Title

Molecular Assembly of Exon Definition Complex and Its Regulation for Alternative Precursor-messenger RNA Splicing by Polypyrimidine-tract Binding Protein PTBP1

Author

Wongpalee, Somsakul

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Molecular Assembly of Exon Definition Complex and Its Regulation for Alternative Precursor-messenger RNA Splicing by Polypyrimidine-tract Binding Protein PTBP1

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

by

Somsakul Wongpalee

2015

ABSTRACT OF THE DISSERTATION

Molecular Assembly of Exon Definition Complex
and Its Regulation for Alternative Precursor-messenger RNA Splicing
by Polypyrimidine-tract Binding Protein PTBP1

by

Somsakul Wongpalee
Doctor of Philosophy in Molecular Biology
University of California, Los Angeles, 2015
Professor Douglas L. Black, Chair

Alternative pre-mRNA splicing (AS) plays a major role in gene expression in higher eukaryotes, especially in vertebrates. Regulation of AS is primarily accomplished by splicing factors that bind to regulatory elements on pre-mRNAs and influence spliceosomal complex assembly. One important splicing factor is the polypyrimidine tract-binding protein 1 (PTBP1), well known for its repressor activity. In repressing an alternative exon, PTBP1 uses different mechanisms depending on the location of its binding sites. PTBP1 can compete with U2AF65 to bind the polypyrimidine tract at the 3' splice site if it contains the PTBP1 binding sites.

Alternatively, when PTBP1 is bound to an exon, it can block the recruitment of U2AF65 across an exon or the cross-intron interactions to an adjacent exon. However, PTBP1 binding sites are often located further within both introns flanking an alternative exon. The mechanism behind this type of alternative exon repression is poorly characterized. Therefore, we used a modified version of an alternative N1 exon of c-src pre-mRNA to characterize this type of repression mechanism. We found that PTBP1 stalls an exon in an early exon complex, inhibiting its transition to an exon definition complex (EDC), which is required for subsequent spliceosomal complex assembly. Using quantitative mass spectrometry to analyze PTBP1-repressed early exon complex, de-repressed early exon complex and the EDC, we identified that PTBP1 accomplishes this, in part, by blocking the association of two ATP-dependent RNA helicases—DDX5 and DDX17, which we found playing an important role promoting efficient formation of the EDC in vitro and exon inclusion in vivo. Interestingly, during the normal transition of the exon from early exon complex to the EDC, we detected the removal of hnRNP proteins and substantial recruitment of SR proteins, concurrent with U2 snRNP recruitment. Not only did these data reveal for the first time the assembly of the EDC, but also suggest that there are remodeling events required prior to the exon definition process. We believe that these remodeling events could be attributed to actions of ATP-dependent RNA helicases DDX5 and DDX17.

The dissertation of Somsakul Wongpalee is approved.

Feng Guo

Hong Zhou

Douglas L. Black, Committee Chair

University of California, Los Angeles

2015

DEDICATION

I dedicate my dissertation to six people I love most—My parents, sister, grandfather, Andrew Snyder and my life partner to-come. I was thinking of them often when I was writing this dissertation. My father and my mother always love me. My father made sure I received the best education available to me. My mother has been taking care and supporting me all the time and allowed me to use her kitchen to do experiments when I was young. I still remember those days when I used her pot, pan and steamer to do plant tissue cultures and ferment watermelon juice. My sister, as a big sister, scarified many things to me. Those included a first computer our parents bought for us. Her sacrifice has paid off; I am now getting a Ph.D. degree. My grandfather, a wise man, showed me the nature around our countryside home and taught me local wisdoms. He implanted a curious mind inside me. I thank Andrew Snyder for being my first love, teaching me how to love, and for running with me. It meant so much to me seeing him on every Sunday while we ran together. I love him, and I always got his back. Lastly, I thank my future life partner. I know he is somewhere out there; we just have not met. And I know this is clichéd, but I want to tell him that I work very hard but make sure I enjoy myself every single day, I travel and explore the world as much as I can, just to see him. I cannot wait to share those stories and life with him.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I want to foremost thank to the Development and Promotion of Science and Technology talented project (DPST) of Thailand, for supporting me financially and educationally since grade ten until the completion of a Ph.D. degree. Being in the program has been a tremendous eye-opening opportunity for my education and personal life. The program has supported and fostered growth of loving science inside of me. I could not be more thankful for all of this.

Six years of my graduate training would not be possible without supports from people, with whom I have interacted both personally and professionally. I have made collaboration with many experts. This has helped to enrich my graduate training and tremendously facilitate discovery of my project. Especially, James A. Wohlschlegel and Ajay Vashisht who analyzed mass spectrometry. Chapter Three is version of "Sharma S, Wongpalee SP, Vashisht A, Wohlschlegel JA, Black DL. *Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly*. Genes & Development. 2014 Nov 15;28(22):2518–31. PMCID: PMC4233244". Shalini Sharma led this study under the supervision of Professor Douglas L. Black. I contributed to the study by identifying and verifying a protein that interacts with stem-loop 4 of U1. James A. Wohlschlegel and Ajay Vashisht, again, helped with mass spectral analyses.

I would like to thank all members of the laboratory of Professor Douglas L. Black, who has helped me with scientific discussion and troubleshooting, and have made pleasant working environment. Importantly, I thank to my boss Professor Douglas L.

Black and Shalini Sharma who have provided all supports and time to train me throughout graduate school.

Finally, I would like to thank Pagkapol Yhew Pongsawakul who has helped me in many aspects since we were both undergrads at The University of Virginia until now. Andrew Ah Young and Maxime Chapon who are my very good friends; I always feel like being home at their place. I thank Andrew Snyder for love, Kay Kyu-Hyeon Yeom and Adrian Hernandez for tremendous supports both personally and work-wise. Thanks all my friends Wen Gu, Wing Lung, Anna Reichardt, Gayle Boxx, Eriko Shimada, Susan Realegeno, Sara Weitz, Jenny Link, Marisabel Oliveros, Sara Weitz, Christian Aguilera, Naphat Chantaravisoot, Sirikarn Surawanvijit, Warin Isvilanonda, Natee Pitiwan, Chitrada Kaweeteerawat, Kevin Roy, Jaspreet Sandhu, Albert Sek, Hoa Bui and Willy Lee, for friendship and all helps. Thank you so much.

VITA

Education

University of California at Los Angeles (UCLA), Los Angeles, CA 2009 - 2015

Molecular Biology Interdepartmental Ph.D. Program (MBIDP)

GPA. 4.00

University of Virginia (UVa), Charlottesville, VA 2005 - 2009

B.S. in Biology with *Highest Distinction*

GPA. 3.92

Honors, Fellowships & Awards

Paul D. Boyer Outstanding Teaching Award, UCLA 2014

Distinguished Major in Biology with *Highest Distinction*, UVa 2009

David A. Harrison III Undergraduate Research Award, UVa 2008

Intermediate Honor, College of Arts and Sciences, UVa 2007

Echols Scholar, College of Arts and Sciences, UVa 2006 - 2009

College Science Scholar, College of Arts and Sciences, UVa 2005 - 2009

Dean's list, College of Arts and Sciences, UVa 2004 - 2006

Royal Thai Scholarship, Royal Thai Government 2004 - 2014

Thailand Representative for SEMEO Young Scientist, Malaysia 2004

Publications

1. Wongpalee SP, Sharma S. The pre-mRNA splicing reaction. Methods Mol. Biol. 2014;1126:3–12.
2. Sharma S, Wongpalee SP, Vashisht A, Wohlschlegel JA, Black DL. Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. Genes & Development. 2014 Nov 15;28(22):2518–31. PMCID: PMC423324
3. Negishi M, Wongpalee SP, Sarkar S, Park J, Lee KY, Shibata Y, Reon BJ, Abounader R, Suzuki Y, Sugano S, Dutta A. A new lncRNA, APTR, associates with and represses the CDKN1A/p21 promoter by recruiting polycomb proteins. PloS one. 2014;9(4):e95216. PMCID: PMC3991591
4. Kunnaja P, Wongpalee SP, Panthong A. Evaluation of anti-inflammatory, analgesic, and antipyretic activities of the ethanol extract from Murdannia loriformis (Hassk.) Rolla Rao et Kammathy. Bioimpacts. 2014;4(4):183–9. PMCID: PMC4298709

CHAPTER ONE

Introduction

PRE-SPLICEOSOMAL COMPLEX ASSEMBLY

Most eukaryotic protein-coding genes are transcribed into precursor-messenger RNAs (pre-mRNAs) in which the two terminal untranslated regions and the protein-coding regions, together called exons, are interrupted by noncoding intervening segments called introns. This pre-mRNA must undergo an RNA processing reaction called splicing that removes introns and ligates the exons to generate a mature translatable mRNA. All of the introns must be excised from a gene transcript prior to the export of the mature mRNA from the nucleus to the cytoplasm for translation. Splicing occurs in two trans-esterification reactions that are catalyzed by a large ribonucleoprotein (RNP) complex called the spliceosome.

The spliceosome is one of the most complex macromolecular machineries in eukaryotic cells, characterized by the dynamic assembly and disassembly of more than ~170 proteins and five Uridine rich small-nuclear RNPs (U snRNPs) in a single splicing reaction (1,2). This dynamics is driven largely by ATP-dependent reactions mediated by eight DExD/H-type RNA-dependent ATPase/helicases conserved from yeast to human. These proteins participate in different stages of splicing to remodel protein-protein, protein-RNA, and RNA-RNA interactions to create catalytic sites that consist of both snRNAs and Prp8 proteins of U5 snRNP (3,4). The U snRNPs function to mark intron/exon boundaries and to engage the spliceosome in catalytic activities. Much of our knowledge of the spliceosome is obtained from the ability to reconstitute splicing in

vitro using optimized intron-defined pre-mRNA substrates—containing a small intron flanked by partial exons—and ability to purify spliceosomal intermediates.

