (Misra et al., 2020). These studies demonstrate the importance of the self-assembly of Rbfox via its C2 region and point towards a need to investigate

what specific regions, motifs, and amino acid residues within C2 contribute to its oligomerization. In this chapter, we used size-exclusion chromatography and co-immunoprecipitations assays to examine what parts and sequences within the C2 region contribute to its self-assembly.

Results

Rbfox2's C2 region contains parts that drive and inhibit its oligomerization

Rbfox2's 59 amino acid C2 region was fused with an N-terminal monomeric Cherry (mCherry) and a C-terminal TEV cleavage site followed by a 6xHistidine tag (mCherry-C2(1-59)-TEV-His) and expressed as a recombinant protein in *E. Coli* (Fig. 3.1A). mCherry was incorporated to increase the solubility of C2 and monitor its migration in the size-exclusion column; the histidine tag was used for purification via a nickel affinity column. SDS-PAGE and total protein staining of the purifications indicates high purify of the mCherry-C2(1-59)-TEV-His (Fig. S3.1A). Besides the major band corresponding to the full-length protein, there was also a minor smaller band, corresponding to a truncation product that was detected by an anti-Histag immunoblot but not by anti-mCherry (Fig S3.1B-C).

The oligomerization state of purified mCherry-C2(1-59)-TEV-His was then assessed via size-exclusion chromatography (SEC) using a Superdex 200 column. Migration through the column was monitored via the absorbance of mCherry at 587 nm. The chromatograph shows 3 major peaks (Fig. 3.1B). The fastest migrating peak at 40 mL, which corresponds to the column's void volume, contains 9% of the total absorbance and likely consists of large protein aggregates. The subsequent peak at 49 mL, migrating close to where a 670 kDa marker protein elutes, consists of 34% of the protein. The third peak contains 57% of the total absorbance and elutes at 85 mL, close to a 44 kD marker proteins. Since the molecular weight of mCherry-C2(1-59)-TEV-His is 35.9 kDa, the 85 mL peak likely corresponds to monomeric species. A 600-700 kD

assembly of this protein would, therefore, be approximately 18-20 copies of the monomer.

In addition, to this fusion protein containing the entire C2 region, we also made fusion proteins of truncation mutants: C2(1-29), which contained the first 29 amino acids of C2, C2(30-59) which contained the last 30 amino acids, and C2 (1-14, 44-59) which contained the first 14 amino acids and last 16 amino acids (Fig. 3.1A). These proteins were expressed and purified in the same manner as the full length C2 and had similar purities as assessed by total protein staining and immunoblotting for mCherry (Fig. S3.1A-B). SEC of the C2(1-29) shows two peaks: a major peak of 83% of the total absorbance migrating at 50 mL and a smaller peak of 17% of the absorbance migrating at 89 mL. In contrast to C2(1-59), C2(1-29) does not come out in the void, indicating this protein does not form large aggregates. However, more of this protein is present as the 600-700 kDa complex, indicating this protein has a higher propensity to oligomerize.

In contrast, C2(30-59) mostly exist in a monomeric state. The SEC profile of this protein has 3 peaks: 4% absorbance at 56 mL, 3% absorbance at 67 mL, and 93% absorbance at 86 mL, demonstrating that the sequence in this part of C2 by itself is insufficient to oligomerize into the 600-700 kDa complex. C2(1-14, 44-59) also has two peaks migrating at 52 mL with 40% absorbance and 88 mL with 60% absorbance. Thus, C2(1-14, 44-59) has a monomeric to oligomerized ratio similar to C2(1-59) but does not form large aggregates that elute in the void volume.

The oligomerization of C2 was also assessed via a co-immunoprecipitation assay. We reasoned since C2 oligomerizes, an untagged version of C2 should bind to and co-immunoprecipitate with an epitope tagged version (Fig. 3.1C). Therefore, we put

97

a FLAG-tag on the N-terminus of the C2 protein. This fusion protein also contained a monomeric enhanced GFP (mEGFP) in its N-terminus and TEV and histidine tag in its C-terminus as before yielding FLAG-mEGFP-C2-TEV-His. The mEGFP was put in place of the mCherry to allow probing for the epitope tagged version of the protein separately from the untagged version (if both the epitope and untagged version of the proteins had mCherry we would not be able to specifically detect the untagged version of the protein. In addition, this tag is needed to enhance the solubility of the protein). For brevity, we will omit the names of the tags at the C-terminus of these proteins and will refer to them as FLAG-GFP-C2 and mCherry-C2.

FLAG-GFP-C2 was pre-bound to beads containing an anti-FLAG antibody. mCherry-C2 was then added to these beads at increasing concentrations of 0.5, 1, 5 and 10 uM. The concentration of FLAG-mEGFP-C2 was fixed at 1 uM. In the eluates, mCherry was detected when it was added at 0.5 uM and increased in intensity for the co-IPs with the 5 and 10 uM concentrations, indicating that the mCherry-C2 protein binds to and co-IPs with the FLAG-mEGFP-C2 protein. We next assessed the co-immunoprecipitation of the truncation mutants. C2(1-29) was detected in elutions starting at 0.5 uM and this signal increased in a concentration dependent manner. Furthermore, the amount of co-immunoprecipitated C2(1-29) was higher compared to C2(1-59), indicating that this protein has a higher affinity to bind to C2(1-59). In contrast. C2(30-59) was not detected in the eluates at any concentration. Co-immunoprecipitated C2(1-14,44-59) was detected starting at 5 uM. Together, these co-IP results show that a copy of C2(1-59) can bind to another copy of C2(1-59), in agreement with the ability of this protein to oligomerize. C2(1-29) can also bind to C2(1-59) at a seemingly higher