Pre-spliceosomal complex assembly (E and A complexes) is of great interest because it is known to be a target for alternative pre-mRNA splicing (5-8). At these early stages, splice sites are not fully paired together, so many splicing factors exploit this chance for alternative splicing. In vitro, the spliceosome assembles in a step-wise manner to form discrete intermediates that are different in size and shape. These intermediate complexes are termed E, A, B, B* and C complexes. Early stages of splicing involve identifying 5' and 3' splice sites. In forming the pre-spliceosomal E complex, U1 snRNP base-pairs to the 5' splice site, while U2AF65 and U2AF35 interact with polypyrimidine-tract and 3' splice site on the opposite end of an intron, respectively. It has been shown that during the E complex formation, the two splice sites already communicate across an intron and are brought close together (9,10). This pairing interaction is likely mediated by many proteins such as SR proteins, which have been shown to interact with both U1 snRNP and U2AF35 (11-13). This complex is the earliest complex that commits the pre-mRNA to the splicing pathway (14). U2 snRNP is then recruited to the branch site sequences at least by interaction with U2AF65 (15,16). In base-pairing with the branch point sequences, U2 snRNP requires the first two ATP-dependent reactions—1) ATP hydrolysis by UAP56, which is also recruited by U2AF65, 2) ATP hydrolysis by DDX46 (Prp5), a protein integral to the U2 snRNP. Precise activity of UAP56 is unclear, however it genetically interacts with proteins that bind to branch point sequence in yeast, suggesting that it might remove those proteins prior to U2 base-pairing (17,18). In contrast, activity of DDX46 is well known, as it is required to

expose a part of U2 snRNA for base-pairing at the branch point sequences (19,20). Integration of the U2 snRNP in the complex results in the formation of the pre-spliceosomal A complex. The 5' and 3' splice sites are shown to be committed to each other in this complex (21). Consistently, additional interaction between U1 snRNP : DDX46 : U2 snRNP and U1 stem-loop4 : U2 snRNP (refer to Chapter 3) are also found to mediate the cross-intron interaction in addition to those mentioned above (22,23). The pre-spliceosomal A complex then recruits tri-snRNPs U4/U6.U5 and Prp19/CDC5 to form the full spliceosomal B complex. The B complex is not catalytically active. Rather, it undergoes substantial rearrangement of both protein and U snRNAs. This results in a loss of U1 and U4 snRNPs from the B complex, U2 snRNA now base-pairs to U6 snRNA; U6 base-pairs to 5' splice site, while U5 snRNA interacts with a 5' exon (2). This resulting complex is call the B* complex and is active for catalyzing the first trans-esterification reaction where the branch point Adenosine attacks the 5' splice site. The first catalytic reaction generates the spliceosomal C complex which completes the second reaction where the 5' splice site attacks the 3' splice site in order to ligate two exon together and release an mRNA. A by-product intron lariat is attached to the post-spliceosomal intron complex and is released as the U2, U5 and U6 snRNPs are recycled for the next round of splicing (2).

EXON DEFINITION

As mentioned above, our current understanding of spliceosomal complex assembly is based on experiments using small intron pre-mRNA containing partial terminal exons, in which the 5' and 3' splice sites are recognized across a short intron in

the pre-spliceosomal E and A. This early assembly is termed intron definition (24). However, in higher eukaryotes like human, average exon length is only 170 nucleotides (nts), separated by considerably long introns; thus, recognizing splice sites across a vast intronic sequences seems to be an incredible pathway. Rather, early spliceosomal complexes in vertebrates are thought to assemble across an exon during a process called exon definition (25,26). Once an exon is recognized as an exon definition complex (EDC), adjacent EDCs will have to pair together across an intron during the process of intron definition, which will be then feed to the intron-defined spliceosomal pathway described above (26,27).

A large body of evidence supports the existence of exon definition in vertebrates (26,28). First, 5' splice site mutation of an internal exon commonly leads to skipping of the exon as well as both flanking introns, rather than retention of the affected intron. Second, if 5' splice site mutation does not lead to exon skipping, a new cryptic 5' splice site is almost invariably located very close to the affected 5' splice sites, suggesting there is a strict communication with an upstream 3' splice site in order to splice in. Third, there is a limit of exon length to support efficient splicing in human nuclear extract, with exon greater than 300 nts splice poorly. This coincides with average exon length in vertebrate, which is less than 300 nts. Lastly, proteins that mediate cross-exon interaction such as SR proteins and U2AF35 are missing in *Saccharomyces cerevisiae* in which their introns are spliced out via intron definition, suggesting the requirement of these proteins in higher eukaryotes.

Exon definition is the term coined by Susan Berget's group in 1990 after an observation that adding a 5' splice site to the second exon enhanced splicing of the

upstream intron (29). Exon definition is a process that vertebrates use for recognizing 5' and 3' splice sites across an exon. This process yields a pre-spliceosomal A-like "α" complex, which is subsequently known as EDC (24,26). However, the term became loosely defined as any RNP complexes assembled on an exon, for example a pre-spliceosomal E-like EDC, containing U1 snRNP. Recently, the EDC was found to contain the tri-snRNPs; this type of EDC is readily to transition to the intron-defined B complex (30). EDC herein will be referred to as an original, pre-spliceosomal A-like α complex, containing U1 and U2 snRNPs. Besides the two U snRNPs, mass spectrometry analyses of isolated EDC from two different exons confirmed that it also contains prominent numbers of Serine-Arginine protein family (SR proteins) and heterogeneous nuclear ribonucleoproteins (hnRNPs), along with other miscellaneous proteins (8,30). SR proteins bind exon sequences, positively affecting recognition of 5' and 3' splice sites and promoting communication between them (31-34). However, the presence of hnRNP proteins in the EDC is not well understood as these proteins commonly known for their negative affect on splicing (35).

Nevertheless, steps to assemble an exon to the EDC are poorly characterized. Besides some pairwise pulldown experiments, an ensemble of protein-protein interaction inside the complex have never been studied. As pre-mRNA splicing in vertebrates employs exon definition process, growing evidence suggest that exon definition is a target for many splicing factors for alternative pre-mRNA splicing (5,6,34,36,37). Thus, dissecting the mechanisms underlying EDC assembly, in particular the protein-protein, protein-RNA and RNA-RNA interactions that are

necessary for complex formation, is a prerequisite for understanding how alternative splicing is accomplished.

SR PROTEINS AND HNRNP PROTEINS

Serine-Arginine protein family (SR protein family) is a family of nuclear phospho-proteins that have many functions in RNA metabolism. They participate in mRNA export, nonsense-mediated mRNA decay, and translation (38-40). Additionally, SR proteins are known for their positive effects on both constitutive and alternative pre-mRNA splicing (27), even though they can have repressive functions (41). The SR protein family is characterized by the presence of one or two RNA binding domains termed RNA Recognition Motif (RRM) on the N terminus, followed by RS (Arginine-Serine rich) domains of at least 50 amino acid with >40% RS content (42). Twelve SR proteins, SRSF1 to SRSF12, were identified in human. Other RS domain-containing proteins exist, but do not belong to the SR protein family because of different domain arrangement, the absence of RRM, or the presence of different RNA binding domain; these proteins are often called SR-related proteins. Some examples of SR-related proteins include TRA2B (SFRS10), RBM39 and LUC7L2. RRM domains mediate binding of SR proteins to RNA targets, whereas RS domains mediate protein-protein interactions among themselves or with RS domains of other proteins (11,13,42). RNA binding motifs of SR proteins have been identified; consensus of these motifs is generally low for a particular SR protein. However, it has been shown that existence of these motifs is higher in an exon compared to intron (41,43). Many studies, using either exon or intron-defined substrates, have demonstrated SR proteins promotes recruitment

of U1 snRNP, U2AF and U2 snRNP to 5', 3' splice sites and branch point sequences, respectively (11,12,15,32,33,44-47). Therefore, given their importance in splicing and their preferred association with exon sequences, SR proteins are believed to promote exon definition by recruiting components of splicing machinery and mediating interaction across an exon. However, when do the SR proteins function, how many SR proteins needed for a given exon, what interaction do they make and what are their relationship with hnRNPs proteins bound to an exon—are open questions that have been intensely investigated.

hnRNP (heterogeneous nuclear ribonucleoprotein) proteins were originally identified as a group of most abundant proteins that associated with hnRNA (pre-mRNA) in vivo (35,48). These originally identified proteins were named 'hnRNP' followed by an alphabet starting from 'A', to indicate their identity (48). Later, more hnRNP proteins were identified; some of them did not follow such nomenclature. Because this operational classification does not reflect their structural identity and biochemistry, it is not surprised that hnRNP class does not have certain numbers of its members and that hnRNP proteins exhibit different structural and functional diversities. They participate in various aspects of RNA metabolism such as telomere biogenesis, polyadenylation, translation, RNA editing and RNA stability (35). hnRNP proteins are also known repressors of alternative splicing, however some studies have reported a role in splicing activation (35,49). Like SR proteins, hnRNP proteins contain a RNA binding domain, specifically RRM. Some hnRNP proteins, such as hnRNPE1 (PCBP1) and hnRNPK, possess KH (K homology) domain as an RNA interacting domain. hnRNP proteins also harbor RGG boxes (repeats of Arginine-Glycine-Glycine), and additional

glycine-rich, acidic amino acids-rich or proline-rich domains (50). Binding sites of hnRNP proteins are typically found inside flanking introns or inside an exon. Upon binding to such sites, hnRNPs interact with other proteins through the RGG boxes or other domains, and employ various mechanisms for splicing inhibition. They can bind inside an exon and prevent cross-exon interactions, which are essential for U2AF recruitment, thus inhibiting formation of the EDC. An example of this type of repression is seen in Fas exon 6 inhibition by PTBP1 (5). hnRNPL binds inside an alternative exon 4 of CD45 pre-mRNA and targets the EDC. However, rather than blocking assembly of the EDC, hnRNPL inhibits a transition from exon definition to intron definition (6). As mentioned above, SR proteins are involved in process of exon definition and communication between the two splice sites across an exon. Thus, it is not surprising to find that hnRNP proteins antagonize effects of SR proteins for splicing repression (51-54). Some hnRNP proteins share the same binding sites as SR proteins, thus, directly competing out SR proteins from their binding sites, such as hnRNPA1 and SRSF1 (55,56). However, many times hnRNP proteins do not share binding site with SR proteins. How this splicing repression is achieved remains to be addressed.

As exon definition is the major pathway for recognizing splice sites in vertebrates, it is emerging that an ultimate goal of splicing repression is to disrupt this process, which many times involve intervening with functions of SR proteins as mentioned. Still, it is likely that hnRNP proteins could employ other mechanisms to achieve the same goal. Given divergent sequences splice sites, exonic and intronic sequences in vertebrates, every exon is expected to have a different splice site strength, sets of SR proteins, hnRNP proteins and other RNA binding factors. A sum of these positive and

negative factors will determine splicing outcomes. Proteomic approach on various model alternative exons under normal and repressive conditions will lead to understanding steps during assembly of the EDC, and may reveal novel factors necessary for EDC activation or repression as well as a common theme of how repressors accomplish splicing repression.

PTBP1 AND ALTERNATIVE SPLICING

Polypyrimidine-tract binding protein 1 (PTBP1) or PTB or hnRNPI, are a member of hnRNP group of proteins that participates in translation, polyadenylation and alternative pre-mRNA splicing regulation (57). PTBP1 was first identified in HeLa nuclear extract as a protein that crosslinked to a polypyrimidine tract of a 3' splice site (58). PTBP1 has an N terminal nuclear shuttling domain and four RNA-binding domains. These RNA binding domains are RRM type, which has a common fold βαββαβ and are identified by two conserved RNP (Ribonucleoprotein) motifs (57,59). PTBP1 binds to a tract of polypyrimidine containing mixture of C and U. This property stems from specificity of each domain that can bind at least a triplet of a mixture of C and U nucleotides (59). With an exception of domain 3 and 4, each domain is separated by a linker, making it independent from one another in binding target sequences (59). Domain 3 and 4 interact with each other at a fixed conformation, orienting their RNA binding surface away from each other. Therefore, in order for the two domains to simultaneously bind to RNA targets, a spacer or a loop of RNA of at least 15 nucleotides are required (57). This is potentially implicated in PTBP1 repression as it is proposed that PTBP1 creates a repressive RNA loop during splicing

repression. In fact, mutational analysis indicated that residues that mediate the interaction between domain 3 and 4 are important for efficient repression (60,61).

PTBP1 is conserved from Drosophila to human (57). This is not surprising, as PTBP1 is well known for its function in regulation of alternative pre-mRNA splicing, an essential mechanism of gene regulation in higher eukaryotes. In vertebrate, PTBP1 is highly expressed in most tissues, except for neuronal tissues, which express a PTBP1 paralog, nPTB (PTBP2). High expression of PTBP1 in non-neuronal tissues is necessary to suppress neuronal gene expression programs in those tissues. It has been shown that knockdown of PTBP1 can trans-differentiate fibroblasts into functional neurons (62). In hematopoietic cells, another PTBP1 paralog, PTBP3 or ROD1 (Regulator of differentiation) is expressed (63).

PTBP1 is most known for its function as a repressor of alternative pre-mRNA splicing. Some studies have also suggested that it can function as an exon activator for certain exons (64,65). Genome-wide crosslinking experiments suggested that as much as 20% of alternative splicing events associate with PTBP1 binding (64). Although the contribution of PTBP1 to all these events is not fully characterized, the number gives an impression of how prevalent PTBP1 associates with pre-mRNA processing. Studies by Xue et al and Llorian et al also confirmed that repression by PTBP1 generally needs multiple binding sites. One of these sites is usually present within an upstream intron; the other will be either inside an exon or a downstream intron. Biochemical studies using various model alternative exons have suggested that PTBP1 is likely to employ different mechanisms during repression, depending on location of its binding sites. In one mode of repression, PTBP1 competes with U2AF65—a general splicing factor—for

binding to a polypyrimidine tract, which is required for 3' splice site recognition. As a result, a 3' splice sites is not recognized, and an exon is skipped (60,66-68). In another mode, an exon 6 of Fas pre-mRNA contains therewithin a PTBP1-binding site. Binding of PTBP1 disrupts U1-dependent U2AF65 recruitment to an upstream polypyrimidine tract (5,69). Another observation on an alternative exon of BCL-X pre-mRNA, found that an SR protein SRSF1 is less cross linked to the exon in the presence of PTBP1 (70). Together the two studies suggested that there are potential interactions inside alternative exon that is inhibited in the presence of PTBP1. In addition, another mode explains binding of PTBP1 to an exon can result in loss of cross-intron interaction between an alternative exon and a downstream exon (7,8). Adding to this complexity, for many alternative exons, PTBP1-binding sites can be distal to repressed exons, and well separate from the polypyrmidine tract (60,64,65). A repressive RNA loop created by PTBP1 has been proposed as a model. This loop is thought to physically interfere with splice site recognition during pre-spliceosomal complex assembly. Still, there is no clear biochemistry data to support existence of an RNA loop during the repression. Nevertheless it may be true that PTBP1 can employ a more sophisticated mechanism to repress an exon other than the aforementioned, given that it can interact with other proteins harboring PTB-RAVER1 Interacting motif (PRI) (71,72) and that repression of some exons require cofactors such as RAVER1 and MBNL1 (71,73,74).

CHAPTER TWO

Molecular Assembly of Exon Definition Complex And Its Regulation for Alternative Precursor-messenger RNA Splicing by Polypyrimidine-tract Binding Protein PTBP1

INTRODUCTION

Alternative pre-mRNA splicing (AS) is a key step during gene expression that shapes transcript repertoire in higher eukaryotes. More than 95% of human multi-exon genes undergo AS (75). This generates an extremely large variety of mRNA isoforms that code for functionally different proteins or mRNA isoforms that have different stabilities. The process of AS is tightly regulated by a combination of splicing factors that are expressed in specific cell types, during different developmental stages, or in response to particular signaling cues (76). Therefore, it is not surprising that misregulation of splicing events is implicated in a variety of diseases, including cancer. Emerging evidence from genome-wide studies have shown large-scale alterations of AS in tumorigenesis (77,78).

In higher eukaryotes, as soon as pre-mRNAs emerge from transcribing RNA polymerase II, many RNA-binding proteins—including an abundant class of protein hnRNP (heterogeneous nuclear ribonucleoprotein)—associate with the pre-mRNA (79). These proteins package the pre-mRNA for subsequent processing such as pre-mRNA splicing and export. In order for two exons to be spliced and then joined to each other, each must first be assembled into a modular ribonucleoprotein exon complex, called an exon definition complex (EDC) (26,29). This pre-spliceosomal complex contains U1

and U2 snRNPs bound to 5' and 3' splice sites, respectively, as well as many other proteins (8,26,30). One class of proteins is the Serine-Arginine (SR) rich protein family, which is thought to bind early to exons via Exonic Splicing Enhancer sequences (ESE), and promote splicing in both constitutive and alternative exons (34,42,80). The binding motif of each SR protein has fairly low consensus, but studies have shown that those sequences are more enriched in the exonic regions (43). Once bound, SR proteins promote binding and interaction between 5' splice sites-bound U1 snRNP and 3' splice sites – bound U2AF and U2 snRNP through their RS domain (81). After being assembled, neighboring EDCs are subsequently joined together across introns during assembly of a pre-spliceosomal A complex (cite). Unlike the assembly of the intron pre-spliceosomal A complex, assembly of the EDC has never been characterized, and it is unclear whether or not EDC proceeds in the same order as the A complex. Most our knowledge of the EDC has been obtained from binary interactions or pull-down studies of major components of the EDC, which lacks temporal and quantitative aspects. Therefore, it is necessary to understand how the EDC is assembled as this will shed light on the mechanism of splicing in higher eukaryotes and how alternative splicing is regulated. Recent studies have also shown that the EDC is a potential target for alternative splicing (5,6,37).

The most common mechanism for AS employs RNA-binding proteins that bind to short elements on a pre-mRNA to alter subsequent spliceosome assembly. One such regulator is the polypyrimidine-tract binding protein (PTBP1), a key splicing factor that represses inclusion of alternatively spliced exons found in many genes (64,65). In humans, a significant number of AS events are associated with PTBP1 (64). The

protein is involved in splicing programs during neuronal development and also in the pathology of some cancers (62,78,82). PTBP1 contains four RNA Binding domains; each can bind to triplets of C and U combinations (83). Repression by PTBP1 requires multiple binding sites, and a repression mechanism being employed is likely to depend on where those sites are located (60,84). In one mode of repression, PTBP1 competes with U2AF65—a general splicing factor—for a polypyrimidine tract required for 3' splice site recognition. As a result, a 3' splice site is not recognized, and an exon is skipped. In another mode, an exon 6 of Fas, a cell surface death receptor, contains PTBP1-binding sites. Inclusion of exon 6 leads to apoptosis, while skipping of the exon generates a soluble isoform of the protein and prevents apoptosis (5). Binding of PTBP1 disrupts U1-dependent U2AF65 recruitment to an upstream polypyrimidine tract (5). Resulting questions remain on how PTBP1 disrupts the communications inside of exon 6, specifically between U2AF65 and U1 snRNP, what other proteins are hindered as a result, and why this repression is dependent on the presence of U1 snRNP. Adding to this complexity, for many alternative exons, PTBP1-binding sites are distal to repressed exons, well separated from the polypyrmidine tract (60,64,65). A repressive RNA loop created by PTBP1 has been proposed as a model, but no biochemical experiment has ever verified if PTBP1 loops an exon during repression. Even this does happen, it alone does not explain the repression mechanism at a molecular level. A lot remains unexplained regarding how PTBP1 mediates silencing of this type of target exons.

Using a modified alternative N1 exon of Src pre-mRNA with PTBP1-binding sites located at least ~50 nts away flanking the N1 exon, we discovered that PTBP1 inhibited

reactivity of both 5' and 3' splice sites of the exon. Binding of PTBP1 blocked an assembly of the EDC by preventing recruitment of U2 snRNP. Interestingly, we found that this blockage occurred even if U2AF65 and U1 snRNP are normally bound binding to the 3' and 5' splice sites, respectively. We purified exon complexes assembled in vitro using RNA containing either wild-type PTBP1 binding sites or mutant sites, and analyzed their constituents using Stable Isotope Labeling by Amino acids in Cell culture coupled with Mass Spectrometry (SILAC-MS). This allowed us to quantitatively examine the changes in the composition between three different states of exon complexes: 1) PTBP1-Repressed (an early exon complex from the wild-type RNA), 2) De-repressed (an early exon complex from the mutant RNA) and 3) the EDC (EDC from the mutant RNA), which resembles a spliceosomal A complex in their mobility and composition. Following the assembly of the EDC from the De-repressed complex, we identified three major structural rearrangements: 1) eviction of hnRNP proteins 2) massive recruitment the SR proteins and 3) U2 snRNP binding to the exon. In the repressed exon complex, PTBP1 appears to cooperate with the proteins MBNL1 and RAVER1 to prevent binding of some positive factors that may be needed prior to the formation of the EDC. Those factors include ATP-dependent RNA helicases, DDX5 and DDX17. We found that these two latter proteins function to promote formation of the EDC in vitro only in the absent of PTBP1. Together, our data reveal for the first time the dynamic assembly of the EDC and may suggest an unprecedented early ATP utilization by the helicases prior to binding of the U2 snRNP to the branch point sequence. Importantly, this suggests that PTBP1 silences an exon by preventing assembly of the

EDC, which is achieved in part by modulating protein compositions and remodeling inside of the early exon complex needed for exon definition.

MATERIALS AND METHODS

Plasmid Construct and In vitro Transcription

An N1 exon and its flanking introns of mouse Src gene (85) was cloned into downstream of a T7 promoter. Plasmids containing N1 exon with either wild-type PTBP1-binding sites (named: N1 of BS719 plasmid) or mutant PTBP1-binding sites (named: N1 of BS719 plasmid) were amplified in E.coli. The plasmids were digest with NotI restriction enzyme to generate templates for in vitro transcription. In vitro transcription reaction was performed at 37°C for 3 hours, with transcription buffer (NEB) and homemade T7 RNA polymerase according to Sharma et al 2008 (8). All transcripts were initiated with an RNA cap analog m7G(5')ppp(5')G (NEB) and uniformly labeled with ^{32}P-α-UTP (Perkin Elmer). RNA transcripts were gel purified in 4% 7.5M urea-PAGE, eluted at room temperature for 3 hours in a gel elution buffer (0.3 M sodium acetate pH 5.2, 1 mM EDTA pH 8.0, 0.1% (v/v) SDS). RNA transcripts were precipitated with ethanol overnight, spun down, and washed with 75% ethanol. Pellets were air dried and re-suspended in DEPC-treated water. For 3X MS2 stem-loop-tagged transcripts, re-suspended RNAs were heated at 65°C for 5 minutes in a heat block and slowly cooled down to 30°C in order to fold the RNAs.