affinity. In contrast, C2(30-59) lacks the ability to bind to C2(1-59). Lastly, C2(1-14,44-59) can also bind to C2(1-59).
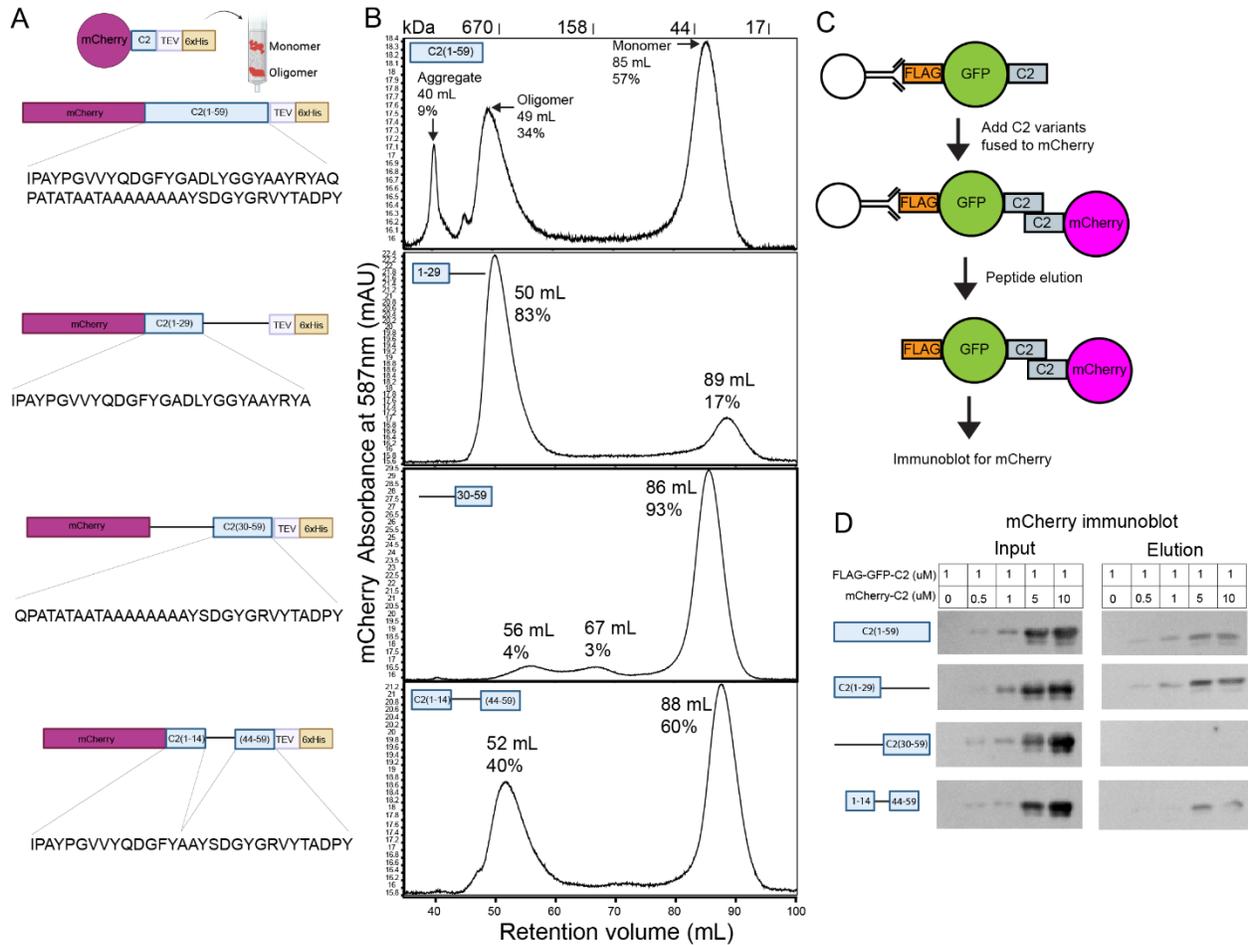
Figure 3.1. Rbfox2's C2 domain contains parts that drive and inhibit its oligomerization.

A) Schematic for the size-exclusion chromatography assay. Diagrams of C2 fusion proteins are also displayed. The C2 domain of Rbfox2 was fused with an n-terminal mCherry tag and a c-terminal TEV and 6xHis tags. The sequence of this domain is shown below the C2(1-59) fusion protein. Truncation mutations and their respective sequences are also shown. B) Size-exclusion chromatographs of the C2 fusion proteins

run on a Superdex 200 column with monitoring of absorbance of mCherry at 587 nm. The elution volumes of marker proteins are displayed above with their respective molecular weights. The volume at the center of each peak is also displayed for the chromatograph. The % refers to the area under the curve for a given peak divided by the total areas for all the peaks within a chromatograph. C) Schematic for a co-binding assay of C2. FLAG-mEGFP-C2-TEV-6xHis was immobilized to beads that have a FLAG antibody. Then mCherry-C2 protein at different concentrations was added while the concentration of the FLAG-mEGFP-C2 was fixed at 1 uM. Beads were washed, proteins were eluted via peptide elution, and the elution were examined via immunoblotting. D) Immunoblot against mCherry for the input and the elutions of the co-immunoprecipitations. The concentrations of the FLAG-GFP-C2 and mCherry-C2 proteins are displayed above each lane.