To generate transcription templates for 5' and 3' exons of Adenovirus Major Late pre-mRNA, T7 promoter-containing forward primers and reverse were used to PCR amplified the templates. Both 5' exon and 3' exon templates encompassed at least 50

bps into the intron from 5' and 3' splice sites, respectively. The PCR amplicons were gel purified in native agarose, gel extracted and used for in vitro transcription as above.

Exon Complex Assembly and Native Agarose Gel

Concentration of ^{32}P labeled transcripts was measured using Nanodrop; molar concentration was calculated. RNP complex was assembled in a standard splicing reaction in 20-uL reaction containing 1 nM of RNA transcript, 2,2 mM $MgCl_2$, 0.4 mM ATP, 20 mM Creatine Phosphate, 1 U/uL RNAseOUT (Life Technology) and 60% v/v HeLa S3 nuclear extract (prepared as in Sharma et al 2008 (8)). The reaction was incubated at 30°C for 30 minutes, or various time as indicated in each figure. 0.067 ug/uL heparin was added to stop the reaction, and the reaction was incubated further at 30°C for 5 minutes. About 10 uL of the reaction was loaded onto 2.5% w/v agarose GTG (Lonza) casted with 25 mM Tris-Glycine pH 8.8. The gel was run using the same buffer at 100 volts for 3 hours at room temperature. The gel was then fixed in 5% methanol-acetic acid for 30 min before vacuum drying for 1 hour. Afterward, heat drying was applied for 2 hours. The gel was exposed to a phosphorus screen (Molecular Dynamics) overnight and scanned using Typhoon (GE).

Trans-splicing

Exon complexes were assembled on cold transcripts (non ^{32}P labeled)—WT N1 exon, mut N1 exon, 5' Adeno ML exon and 3' Adeno ML exon—under a standard splicing reaction for 20 minutes (see above) with the following exception: 1) all reactions was 10 uL and 2) 2 nM of N1 exons and 50 nM of Adeno ML exons were used for

assembling exon complexes. After assembling, equal amounts of N1 exon complexes were mixed with either 3' Adeno ML exon complex (for 5' splice site test) or 5' Adeno ML exon complex (for 3'splice site test). The combined reactions were supplemented with another dose of ATP and Creatine phosphate, and further incubated at 30° C for 3 hours. The reactions were treated with 0.5% v/v SDS and Proteinase K (Roche) for 20 minutes at 37° C. Total RNA was extracted from the reactions by acidic phenol (Sigma). RNA was ethanol precipitated (as above) and re-suspended in DEPC-treated water. RNA was used for primer extension analysis with 5' ^{32}P labeled primer annealing to 3' end of spliced products. The primer extension products were resolved in 10% 7.5M urea-PAGE, heat dried and exposed to a phosphorus screen overnight as above.

Purification of Exon Complexes

Wild-type and mutant exons containing 3X MS2 stem-loops at 3' end were pre-incubated with a fusion protein of MS2 bacteriophage coat protein and Maltose-binding protein (MS2-MBP) before assembling exon complexes. The RNAs were assembled in large scale splicing reaction with 10 nM RNA under a standard splicing reaction as above for 30 minutes. The reaction was chilled on ice for 5 minutes before spinning at 20,000 g for 10 minutes at 4° C to remove large particulates. Heparin was omitted. Each 400 uL of the reaction was layered on top of one 15-45% w/v glycerol gradient prepared with 20 mM HEPES pH 7.9, 80 mM potassium glutamate, 2.2 mM $MgCl_2$ and 0.2 mM EDTA pH 8.0, using BioComp station. The gradient was spun in SW41 rotor (Beckmann) at 37,000 rpm at 4° C for 15:30 hours, then fractionated into 25 fractions

using BioComp station. Abundance of complexes was determined in each fraction using scintillation counting.

Fractions containing relevant exon complexes were pooled together and subjected to MBP affinity purification at 4° C. To this end, the pooled fractions were passed three times through appropriate amount of amylose beads (NEB) that had been pre-equilibrated in a wash buffer (20 mM HEPES pH 7.9, 80 mM potassium glutamate, 2.2 mM $MgCl_2$ and 0.2 mM EDTA pH 8.0). The beads were washed with 20 column volumes with the wash buffer. Exon complexes were eluted with one column volume of wash buffer containing 40 mM maltose by incubating at 4° C for 30 minutes. Yields of the eluted complexes were determined by scintillation counting against known amount of RNA substrates.

4′-aminomethyl Trioxsalen (AMT) Crosslinking

20 ug/mL of AMT was directly added to exon complexes after splicing reaction was finished. Crosslinking was initiated by irradiating the mixture directly with 366 nm UV lamp on ice for 15 minutes. Tubes' lids were kept open throughout the crosslinking procedure. After crosslinking, the mixture was treated with 0.5% v/v SDS and Proteinase K (Roche) for 20 minutes at 37° C. Total RNA was extracted from the reactions by acidic phenol (Sigma). RNA was ethanol precipitated (as above) and re-suspended in DEPC-treated water. RNA was used for primer extension analysis, with 5' ^{32}P labeled primer annealing to an intron downstream of the N1 exon. The primer extension products were resolved in 8% 7.5M urea-PAGE, heat dried and exposed to a phosphorus screen (Molecular Dynamics) overnight as above.

Primer Extension

All primer extension analyses were as same as in Carey et al 2013, except that SuperScript III (Life Technology) was used and that incubation temperature was raise to 50° C (86). After primer extension, the extension products were extracted with phenol:chloroform:isoamyl alcohol 25:24:1(Sigma) and ethanol precipitated. Pellets were washed with 75% ethanol, air dried and re-suspended in RNA loading dye. The product were heated at 65° C before resolving in 8% 7.5M urea PAGE. The gel was heat dried and exposed to a phosphorus screen overnight as above.

DNA-directed RNAseH Cleavage of U1 and U2 snRNPs

To specifically degrade U snRNPs, complementary DNA oligos were synthesized from IDT company. U1 and U2 oligos used were targeting position 1-15 nts of U1 and U2 snRNA (87). The DNA sequences were following: U1 = CTGCCAGGTAAGTAT; U2 = AGGCCGAGAAGCGAT. DNA oligo target GAPDH was used as a negative control: GAGGTCAATGAAGGGGTCAT. Procedure for DNA-directed RNAseH cleavage was as same as used by Merendino et al 1999, except that RNAseH from NEB was used (88). The treated nuclear extracted were then directly used for experiments or stored at -80° C.

SILAC and Preparation of HeLaS3 Nuclear Extract

For SILAC experiments, minimum essential medium Eagle (similar to M8028, Sigma) without sodium bicarbonate, glutamate, arginine, and lysine was ordered from AthenaES. 13C-arginine (R6) and 13C-lysine (K6) were purchased from Cambridge

Isotope Laboratories. The complete medium was reconstituted with 2 mg/mL sodium bicarbonate, 4 mM glutamate, 1X nonessential amino acids (Life Technology), 5% PBS-dialyzed newborn calf serum (Omega Scientific), 1X penicillin–streptomycin (Life Technology), and either ‘‘light’’ (R0K0) or ‘‘heavy’’ (R6K6) arginine and lysine at 12.6 g/L and 7.25 g/L, respectively. HeLa S3 cells were grown in medium containing either R0K0 or R6K6 for five to six doubling cycles to obtain high labeling efficiency (>99%). The cells were harvested in log phase, processed for nuclear extraction, and stored at 80°C as described Dignam 1990 and Black 1992 (85,89).

Mass Spectrometry

Protein mixtures were reduced, alkylated, and digested by the sequential addition of lys-C and trypsin proteases as previously described (90). The digested peptide mixture was fractionated online using strong cation exchange and reverse-phase chromatography and eluted directly into a LTQ-Orbitrap mass spectrometer (Thermofisher) (90,91). Tandem MS (MS/MS) spectra were collected and subsequently analyzed using the ProLuCID and DTASelect algorithms (92,93). Database searches were performed considering both the light and heavy versions of each peptide. Protein and peptide identifications were further filtered with a false positive rate of <5% as estimated by a decoy database strategy (94). NSAF values were calculated as described (95). SILAC ratios for each peptide pair were determined using the Census algorithm (96,97).

siRNA Knockdown, RNA Extraction and cDNA synthesis

A reporter minigene—wild-type N1 exon that was flanked by constitutive exon 1 and 2 of human beta globin—was inserted into Flp-In T-REx HeLa cells (Life Technology) to generate stable cells by following a manufacturer's protocol. The reporter was under the control of Tet repressor. After stable HeLa cell lines were established, they were plated at $5x10^5$ cells/plate (in 6-cm plate) in DMEM+10%FBS medium, and reverse transfected with siRNAs (see sequences below) using RNAiMax (Life Technology) according to manufacturer's protocol. Final concentration of siRNA was 18.75 nM in the medium. The cells were maintained in an incubator at 37° C, 5% CO_2 for 48 hours before inducing expression of the reporter with 1 ug/mL doxycycline. The cells were induced to eight hours and then directly lysed with 1 mL TRIzol (Life Technology) for total RNA. The lysates could be immediately used or kept at -80° C. To extract RNA, TRIzol lysates were mixed with 1 volume of 100% ethanol and preceded according to Direct-zol RNA miniprep (Zymo Reseach)

One ug of total RNA in each knockdown was used for analyzing inclusion of the N1 exon by primer extension. For analysis of other alternative exons, two ug of total RNA and 100 ng of random hexamers were used to synthesize cDNA by SuperScript III (Life Technology) according to manufacturer's protocol. The cDNA was used in PCR (Phusion Hot Start II, Thermo Scientific) with ^{32}P 5' labeled gene specific primers as previously used by Xue et al 2009 (64). PCR reaction was treated with 0.5% v/v SDS and proteinase K (Roche) for 20 minutes at 37° C before resolved in 4% 7.5M urea-PAGE. The gel was heat dried and exposed to phosphorus screen (Molecular Dynamics) overnight.

siRNAs were synthesized by Bioland Scientific. Their targeted sequences (sense) are below:

siControl : ACGUGACACGUUCGGAGAA

siPTBP1 : CAGUUUACCUGUUUUUAAA (Linares et al 2015, in preparation)

siRAVER1 : UUGCCCAGCAGGUCCGACU (Chen et al 2013 (98))

siMBNL1 : AACACGGAAUGUAAAUUUGCA (Ho et al 2004 (99))

siDDX5/17 : GGCUAGAUGUGGAAGAUGU (Dardenne et al 2012 (100))

Western Analysis and Antibody

Western blots were performed on total protein from siRNA knockdown cell cultures (56 hours post transfection) lysed in RIPA buffer supplemented with protease inhibitors (Roche) and NucA nuclease (homemade). Lysates were diluted in 5X SDS loading buffer, heated for 5 min at 95°C. 40 ug of total protein was resolved in 8% polyacrylamide Laemmli SDS PAGE gels. Total protein was transferred onto a PVDF membrane (EMD Millipore). The membrane was blocked 5% skim milk in PBST and probed with primary antibodies overnight at 4°C. The membrane was then washed with PBST and incubated with 5% skim milk containing 1:5000 Cy3 or C5 conjugated secondary antibody. After washing, the membrane was directly scanned using Typhoon (GE). The followings were primary antibody used: mouse anti-PTBP1 (clone BB7, homemade), rabbit anti-RAVER1 (A303-939A, Bethyl), mouse anti-MBNL1 (clone 3E7, LSBio), rabbit anti-DDX5/17 (custom made from Genscript), mouse anti-U2AF65 (clone MC3, Sigma), mouse anti-GAPDH (clone 6C5, Life Technology) and mouse anti-U1-70K (clone CB7, homemade).

Purification of Recombinant DDX5 and DDX17

Human cDNAs expressing DDX5 and DDX17 were tagged with maltose-binding protein (MBP) at their N termini in order to make them solubilized in bacterial expression system and to use as an affinity tag. 6X histidine tag was also cloned into their C termini for affinity purification purpose. The constructs were inserted into pMal-c2 vector and transformed into BL21 Rosetta E. coli (Novagen). Three liters of cultures were induced with 0.5 mM IPTG while in a log phase (OD_{600} = 0.4-0.6) for three hours at 37°C. Cells were harvested and washed with PBS and proceeded for amylose affinity column (5-mL HPLC column, Pharmacia) purification as described by Jurica et al 2002 (101) with exceptions that 1) 10 column volumes of only AB1 buffer was used for washing the amylose column, 2) AB1 containing 40 mM maltose and 10 mM imidazole but no EDTA was use instead of ABE buffer. Fractions containing eluted protein were loaded into 5-mL HPLC Nickel NTA column (Pharmacia). The column was washed with 10 column volumes of a buffer containing 20 mM HEPES pH 7.9, 200 mM KCl, 50 mM imidazole. The bound proteins were eluted with the same buffer containing increasing gradient of imidazole from 50-500 mM (pH was monitored as high imidazole changes pH). Fractions containing proteins were analyzed on 8% SDS-PAGE and stained with Imperial stain (Life Technology). Fractions containing DDX5 or DDX17 were pooled and concentrated with Amicon Ultra-15 (50 KDa cutoff) before loading onto Superdex 200 HPLC pre-equilibrate with a buffer containing 20 mM HEPES pH 7.9, 80 mM potassium glutamate, 0.2 mM EDTA and 10 % glycerol. Eluted fractions containing proteins were analyzed again on 8% SDS-PAGE and stained with Imperial stain (Life

Technology). Only fractions containing monomeric form of DDX5 and DDX17 were pooled, added with an addition of glycerol to 20% and stored at -80 °C.

RESULTS

PTBP1 inhibits formation of the exon definition complex

Repression by PTBP1 requires multiple binding sites (102). These sites often reside within introns that flank a repressed exon and can be distal to an exon. An N1 exon of mouse Src pre-mRNA is a 23-nt alternative exon coding for a small peptide inserted into a SH3 domain of Src protein tyrosine kinase (103). Splicing of the N1 exon is repressed by PTBP1 which binds to CU-rich elements located downstream of the exon and immediately upstream of the 3' splice site overlapping a polypyrimidine tract (104,105). To begin characterizing the mechanism of PTBP1 repression when bound to flanking distal sites, we modified a PTBP1-repressed N1 exon by shifting upstream PTBP1-binding sites to 25-nt upstream of the branch site sequences (**Figure 2-1A**). We then inserted a strong polypyrimidine tract of Adenovirus Major Late pre-mRNA upstream of the 3' splice site AG in replacing the PTBP1-binding sites. In this way, we can rule out the possibility of PTBP1 competing with U2AF65 to bind to the polypyrimidine tract. The downstream PTBP1-binding sites were kept intact throughout the study. We analyzed ribonucleoprotein (RNP) complexes assembled on the wild-type N1 exon by incubating the RNA in a standard splicing reaction and resolving it in a native agarose gel. The analysis showed that there was only one complex formed on the N1 exon, i.e. an early exon complex, which is also called a heterogeneous complex (**Figure 2-1B** lane 1-4). Some reductions of early exon complex were observed due to

degradation. However, when the upstream PTBP1-binding sites were mutated (102), the early exon complex transformed to become a slower migrating exon complex as time progressed (**Figure 2-1B** cf. lane 1-4 to 5-8). We confirmed that this effect was due to PTBP1 by using PTBP1-depleted HeLa S3 nuclear extract to assemble RNP complexes on the wild-type N1 exon. As expected, PTBP1 depletion allowed for the formation of the slower migrating complex even on the wild-type N1 exon (**Figure 2-1C** cf. lane 3 to 4). Formation of the complex on the mutant RNA was not affected by PTBP1 depletion (**Figure 2-1C** cf. lane 5 to 6). This suggested that PTBP1 prevents formation of the slower migrating exon complex.

The slower migrating exon complex has a similar migration pattern to an EDC form across an exon, which contains U1 and U2 snRNPs (26,29). We verified that this complex contained both U snRNPs by degrading them from HeLa S3 nuclear extract using DNA-directed RNAseH cleavage. DNA oligo-mediated RNAse H degradation of either U1 or U2 snRNPs resulted in the inhibition of formation of the slower migrating complex, suggesting that the slower migrating complex was in fact the EDC (**Figure 2-1D** cf. lane 1-2 to 3-4 and 5-6). We renamed each exon complex according to their different regulatory states: **Repressed** (an early exon complex from the wild-type RNA), **De-repressed** (an early exon complex from the mutant RNA) and **EDC** (the slow migrating complex from the mutant RNA). Together, these data demonstrated that PTBP1 inhibits the formation of Exon Definition Complex.

PTBP1 repression inhibits reactivity of both 5' and 3' splice sites

To assess the consequence of PTBP1 binding that resulted in the blockage of EDC formation, we analyzed reactivity of both 5' and 3' splice sites of the N1 exon for splicing using trans-splicing assays. Trans-splicing measures the ability of a given exon to splice in *trans* to another exon on a separate RNA molecule. Thus, it is thought to represent splicing of multi-exon pre-mRNAs in higher eukaryotes, in which short exons are intervened with extremely long introns (26). We assembled in vitro exon complexes on the N1 exon containing either wild-type or mutant PTBP1-binding sites as previously mentioned. Reactivity of the N1 5' and 3' splice sites were measured by mixing the preformed exon complexes to another preformed exon complex containing either 3' or 5' splice sites, respectively (**Figure 2-2A**). We used both 5' and 3' exons of Adenovirus major late pre-mRNA for this purpose. After mixing, incubation was allowed to continue under a standard splicing reaction conditions for the trans-splicing reaction to occur. Spliced products were analyzed by primer extension using primers that annealed to the 3' end of the spliced RNA products. The results revealed that both 5' and 3' splice sites of the N1 exon were much less reactive for splicing in the wild-type compared to the mutant PTBP1-binding sites (**Figure 2-2B** and **2C** both cf. lane 3 to 4). This indicated that both splice sites are compromised in the PTBP1-bound exon complex. Together with previous data, we found that PTBP1 bound to distal sites deactivates 5' and 3' splice sites and hampers formation of the EDC.

PTBP1 blocks U2 snRNP—but not U2AF65 nor U1 snRNP—from associating with the exon

Exon Definition Complex is an ATP-dependent pre-spliceosomal A-like complex, consisting of two large snRNP particles U1 and U2 that base-pair to 5' splice site and branch point sequence, respectively (8), and U2AF heterodimer (U2AF65 and U2AF35) bound to the polypyrimidine tract and 3' splice site AG di-nucleotide. U2AF65 interacts with U2 snRNPs and is required for the recruitment to the branch point sequence (15,16). Obstruction of assembly of the EDC by PTBP1 seen in figure 2-1B and 1C could be a result of loss in binding of the U snRNPs to the exon. To determine base-pairing interactions of the U snRNPs to the N1 exon, we employed a double-stranded nucleic acid crosslinker 4′-aminomethyl trioxsalen (AMT). AMT is a tricyclic compound that intercalates into double-stranded regions of DNA or RNA. Upon a long wavelength UV irradiation (365 nm), the compound covalently crosslinks adjacent pyrimidine bases on opposite strands together (106). The crosslinked position can be precisely mapped using primer extension, which will stop at the base adduct site. We assembled exon complexes on both the wild-type and mutant RNAs, crosslinked them with AMT and UV and isolated RNA from the reaction for primer extension using a common primer that base-paired to a downstream intron of the N1 exon. Primer extension revealed stopped products at the 5' splice site. These products that was dependent on U1 snRNP, suggesting that they were 5' splice site-U1 snRNA base pairing (**Figure 2-3A** cf. lane 2 to 4 and 6 to 8). We detected no decrease in U1 snRNA base-pairing to the 5' splice site in the wild-type as compared to the mutant RNAs. In fact, we observed slightly more base-pairing interactions on the WT RNA compared to the mutant RNA (**Figure 2-**

3A cf. lane 2 to 6 and quantification on the right). Unfortunately, we could not detect branch point sequence-U2 snRNA crosslinked products (data not shown). This was likely due to low consensus match of the branch point sequence of the N1 exon. To get around this problem, we therefore tagged both RNAs with MS2 stem-loops for purification of exon complexes. Before assembly of the exon complexes, the RNAs were pre-bound to a fusion protein consisting of Bacteriophage MS2 coat protein and maltose binding protein (MS2-MBP). The complexes were assembled on the MS2-MBP–bound RNAs, resolved by ultracentrifugation on glycerol density gradients, then followed by affinity purification on amylose resin (**Figure 2-3B**). RNA from each complex was extracted, resolved in urea-PAGE (Polyacrylamide Gel Electrophoresis) and stained with SyberGold. As expected, we found that EDC contains both U1 and U2 snRNAs in a precise 1:1 stoichiometry relative to the N1 exon (**Figure 2-3C**). We found only U1 snRNA in both repressed and de-repressed exon complexes. Since the RNA with wild-type PTBP1 binding sites formed only the repressed complex, this suggested that it is the U2 snRNP that is prevented from binding to the RNA when PTBP1 is bound to the exon. Previous studies found that when PTBP1 binds inside an alternative exon, it prevents U2AF65, which is important for recruitment of U2 snRNP, from binding to an upstream polypyrimidine tract (5,107). However, unlike the previous reports, we found that the level of U2AF65 was unchanged regardless of whether or not PTBP1 was found inside the exon complexes (**Figure 2-3D**). Collectively, these data indicated that PTBP1 keeps the exon in the repressed complex, preventing U2 snRNP integration to form EDC. This is achieved even if U1 snRNP and U2AF are unaffected by PTBP1.