Differently spaced tyrosine residues in Rbfox2's C2 domain have distinct effects on its self-assembly

Previously, it was reported that mutating ten tyrosines to serines in C2 abrogated Rbfox's self-assembly (Ying et al., 2017). However, the contribution of individual tyrosines to this process remain unclear. Some tyrosines are clustered together with distinct spacing patterns: tyrosines 2-4 are spaced apart by four amino acids, tyrosines 5-7 are spaced apart by less than three amino acids, and tyrosines 8-10 have a spacing of three amino acids. Therefore, we investigated how C2 self-assembles when all tyrosines or clusters of differentially spaced tyrosines are mutated (Figure 3.2A).

SEC of a recombinant C2 fusion protein with all nine of these tyrosine residues mutated to serines, C2(2-10YS), shows this protein migrates as a single monomeric species at 82 mL (Figure 3.2A-B). A C2(2-4YS) protein primarily migrated as a monomer with 93% at 84 mL. Additionally, 3% of this protein migrated with the void volume at 40 mL, 1% eluted around 45 mL, and 3% eluted as a broad peak centered 52 mL. SEC of C2(5-7YS) shows 4% of this protein eluting at 40 mL retention in the void volume, 1% at 45 mL, 2% at 54 mL, and 93% at 84 mL. Both the C2(2-4YS) and the C2(5-7YS), therefore, formed large aggregates in addition to smaller oligomeric complexes. A C2(8-10YS) mutant migrated exclusively as a monomer at 84 mL, giving a similar chromatographic profile as the C2(2-10YS).

The binding of the tyrosine mutants to the wildtype C2 was assessed by the co-immunoprecipitation assay (Figure 3.1C). The C2(2-10YS) protein did not co-

101

immunoprecipitate with the C2(1-59) protein at any of the concentrations tested. In

contrast, all three of the C2(2-4YS), C2(5-7YS), and C2(8-10YS) were detected when

they were added at both 5 and 10 uM. The C2(8-10YS) also shows a faint band present

at 0.5 and 1 uM. Overall, the co-binding of the three proteins is not substantially

different from one another and the differences can be due to variations in the

immunoblot procedure.
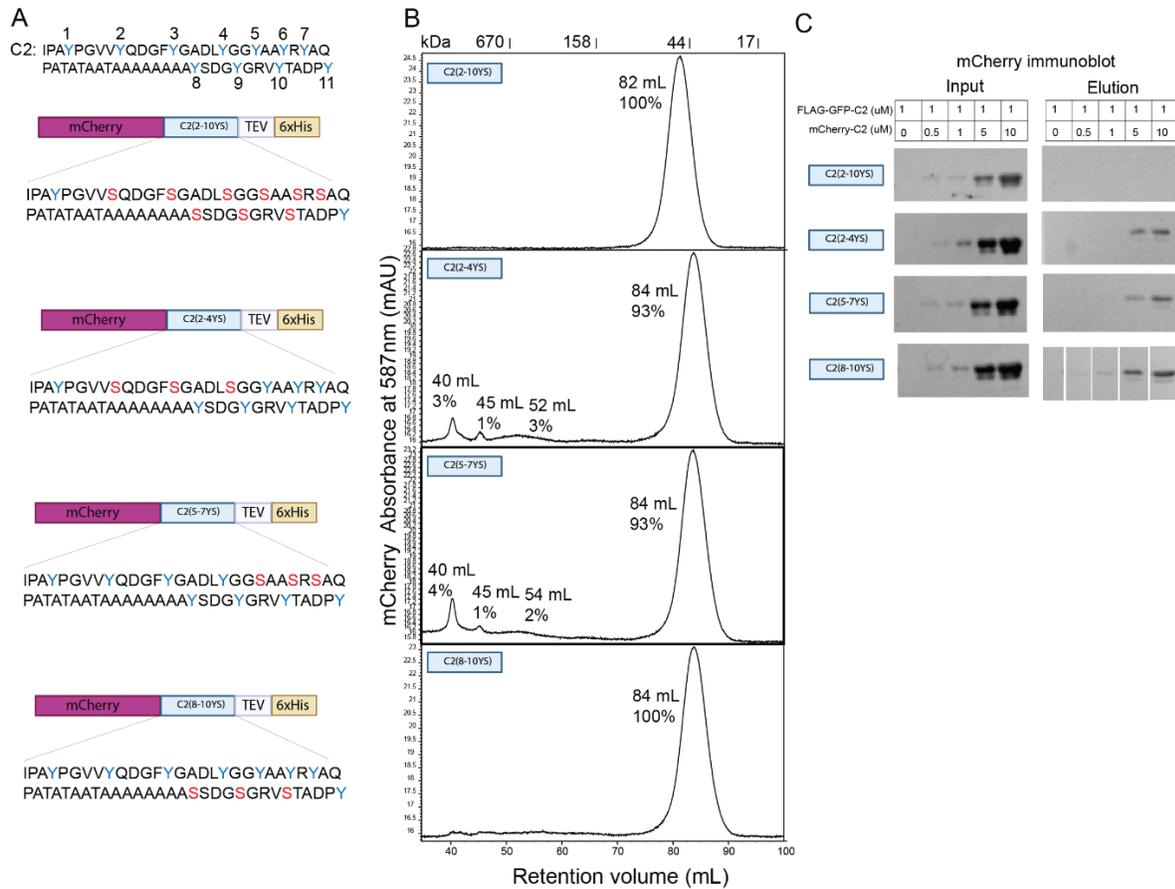
Figure 3.2. Differently spaced tyrosine residues in Rbfox2's C2 domain have specific effects on its self-assembly.