Quantitative proteomics reveals PTBP1 co-repressors, its target proteins and molecular assembly of EDC

Molecular interactions exist inside the EDC; these include, SR-SR proteins, U1 snRNP-SR proteins and U2AF65, U2AF65 and U2 snRNP, SR proteins and U2AF35, for example (11,13,15,31,47). The recruitment of U2 snRNP depends on U2AF65 (15,16). However, our finding that U2 snRNP failed to be recruited even if U2AF65 was bound to the exon intrigued us to realize that there must be an unappreciated network of interactions within the EDC and that PTBP1 may likely disrupt some of these interactions or their assembly pathway. In fact, a similar U2AF65-independent U2 snRNP blockage has been reported (108). In order to understand an assembly pathway of the EDC and how PTBP1 achieves the repression, we set up quantitative proteomics of the three purified exon complexes. To this end, we used SILAC coupled with mass spectrometry. Two populations of HeLa S3 were grown in cell culture media containing either **light** (^{12}C) or **heavy** (^{13}C) versions of Arginine and Lysine. Nuclear extracts were prepared from the two populations and used for assembling exon complexes (**Figure 2-4A)**. The wild-type RNA was assembled to the repressed complex in the heavy nuclear extract, while the mutant RNA was assembled to the de-repressed and EDC in the light nuclear extract. Exon complexes were two-step purified as above. Before being analyzed by mass spectrometry, two sample pairs were generated. We mixed 1:1 molar ratio of the heavy repressed complex to the light de-repressed complex, and also 1:1 molar ratio of the heavy repressed complex to the light EDC. This enabled us to use values in the heavy repressed complex to normalize the two sample pairs, so that we could generate a ratio of the relative abundance of a given

protein across three different exon complexes. We then used hierarchical clustering to group proteins according to their similar dynamic behavior across two transitions: from repressed state to de-repressed state, and from de-repressed state to exon definition. Unexpectedly, transitioning from de-repressed state to exon definition, we found that many proteins are depleted (**Figure 2-4C)**. This includes many hnRNP proteins such as hnRNPA1, hnRNPA2B1, hnRNPH, hnRNPK, PTBP1 and other abundant RNA binding proteins. Conversely, many U2 snRNP proteins are recruited during this transition, as expected. Nevertheless, we observed an unprecedented massive recruitment of many members of the SR protein family such as SRSF3, 6, 7, 11, 12, and including SR-related proteins, TRA2B, RBM39 and LUC7L2.

When comparing the repressed exon complex to the de-repressed complex, we found two proteins—RAVER1 and MBNL1—significantly decreased in the absence of PTBP1 (**Figure 2-4B)**. The two proteins were previously reported to be PTBP1 co-repressors in some exons (72-74). In contrast, there were many proteins that increased concurrent with low levels of PTBP1 in the de-repressed exon complex. Two proteins of interest are ATP-dependent RNA helicase DDX5 and its paralog DDX17, which may function as exon remodelers to promote exon inclusion. We found that association of these two proteins was also interesting as they had low levels in the PTBP1-repressed exon complex, high levels in the de-repressed exon complex, and then decreased levels again in the EDC, suggesting they are specific to an early exon complex that is not regulated by PTBP1.

We were able to confirm some of the mass spectrometry results by Western analysis (data not shown). Unfortunately, many proteins of interest cannot be probed

by Western due to low amounts of purified exon complexes and quality of antibodies. Thus, we performed another SILAC-MS using PTBP1- or mock-depleted light HeLa S3 nuclear extracts. Only the wild-type RNA was used to assemble all three exon complexes from the light nuclear extracts. Prior to MS analysis, each purified exon complex was spiked with mixed heavy exon complexes purified separately to serve as a normalization method. Between two experiments, ~78% of identified proteins overlapped. Loss of proteins in the second experiment was potentially due to low amount of the spike standard being used (data not shown). We found that fold changes of protein level were, in general, moderate in the second experiment. However, when we generated a matched heatmap, we noticed fairly correlated patterns of many proteins across all exon complexes as the previous analysis (**Figure 2-4B** and **C)**. Those patterns included lower amounts of many detected hnRNPs in the EDC and a higher abundance of U2 snRNP proteins and SR proteins in the EDC. In addition, this analysis confirmed relative abundance of PTBP1, RAVER1, MBNL1, DDX5 and DDX17 in all three exon complexes.

Knockdowns of RAVER1, MBNL1 and DDX5/17 confirm their role as PTBP1 co-repressors and as exon activators, respectively

The reciprocal abundance of DDX5/17 in the PTBP1-deficient de-repressed exon complex hinted toward an antagonistic relationship between them and also hinted at their role in promoting exon inclusion. Likewise, the enrichment of RAVER1 and MBNL1 in the PTBP1-repressed exon complex suggested their collaborative role in repressing the exon. To test these hypotheses in vivo, we performed siRNA knockdown

of those proteins in a HeLa cell line and measured splicing outcome of a three-exon minigene reporter, which contains wild-type N1 exon as an alternative exon (**Figure 2-5A)**. We used primer extension to assess inclusion of the N1 exon 56 hours after the siRNA knockdowns. Results showed that in control siRNA knockdown, the N1 exon was generally skipped, with only ~40% inclusion (**Figure 2-5B)**. Knocking down of PTBP1 significantly increased inclusion of the N1 exon to ~90%, while knocking down RAVER1 and MBNL1 increased inclusion level to ~55%. Consistent with MS analysis, these data support that in vivo the N1 exon is strongly repressed by PTBP1 and that RAVER1 and MBNL1 also contribute to the repression but to a lesser extent. In contrast, simultaneous knockdown of DDX5 and DDX17 using a common siRNA lowered the inclusion level to ~25% (**Figure 2-5B)**. This suggested that both proteins function to promote splicing of the N1 exon.

PTBP1 prevents DDX5/17-mediated efficient formation of EDC

We were particularly interested in the function of DDX5/17 in alternative splicing because the involvement of ATP-dependent RNA helicases in early spliceosomal assembly remains largely unknown. To determine the role of DDX5/17 in promoting exon inclusion observed in vivo, we assessed an effect of DDX5/17 in exon complex formation in vitro. HeLa S3 nuclear extracts were either DDX5/17- or mock-depleted by antibodies, then used for assembling exon complexes on the N1 substrates as above. Efficiency of immuno-depletion was modest, with 60%, 80% and 90% depletion of DDX5 (p68), DDX17 (p72) and DDX17 (p82), respectively (**Figure 2-6A** top,left**)**. Nevertheless, the depletion yielded some reduction of EDC assembly on the mutant

RNA. Importantly, adding back bacterially purified DDX17 to the DDX-deficient extract fully restored assembly of the EDC (**Figure 2-6A** cf. lane 4 to 5 and 5 to 6**)**. In contrast, the wild-type RNA could not form the EDC regardless of adding back the purified DDX17 (**Figure 2-6A** cf. lane 1-3 to 4-6**)**. We used another HeLa S3 nuclear extract, which could not support formation of the EDC, to test the effect of DDX5/17. The extract failed to promote EDC on both wild-type and mutant N1 RNAs (**Figure 2-6B** lane 1-2**)**. However, supplementing the extract with either bacterially purified DDX5 or DDX17 restored EDC assembly (**Figure 2-6B** cf. lane 2 to 4,6**)**. Notably, the effect of DDX5 and 17 was only on the mutant but not on the wild-type RNA (**Figure 2-6B** cf. lane 1 to 3,5**)**. Adding the other RNA-binding protein RbFox1 (a negative control) failed to promote assembly of EDC on neither the wild-type nor the mutant (**Figure 2-6B** lane 1-2**)**. Collectively, these data indicated that DDX5 and DDX17 promote efficient assembly of EDC only in the absence of PTBP1.

The EDC enhancing effect of DDX5 and DDX17 intrigued us to test whether this effect is specific to the N1 exon or general to other alternative exons. To this end we tested EDC formation on a synthetic 46-nt exon containing neutral sequences predicted to neither promote nor obstruct its own splicing (109). Native gel analysis clearly showed that both DDX5 and DDX17, not RbFox1, could promote assembly of the EDC even on this synthetic exon (**Figure 2-6C)**. To extend this finding, we tested splicing of other endogenous exons in vivo after knocking down DDX5/17 with the siRNA in HeLa cells (**Figure 2-6D)**. While some of these exons may or may not be negatively regulated by PTBP1, RAVER1 or MBNL1, they were all positively regulated by DDX5 and DDX17, as treatment with the siDDX5/17 resulted in increased exon skipping.

Together, these data indicated that DDX5 and DDX17 function to promote efficient assembly of EDC on many exons.

DISCUSSION

Alternative pre-mRNA splicing is an essential means to regulate gene expression in vertebrates. This process is largely regulated by *trans* factors that bind to pre-mRNAs to either promote or obstruct splicing of alternative exons. One such important and prevalent splicing factor is PTBP1 whose function is well known as a repressor. We found that when bound to distal sites both upstream and down stream from the N1 exon, PTBP1 blocks assembly of the Exon Definition Complex (**Figure 1**), which results in inactivation of both 5' and 3' splice sites (**Figure 2**). The inactivation of the 5' splice site may result from PTBP1 competitively interacting with a stem-loop 4 of U1 snRNA, which is required for the cross-intron interaction between U1 and U2 snRNPs (23,110). The inactivation of the 3' splice site is potentially due to blockage of U2 snRNP recruitment to the exon (**Figure 2-3C**). Surprisingly, unlike U2AF65-dependent PTBP1 repression reported previously (5,7,107), we found that this type of repression did not use the same mechanism to block U2 snRNP binding to the branch point sequence. This blockage was achieved even if U2AF65 remained associated to the polypyrimidine tract (**Figure 2-3D**). It is likely that there are intricate interactions within the early exon complex required for U2 snRNP recruitment, besides its interaction with U2AF65. One important type of interaction is of RS-RS domain interactions of SR proteins, which have been demonstrated to be essential for splicing by mediating communication between 5' and 3' splice sites across both exon and intron (11,13,44,46,69,111). In fact, our

quantitative MS revealed for the first time that an unexpected substantial amount of SR and SR-related proteins were not recruited until the assembly of EDC, even though we detected their earlier binding to the exon (**Figure 2-4B and C**). It is plausible to envisage that a fraction of SR proteins bind early to the exon in the early exon complex to help anchor U1 snRNP and U2AF (11,13); however, to efficiently recruit U2 snRNP and define the exon, more elaborate RS-RS domain interactions are required to bridge the communication between 5' and 3' splice sites and to form a functional EDC. It is also possible that those early SR proteins provide binding patches for other factors, yet to be identified, which are constitutive players of the EDC. One potential candidate is TRA2B. Regardless, these notions open possibilities for regulation of alternative splicing, as disrupting these interactions may result in exon exclusion. Some pieces of evidence were reported in certain exons that the presence of hnRNPA1 and PTBP1 leads to decrease of binding of SR proteins (70,112).

The SILAC-MS analyses provided us an insight into fate of hnRNP proteins during splicing. hnRNP proteins are an abundant class of proteins that associate very early to a pre-mRNA as it emerges from a transcribing RNA polymerase II (79,113). Their association fulfills either regulatory purposes or structural support as the pre-mRNA is packaged for subsequent processing. However, this presents an obstacle for assembling the EDC because these hnRNPs may hinder SR proteins from binding or may sterically clash with the massive U2 snRNP. At the branch point sequence, UAP56 (DDX39B) protein is believed to promote U2 snRNP recruitment by removing nearby proteins (18), but it has not been detected in all exon complexes, including ours (8,30,114). Nevertheless, it is very likely that other compositional remodeling events

occur within an exon and neighboring intronic sequences to remove hnRNP proteins. In fact, we detected a widespread decrease of most, if not all, hnRNP proteins and abundant RNA-binding protein such as YBX1 and YBX3 during the transition to the EDC (**Figure 2-4B and C**). This finding provides us with a better understanding of dynamics of hnRNP proteins as the exon is more committed into a splicing pathway.

The answer to how PTBP1 sequesters the exon in the early complex partly came to light when we compared proteomic data between the Repressed and De-repressed exon complexes. Mutating only one tract of PTBP1 binding site significantly changed the proteomic profile of the Repressed exon complex (**Figure 2-4B**). These changes were unlikely to be explained simply only as mutant-specific protein binders because when we repeated the MS analyses with the three exon complexes differently assembled using only WT RNA and mock- or PTBP1-depleted nuclear extracts, the results were fairly similar, even though degrees of fold changes were variable in many proteins (**Figure 2-4C**). Therefore, this suggested that PTBP1 could modulate exon proteomics. Of these changes, we detected a significant decrease of RAVER1 and MBNL1 as PTBP1 departs from the Repressed exon complex. Consistent with earlier studies suggesting their roles as PTBP1 co-repressors (72-74), RAVER1 and MBNL1 contributed to repressing the N1 exon in vivo (**Figure 2-5B**). Interestingly, we noticed an inverse relationship between PTBP1 and ATP-dependent RNA helicases DDX5 and DDX 17 among others (**Figure 2-5B and C**). Knockdown of DDX5/17 indicated that they antagonize PTBP1 by promoting N1 inclusion (**Figure 2-5B**). The two helicase paralogs are particularly attractive to us because 1) they, especially DDX5, have been found in all exon complexes reported so far (1,8,30,114), and they have found to be co-

purified with U1 and U2 snRNPs (115,116), 2) their association to an early exon complex (in non-PTBP1-regulated exon complex) implies to us an unappreciated early ATP-dependent remodeling on pre-mRNA, before the first known ATP hydrolysis used by U2 snRNP to base-pair to branch point sequence (20). In fact, we observed extensive compositional remodeling events of the early exon complex transitioning to the EDC—those events are the global eviction of hnRNP proteins and abundant RNA-binding proteins, the substantial recruitment of SR and SR-related proteins, and the recruitment of U2 snRNPs. It may be true that during this transition, RNA-bound hnRNP proteins need to be removed from the RNA in order to make room for the binding of SR proteins and U2 snRNP. Yet, the extent of these events that results from DDX5 and DDX17 activities needs to be determined.

We hypothesize that DDX5 and DDX1 possess a general role in promoting formation of the EDC during an early stage of splicing, besides their roles in alternative splicing of specific exons (115,117,118). In fact, both proteins can facilitate assembly of the EDC even on a synthetic exon (**Figure 2-6C**) as well as promote splicing of other tested alternative exons (**Figure 2-6D**). Recent studies by Dardenne et al. using exon arrays to assess splicing changes after knocking down DDX5/17 revealed that large numbers of alternative exons are normally activated by both proteins (100,119). Even though they also found that many alternative exons are repressed by the proteins, it is possible that these are indirect effects of DDX5 and DDX17, or that exon definition process of flanking constitutive exons are more dependent on the two proteins.

As exon definition is the major pathway for recognizing splice sites in vertebrates, it is becoming clear that an ultimate goal of splicing repression is to disrupt this process,

which many times involve directly intervening with binding of SR proteins. Still, it is likely that repressors, including hnRNP proteins, employ other mechanisms to achieve the same goal. Given divergent sequences of splice sites, exonic and intronic sequences in vertebrates, every exon is expected to have a different splice site strength, sets of SR proteins, hnRNP proteins and other RNA binding factors. A sum of these positive and negative factors will determine splicing outcomes. Proteomic approach on various model alternative exons under normal and repressive conditions will lead to an understanding of the steps during assembly of the EDC, revealing factors necessary for the process or factors needed for repression. We showed here how such an approach could be employed to study complex assembly of the EDC and repression mechanism by PTBP1. With knowledge from various models, we may be able to see a common theme of how repressors accomplish splicing repression.

In conclusion, we believe that DDX5 and DDX17 play an important role in splicing by remodeling protein compositions during the transition from the early exon complex to the EDC. By blocking the association of DDX5 and DDX17 to the early exon complex, PTBP1 can sequester the exon in the repressed state. We suspect that there are unknown properties still to be determined in the PTBP1-repressed exon complex that veer DDX5 and DDX17 away from the exon.