A) Diagrams of fusion recombinant proteins containing different tyrosine to serine mutations in C2. The sequence of the wildtype C2 is displayed at the top and all the tyrosines in the sequence are displayed in blue. The serine mutations are shown in red in the sequences below. B) Size-exclusion chromatographs of the C2 proteins run on the superdex 200 column with absorbance of mCherry monitered at 587 nm. The retention volumes and percentages are displayed as described in the legend for figure 3-1D. C) mCherry immunoblots for co-immunoprecipation of the C2 tyrosine mutants

103

with the FLAG-GFP-C2 protein. Concentrations of the proteins are indicated above

each lane.

Discussion

Sequence determinants of C2 oligomerization

Homotypic interactions mediated by low complexity domains of RNA binding proteins contribute to their physiological and pathological roles. Previously, it was shown that Rbfox's C-terminal domain contains 10 tyrosines within its C2 region that drive its self-assembly and are important for its ability to stimulate splicing. Here, we developed two tandem approaches to examine the sequence determinants of Rbfox's self-assembly. Using size-exclusion chromatography and co-immunoprecipitation studies we tested the oligomerization of different truncation mutants of the C2 region. We found that the first half of C2, C2(1-29), oligomerizes even more than full length C2. In contrast, the last half of the C2, C2(30-59), loses its ability to oligomerize. A recombinant protein containing the first fourteen amino acids and the last sixteen amino acids, C2(1-14, 44-59), was also able to oligomerize similar to the full-length protein but not as well as C2(1-29).

Six tyrosines are in C2(1-29) compared to five tyrosines in C2(30-59). Given the large differences between the oligomerization states of the two proteins, it is unlikely that the number of tyrosines by itself is sufficient to explain their different properties. Therefore, there might be other amino acid residues or structural features within C2(1-29) that drive its self-assembly. The B40 exon which was previously shown to be important for Rbfox2's self-assembly is located within C2(1-29). When this exon is swapped with M43, Rbfox's ability is greatly diminished even though there are only a difference of three tyrosines between B40 and M43 (Ying et al., 2017). Therefore, there might exist motifs or structural features within the B40 that are essential for

oligomerization of Rbfox. SEC of a recombinant protein with just the B40 exon, in the same mCherry-TEV-His backbone as the other C2 proteins forms, indicates this protein can oligomerize (data not shown). Examining oligomerization of B40 truncation mutants might yield a minimal motif that drives self-assembly.

An intriguing feature of C2(1-29) is its ability to oligomerize more than the full length C2(1-59). Right after where the C2(1-29) sequence ends there is a proline followed by a repeating sequence of alanines. This rigid proline perhaps separates features that drive self-assembly from those that act antagonistically. It would be interesting to assess the oligomerization of recombinant proteins containing C2(1-29) fused to a proline and repeating alanines.

The tyrosines in C2 are clearly important for its self-assembly. Mutation of nine of these tyrosines to serines yields an exclusively monomeric protein. However, the C2(2-4YS) and C2(5-7YS) proteins were still able to homo-oligomerize, although to a much smaller extent. These proteins also co-immunoprecipitated with an epitope tagged wildtype C2 fusion protein. Interestingly, the C2(8-10YS) protein also loses the ability to oligomerize, yet co-immunoprecipitates with wildtype C2. Homo-oligomerization of C2 into a stable 600-700 kDa complex might therefore require different interactions from those that lead to co-immunoprecipitation.

Structural studies should also be considered in future studies of C2 self-assembly. SEC of full length C2 shows large aggregates, a 600-700 kDa oligomer, and monomeric species. These states are likely in equilibrium, with the large aggregates very heterogenous, complicating potential structural studies. In contrast, C2(1-29) exists as either a 600-700 kDa oligomer or a monomer. Since the peak of 600-700 kDa

oligomer is tightly defined in the chromatograph and not broadly distributed, it might be sufficiently homogenous for structural studies. Examining C2(1-29)with electron microscopy would be a reasonable starting point to assess the feasibility of this approach. An alternative approach to gaining structural information would be to test more truncation mutants until a smaller region of 6-10 amino acids is defined. Regions of this size that form homotypic interactions can be amenable to X-ray crystallography as previously applied to FUS, hnRNP A1, and nup98 (Hughes et al., 2018).

Co-immunoprecipitation assay of C2

Most studies utilize droplet formation assay to examine what features of a protein cause it to phase separate. However, the Rbfox proteins by themselves do not seem to form liquid droplets using these assays and instead form aggregates that appear to be more solid; this property makes the Rbfox proteins hard to study via droplet assays. We reasoned that if a protein self-assembles through homotypic interactions an untagged copy of a protein should co-immunoprecipitate with an epitope tagged copy. Therefore, we developed a co-immunoprecipitation assay to test binding of C2 mutants to wildtype C2. This assay shows clear differences of binding between the deletion mutants that also agree with the findings from size-exclusion chromatography. However, at this point this assay is qualitative and can be refined to give out quantitative measurements of the binding affinities. On approach would be to utilize a binding assay used to determine the affinity of GST-2/3/A, which is part of the tail of Acanthamoeba myosin-I, for the amoeba Arp2/3 complex (Pollard, 2010). With this approach, the mCherry-C2 concentration would be fixed and then different concentrations of the FLAG-GFP-C2 bound to beads would be added to this solution starting from a concentration suspected to be below the

K_d and increasing the concentration until all the mCherry-C2 gets bound to beads. This approach would allow for construction of a binding curve and determination of an equilibrium constant.

Droplet formation assays are also limited to testing one mutation at a time. However, the co-immunoprecipitation assay we have developed has the potential to be multiplexed to allow testing of hundreds to thousands of variants. There have been recent developments of peptide barcodes that allow one to tag different proteins with unique identifiers and test their binding to a ligand simultaneously (Egloff et al., 2019). The co-immunoprecipitation assay we have described here is amenable to this approach. High throughput DNA synthesis can be used to synthesize sequences of C2 variants with distinct barcodes. These proteins can then be expressed in a pool and purified. The binding of this pool of barcoded proteins to C2 can then be assessed via co-immunoprecipitation and analysis with mass spectrometry. However, the co-immunoprecipitation assay would need to be quantitative for this approach to be plausible. As aforementioned, development of a quantitative co-immunoprecipitation assay would be very useful.

Methods

Table 3.1. Antibodies

mCherry	Thermo Fisher Scientific	MA5-15257
GFP	Abcam	ab290
Histag	Thermo Fisher Scientific	MA1-21315

Plasmid construction

The mEGFP in the pET28a-mEGFP-TEV-6xHis plasmid that was previously constructed

was replaced with an mCherry via restriction cloning to yield pET28a-mCherry-TEV-

6xHis (Ying et al., 2017). C2 region of Rbfox2 was PCR amplified from a pET28a-

Rbfox2-TE with primers containing EcoRI and XhoI cut sites. This amplicon was then

digested and ligated into the pET28a-mCherry-TEV-6xHis backbone using EcoRI and

XhoI to yield pET28a-mCherry-C2-TEV-6xHis. The C2 deletion and tyrosine mutants

were derived from this parental vector via site-directed mutagenesis PCR using primers

shown in table 1. The C2 fragment was also cloned into the pET28a-mEGFP-TEV-6xHis

plasmid via restriction cloning as described above to yield pET28a-mEGFP-C2-TEV-

6xHis.

Recombinant protein expression and purification

Proteins expression was induced in BL21(DE3) *E. Coli* with 0.5 mM IPTG at 18C for 16

hours. Bacterial pellets were pelleted via centrifugation, resuspended in binding buffer

(50 mM Tris-HCl pH 7.5, 500 mM NaCl, 25 mM imidazole), and lysed via sonication.

This lysate was cleared via centrifugation and the supernatant was subsequently

applied to a HisTrap HP 5ml column (Cytiva Life Sciences) equilibrated with the binding

buffer via the AKTA system (Cytiva Life Sciences). The columns were washed with 5

column volumes of the binding buffer and then a linear gradient of the elution buffer (50

mM Tris-HCl pH 7.5, 500 mM NaCl, 500 mM imidazole) was applied to the column to elute the proteins. Elution of proteins was monitored via absorbance of mCherry at 587 nm and the fractions that contained solutions that absorbed in this range were pooled together. These pooled fractions were concentrated via Amicon Ultra centrifugal filters, (Millipore), buffer exchanged with storage buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl), flash frozen with liquid nitrogen, and stored at -80°C. Concentration of proteins was determined via absorbance at 280nm.

Size exclusion chromatography

Superdex HiLoad 16/600 200 prep grade column (Cytiva) was equilibrated with storage buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl). A solution containing 250 ug of each protein of interest in storage buffer was prepared and centrifuged at 20k rcf for 10 mins. The supernatant was applied to the column and the absorbance at 587 nm was monitored to determine the elution fraction of proteins. The marker proteins consisted of a standard containing thyroglobulin, bovine γ-globulin, chicken ovalbumin, equine myoglobin, and vit B12, MW 1,350–670,000 (Bio-rad Gel Filtration Standard #1511901).

Co-immunoprecipitation assays

FLAG-GFP-C2 was mixed with anti-FLAG magnetic agarose beads and rotated for one hour at room temperature. The beads were subsequently washed three times with binding buffer and distributed to yield 6.25 ul of beads containing 1 uM FLAG-GFP-C2 per immunoprecipitation. Different concentrations of mCherry-C2 were added to the beads for each condition and this mixture was rotated for one hour at room temperature. Post-immunoprecipitation washes were performed three times with the binding buffer. Elution buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1.5 mg/mL 3x FLAG peptide) was

then added to the beads and this solution was mixed for a total of fifteen minutes at

1000 rpm in intervals of fifteen seconds on and forty-five seconds off. The supernatant

was collected as the eluate.

Supplementary data
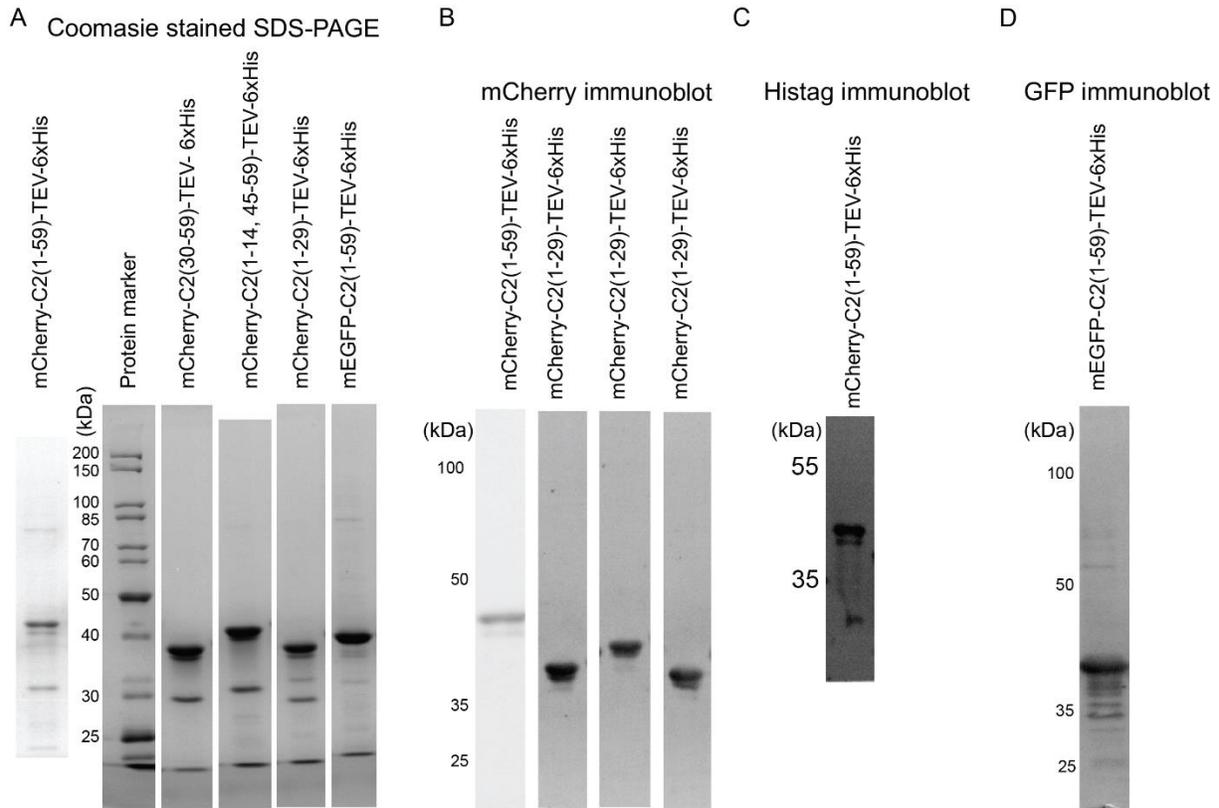
Supplementary Figure 3.1. Protein analysis of C2 fusion proteins.

A) Coomasie stained SDS-PAGE of mCherry-C2(1-59)-TEV-6xHis, mCherry-C2(30-59)-TEV-6xHis, mCherry-C2(1-14, 45-59)-TEV-6xHis, mCherry-C2(1-29)-TEV-6xHis, mEGFP-C2(1-29)-TEV-6xHis. B) Immunoblots of mCherry-C2(1-59)-TEV-6xHis, mCherry-C2(30-59)-TEV-6xHis, mCherry-C2(1-14, 45-59)-TEV-6xHis, mCherry-C2(1-29)-TEV-6xHis with an anti-RFP antibody. C) Immunoblot of mCherry-C2(1-59)-TEV-6xHis using an anti-Histag antibody. D) Immunoblot of mEGFP-C2(1-29)-TEV-6xHis using an anti-GFP antibody.

References

Choi, J.-M., Holehouse, A. S., & Pappu, R. V. (2020). Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual Review of Biophysics*, *49*(Volume 49, 2020), 107–133. https://doi.org/10.1146/annurev-biophys-121219-081629

Conboy, J. G. (2017). Developmental regulation of RNA processing by Rbfox proteins. *WIREs RNA*, *8*(2), e1398. https://doi.org/10.1002/wrna.1398

Egloff, P., Zimmermann, I., Arnold, F. M., Hutter, C. A. J., Morger, D., Opitz, L., Poveda, L., Keserue, H.-A., Panse, C., Roschitzki, B., & Seeger, M. A. (2019). Engineered peptide barcodes for in-depth analyses of binding protein libraries. Nature Methods, 16(5), 421–428. https://doi.org/10.1038/s41592-019-0389-8

Hughes, M. P., Sawaya, M. R., Boyer, D. R., Goldschmidt, L., Rodriguez, J. A., Cascio, D., Chong, L., Gonen, T., & Eisenberg, D. S. (2018). Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science*, *359*(6376), 698–701. https://doi.org/10.1126/science.aan6398

Kato, M., Zhou, X., & McKnight, S. L. (2022). How do protein domains of low sequence complexity work? *RNA*, *28*(1), 3–15. https://doi.org/10.1261/rna.078990.121
Lu, J., Ge, P., Sawaya, M. R., Hughes, M. P., Boyer, D. R., Cao, Q., Abskharon, R., Cascio, D., Tayeb-Fligelman, E., & Eisenberg, D. S. (2024). Cryo-EM structures of the D290V mutant of the hnRNPA2 low-complexity domain suggests how D290V affects phase separation and aggregation. *Journal of Biological Chemistry*, *300*(2). https://doi.org/10.1016/j.jbc.2023.105531

Misra, C., Bangru, S., Lin, F., Lam, K., Koenig, S. N., Lubbers, E. R., Hedhli, J., Murphy, N. P., Parker, D. J., Dobrucki, L. W., Cooper, T. A., Tajkhorshid, E., Mohler, P. J., & Kalsotra, A. (2020). Aberrant Expression of a Non-muscle RBFOX2 Isoform Triggers Cardiac Conduction Defects in Myotonic Dystrophy. *Developmental Cell*, *52*(6), 748-763.e6. https://doi.org/10.1016/j.devcel.2020.01.037
Murray, D. T., Kato, M., Lin, Y., Thurber, K. R., Hung, I., McKnight, S. L., & Tycko, R. (2017). Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell*, *171*(3), 615-627.e16. https://doi.org/10.1016/j.cell.2017.08.048

Murthy, A. C., & Fawzi, N. L. (2020). The (un)structural biology of biomolecular liquid-liquid phase separation using NMR spectroscopy. *Journal of Biological Chemistry*, *295*(8), 2375–2384. https://doi.org/10.1074/jbc.REV119.009847

Oldfield, C. J., & Dunker, A. K. (2014). Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annual Review of Biochemistry*, *83*(Volume 83, 2014), 553–584. https://doi.org/10.1146/annurev-biochem-072711-164947

Pollard, T. D. (2010). A Guide to Simple and Informative Binding Assays. *Molecular Biology of the Cell*, *21*(23), 4061–4067. https://doi.org/10.1091/mbc.e10-08-0683
Shin, Y., & Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. *Science*, *357*(6357), eaaf4382. https://doi.org/10.1126/science.aaf4382
Wright, P. E., & Dyson, H. J. (2015). Intrinsically Disordered Proteins in Cellular Signaling and Regulation. *Nature Reviews. Molecular Cell Biology*, *16*(1), 18–29. https://doi.org/10.1038/nrm3920

Ying, Y., Wang, X.-J., Vuong, C. K., Lin, C.-H., Damianov, A., & Black, D. L. (2017). Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell*, *170*(2), 312-323.e10. https://doi.org/10.1016/j.cell.2017.06.022

Zhao, B., Katuwawala, A., Oldfield, C. J., Hu, G., Wu, Z., Uversky, V. N., & Kurgan, L. (2021). Intrinsic Disorder in Human RNA-Binding Proteins. *Journal of Molecular Biology*, *433*(21), 167229. https://doi.org/10.1016/j.jmb.2021.167229

Zhou, X., Sumrow, L., Tashiro, K., Sutherland, L., Liu, D., Qin, T., Kato, M., Liszczak, G., & McKnight, S. L. (2022). Mutations linked to neurological disease enhance self-association of low complexity protein sequences. *Science (New York, N.Y.)*, *377*(6601), eabn5582. https://doi.org/10.1126/science.abn5582

Chapter 4: Conclusions and Future Directions

There has been a longstanding interest in deciphering a "splicing code" that determines choices of splice sites in different cell types and biological states (Bao et al., 2019; Barash et al., 2010; Fu, 2004). A key to this code will be to understand how regulatory RNA-binding proteins (RBPs) bind to their cis-regulatory RNA motifs. Most studies have focused on generating binding maps of individual RBPs and correlating their binding patterns with their activity. However, it is known that RBPs can form complexes with one another and influence how each other bind to RNA and it has been more difficult to assess how combinations of RBPs recognize their transcriptomic targets (Ule & Blencowe, 2019). In chapter 2, we demonstrate that the Rbfox1/LASR complex binds to its transcriptomic targets via interactions with RNA motifs that bind to Rbfox and the LASR subunits hnRNP M, hnRNP H/F, hnRNP C, and Matrin3. These motifs often occur in homotypic or heterotypic tandem repeats, likely facilitating a high affinity interaction by the complex through multi-site contacts of the RNA binding domains in the complex with the repeats of the RNA motifs.

We delineated the Rbfox binding sites by examining the RNA in complex with LASR bound to an RNA binding mutant F125A Rbfox1. However, we have yet to perform experimental analysis of sites bound by LASR. One approach to this problem, would be to perform CLIP on each individual subunit of LASR from the same cell type where the IP-seq was performed. This approach would allow for comparison of crosslinked sites of individual subunits of LASR to the IP-seq protected sites for the entire complex to better understand which sites protected by the complex interact with which specific subunits. Although this assay would be of great value, the CLIP would be

116

for the entire population of a given subunit and most LASR subunits exist in additional states besides LASR, complicating the interpretation of this data.

An alternative approach would be to purify the LASR complex after UV crosslinking cells and perform a tandem purification of a subunit of interest after the initial round of purification. We have already demonstrated that the complex can indeed be purified after UV crosslinking (data not shown) but have not tried a tandem purification of a subunit after this initial round of purification. A challenge in a tandem purification is to achieve high purity for the subunit of interest. A given subunit must either be tagged via an affinity tag or purified with an endogenous antibody. Purification with an endogenous antibody limits the wash conditions used in the purification. We have observed that the LASR complex is resistant to at least washes with 1M NaCl. Therefore, it is probably a better choice to use an affinity tag to increase the stringency of the wash conditions. In particular, the HaloTag and Spytag systems are attractive options as they allow covalent capture of a protein of interest and therefore washes with denaturing conditions (Guo et al., 2024).

Another approach would be to do IP-seq on LASR containing different RNA binding mutants of its subunits. This approach is not as straightforward as we have done with Rbfox, since the LASR subunits hnRNP M, hnRNP H/F, hnRNP C, and Matrin3 all contain multiple RNA binding domains. Furthermore, their RNA binding domains are less characterized than Rbfox's RNA binding domain and it is not clear how to generate mutants with reduced affinities for their motifs. Therefore, additional biochemical characterization of the RNA binding domains of the LASR proteins would be helpful before doing IP-seq with mutants.

An alternative approach would be to knockdown or knockout subunits of interest and then perform IP-seq in cells where these subunits are not expressed. These proteins can readily be knockdown down via shRNA and siRNA strategies. However, these strategies might be challenging to scale-up for IP-seq experiments that typically require more material needed than a typical knockdown experiment. Alternatively, a degron system can be applied where individual subunits are homozygously tagged with a degron tag and addition of an inducer would trigger rapid degradation of all the tagged proteins (Yesbolatova et al., 2020). We have already applied this system to test degradation of hnRNP M and have seen positive results (data not shown). A final approach would be to use CRISPR/Cas9 to knockout a subunit of interest. This approach might be complicated since some of the subunits will be essential for the cell's growth and survival. Nonetheless, there is at least one report of a HEK293 knockout line of hnRNP M (Cao et al., 2019). Intriguingly, knocking down each subunit raises additional questions: will LASR still be intact after a given subunit is absent? What are the core subunits of LASR? How does absence of a subunit affect splicing of LASR targets? These questions by themselves will constitute a large undertaking and can build an understanding of what the roles of individual subunits of LASR are.

We observed that there are cis-regulatory RNA modules bound by Rbfox1/LASR that regulate a set of exons in HEK293 cells. But it is not clear what the physiological roles of these modules in different cellular processes and cell types are. Rbfox/LASR is present in the brain. It will be interesting to do IP-seq of the Rbfox/LASR complexes in the brain and analyze their regulatory sites. Rbfox activates a set of exons in neuronal development through association with secondary motifs (Begg et al., 2020). We found

that recruitment to these secondary sites can be affected by binding of LASR to other nearby elements in HEK293 cells. Therefore, it will be interesting to look at what motifs are enriched adjacent to functional Rbfox secondary motifs in development. There has been little investigation of LASR outside of the brain. Therefore, it is important to assess the composition of LASR in different tissues. It is clear that Rbfox plays essential roles in heart, muscle, liver and the pancreas. But it is not known if it associates with LASR in these tissues and if it does what the composition of the LASR present in these tissues are. Assessing the composition of LASR subunits in different tissues and their RNA binding can provide important information about the roles that these individual subunits play.

Finally, as discussed in chapter 2, it would be interesting to assess how the self-assembly of Rbfox affects binding of Rbfox/LASR to its targets. Is each protected site bound to a single Rbfox/LASR complex or can the same complex protected multiple sites? Furthermore, how does inhibiting the self-assembly of the complex affect its target recognition and splicing activity? Determining the sequences that drive Rbfox's self-assembly as outlined in chapter 3 can guide making mutants that lack this self-assembly and subsequent tests of their binding and activity. Furthermore, if a mutant Rbfox does not homo-oligomerize yet still binds to LASR, it can be used to purify unit complexes of Rbfox/LASR that are homogenous enough for structural studies with Cryo-EM.

References

Bao, S., Moakley, D. F., & Zhang, C. (2019). The Splicing Code Goes Deep. *Cell*, *176*(3), 414–416. https://doi.org/10.1016/j.cell.2019.01.013

Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B. J., & Frey, B. J. (2010). Deciphering the splicing code. *Nature*, *465*(7294), 53–59. https://doi.org/10.1038/nature09000

Begg, B. E., Jens, M., Wang, P. Y., Minor, C. M., & Burge, C. B. (2020). Concentration-dependent splicing is enabled by Rbfox motifs of intermediate affinity. *Nature Structural & Molecular Biology*, *27*(10), 901–912. https://doi.org/10.1038/s41594-020-0475-8

Cao, P., Luo, W.-W., Li, C., Tong, Z., Zheng, Z.-Q., Zhou, L., Xiong, Y., & Li, S. (2019). The heterogeneous nuclear ribonucleoprotein hnRNPM inhibits RNA virus-triggered innate immunity by antagonizing RNA sensing of RIG-I-like receptors. *PLOS Pathogens*, *15*(8), e1007983. https://doi.org/10.1371/journal.ppat.1007983

Fu, X.-D. (2004). Towards a Splicing Code. *Cell*, *119*(6), 736–738. https://doi.org/10.1016/j.cell.2004.11.039

Guo, J. K., Blanco, M. R., Walkup, W. G., Bonesteele, G., Urbinati, C. R., Banerjee, A. K., Chow, A., Ettlin, O., Strehle, M., Peyda, P., Amaya, E., Trinh, V., & Guttman, M. (2024). Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA in vivo. *Molecular Cell*, *84*(7), 1271-1289.e12. https://doi.org/10.1016/j.molcel.2024.01.026

Ule, J., & Blencowe, B. J. (2019). Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Molecular Cell*, *76*(2), 329–345. https://doi.org/10.1016/j.molcel.2019.09.017

Yesbolatova, A., Saito, Y., Kitamoto, N., Makino-Itou, H., Ajima, R., Nakano, R., Nakaoka, H., Fukui, K., Gamo, K., Tominari, Y., Takeuchi, H., Saga, Y., Hayashi, K., & Kanemaki, M. T. (2020). The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nature Communications*, *11*(1), 5701. https://doi.org/10.1038/s41467-020-19532-z