FIGURES

Figure 2-1 PTBP1 inhibits formation of the Exon Definition Complex

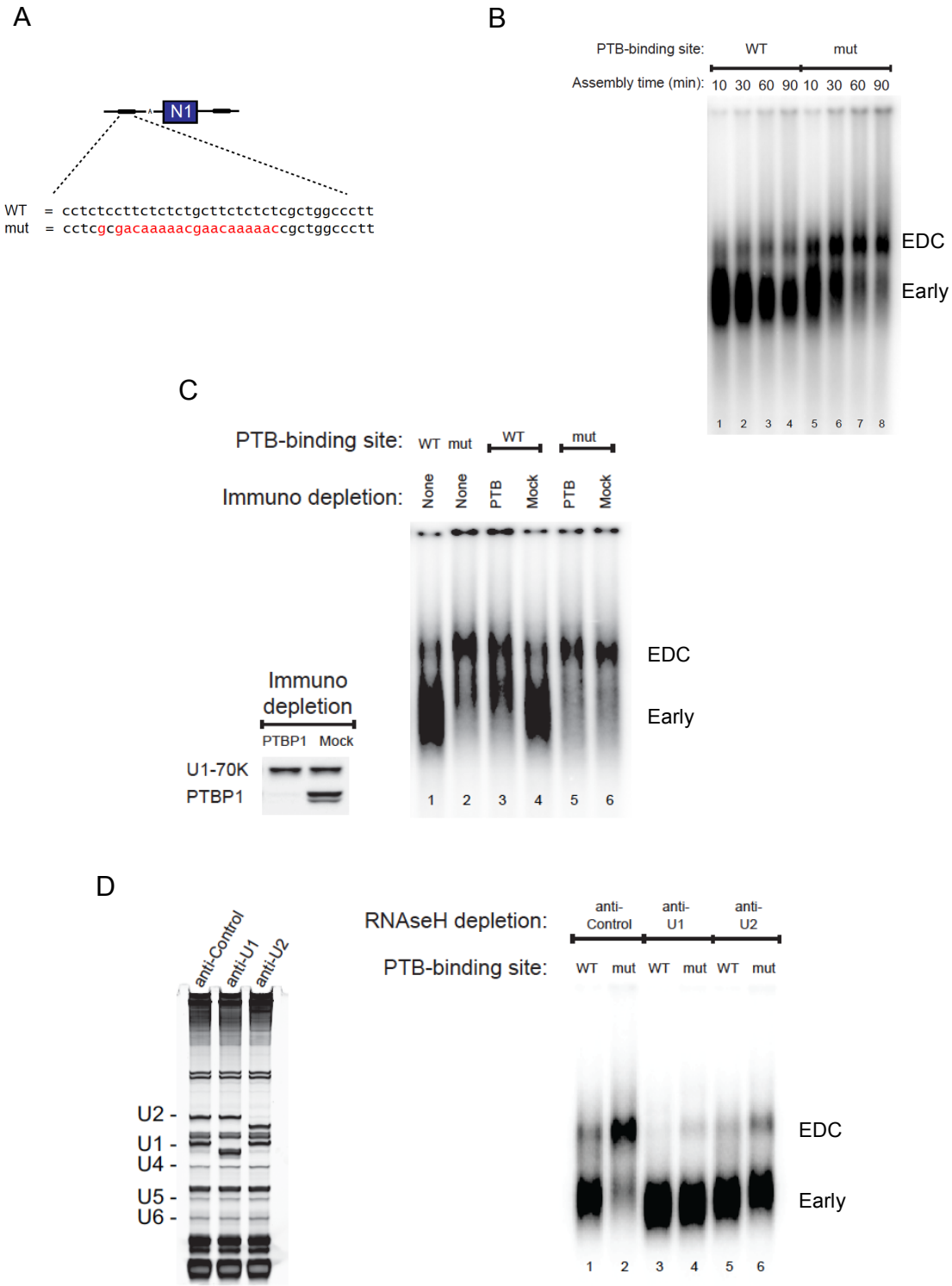


Figure 2-2 PTBP1 repression inhibits reactivity of both 5’ and 3’ splice sites

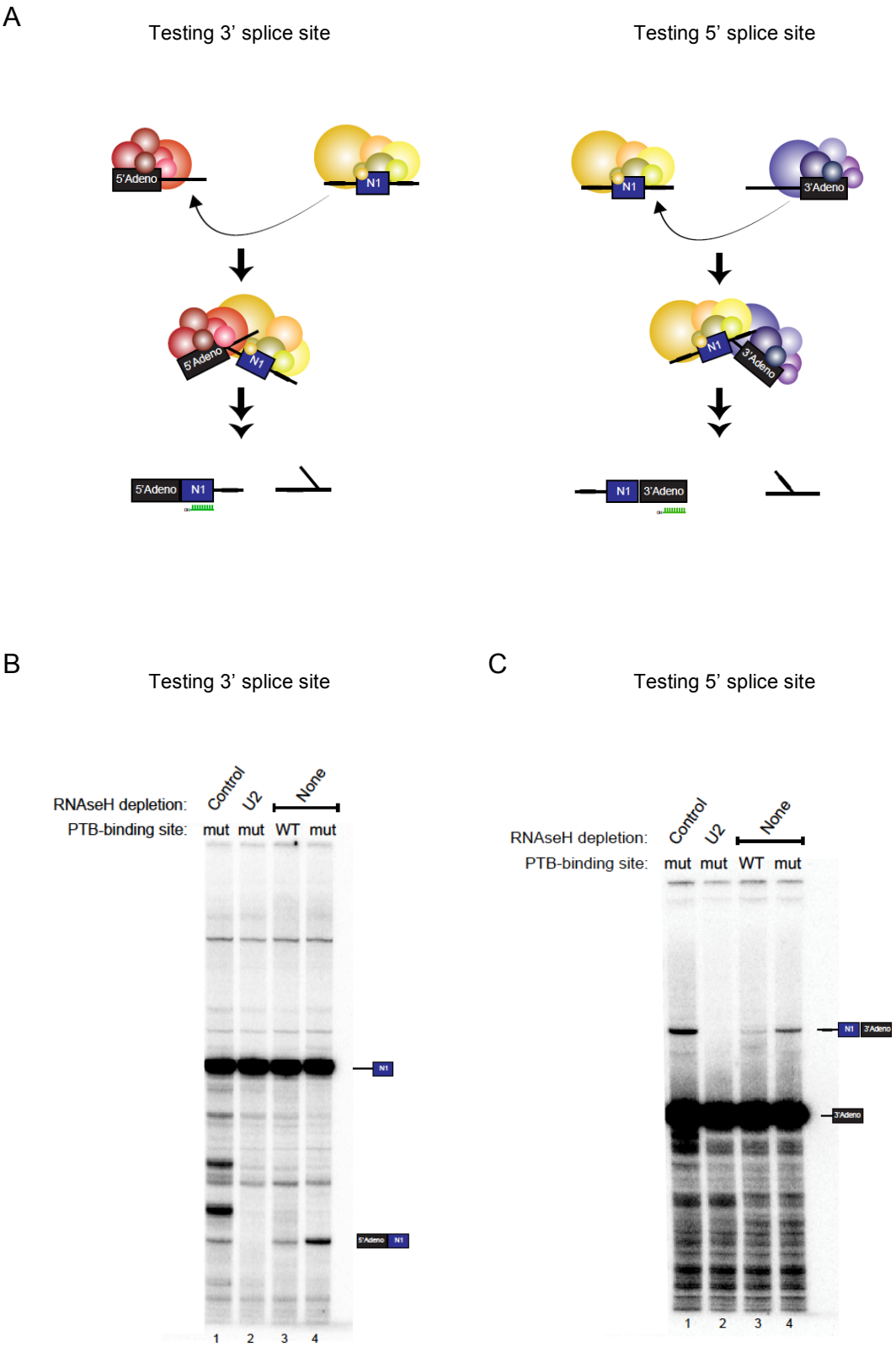
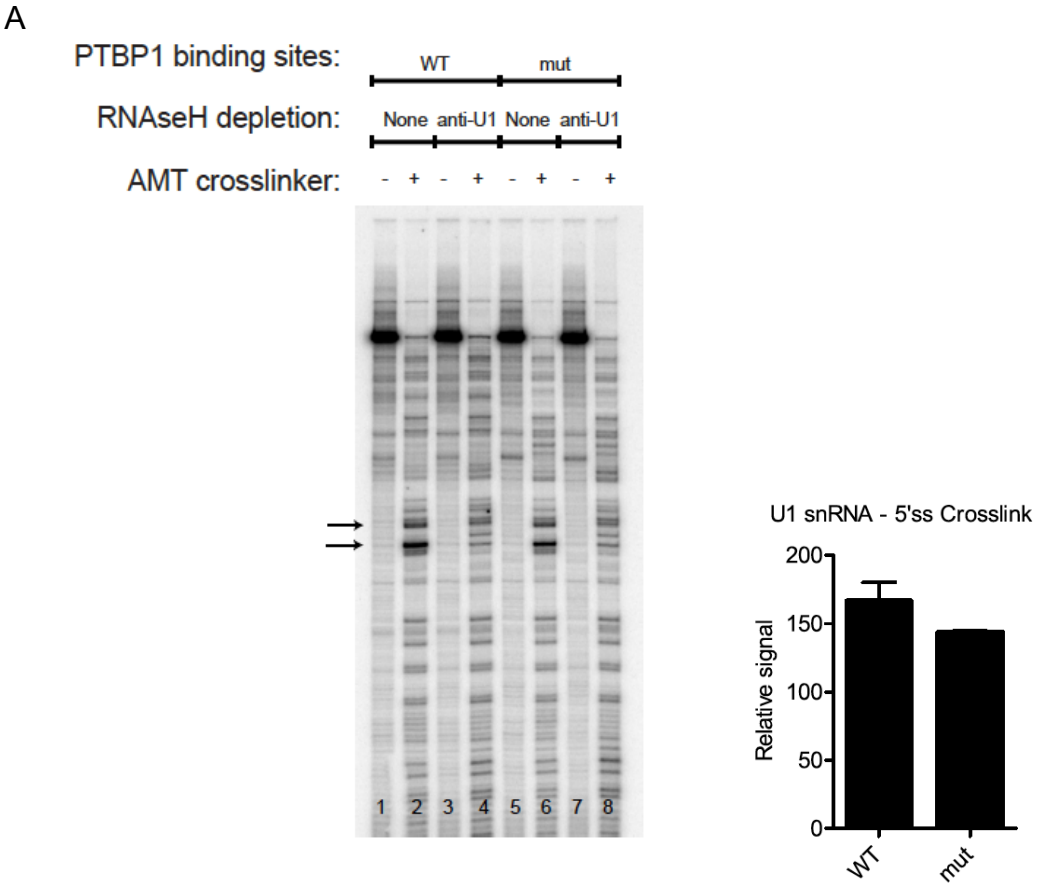
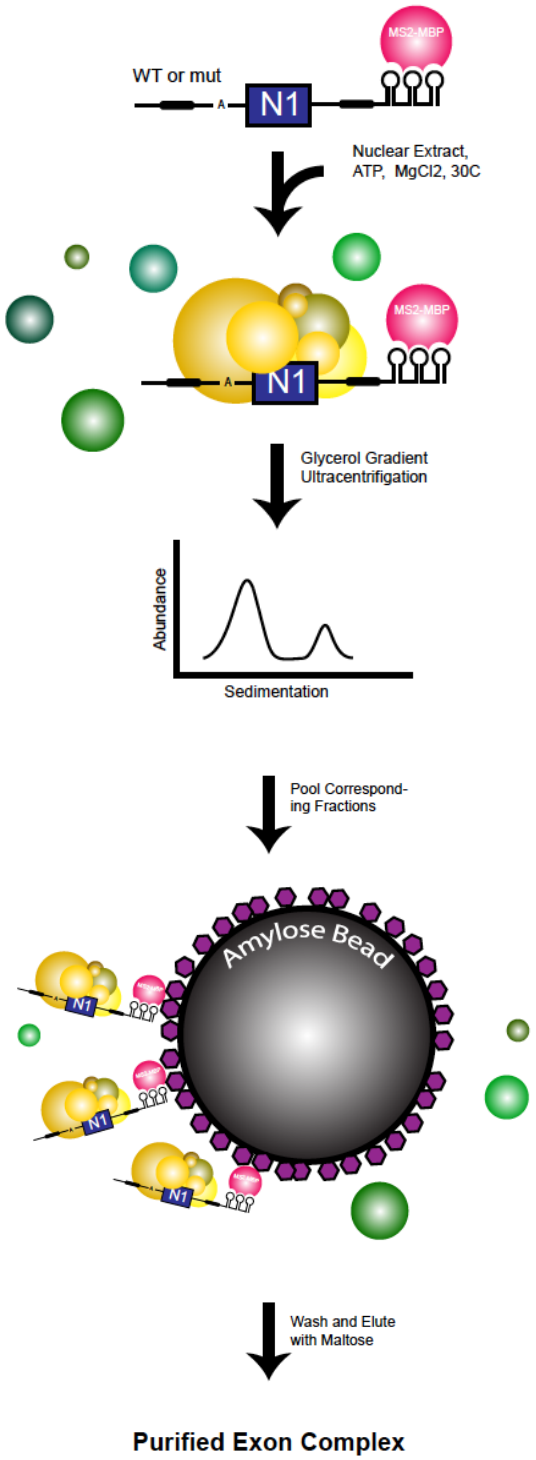


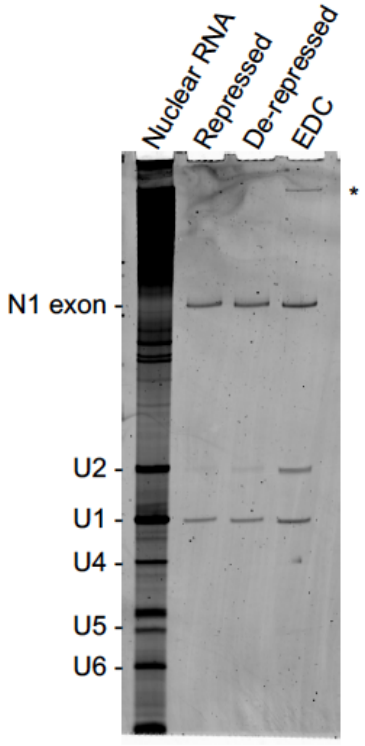
Figure 2-3 PTBP1 blocks U2 snRNP—but not U2AF65 nor U1 snRNP—from associating with the exon

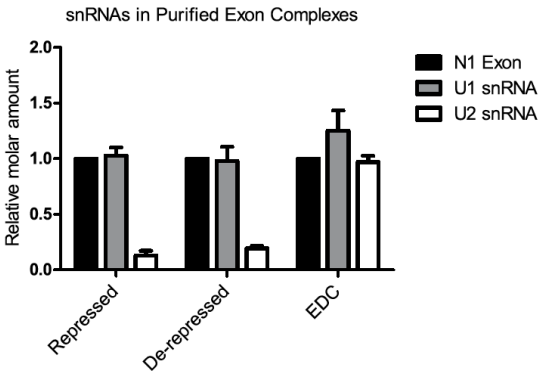


B


WT or mut
MS2-MBP
N1
Nuclear Extract,
ATP, MgCl2, 30C
Glycerol Gradient
Ultracentrifigation
Abundance
Sedimentation
Pool Correspond-
ing Fractions
Amylose Bead
Wash and Elute
with Maltose
Purified Exon Complex

C


Nuclear RNA
Repressed
De-repressed
EDC
N1 exon
U2
U1
U4
U5
U6


snRNAs in Purified Exon Complexes
Relative molar amount
2.0
1.5
1.0
0.5
0.0
N1 Exon
U1 snRNA
U2 snRNA
Repressed
De-repressed
EDC

D

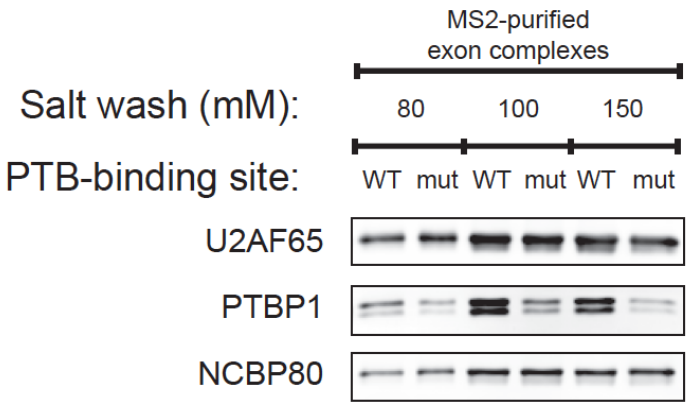
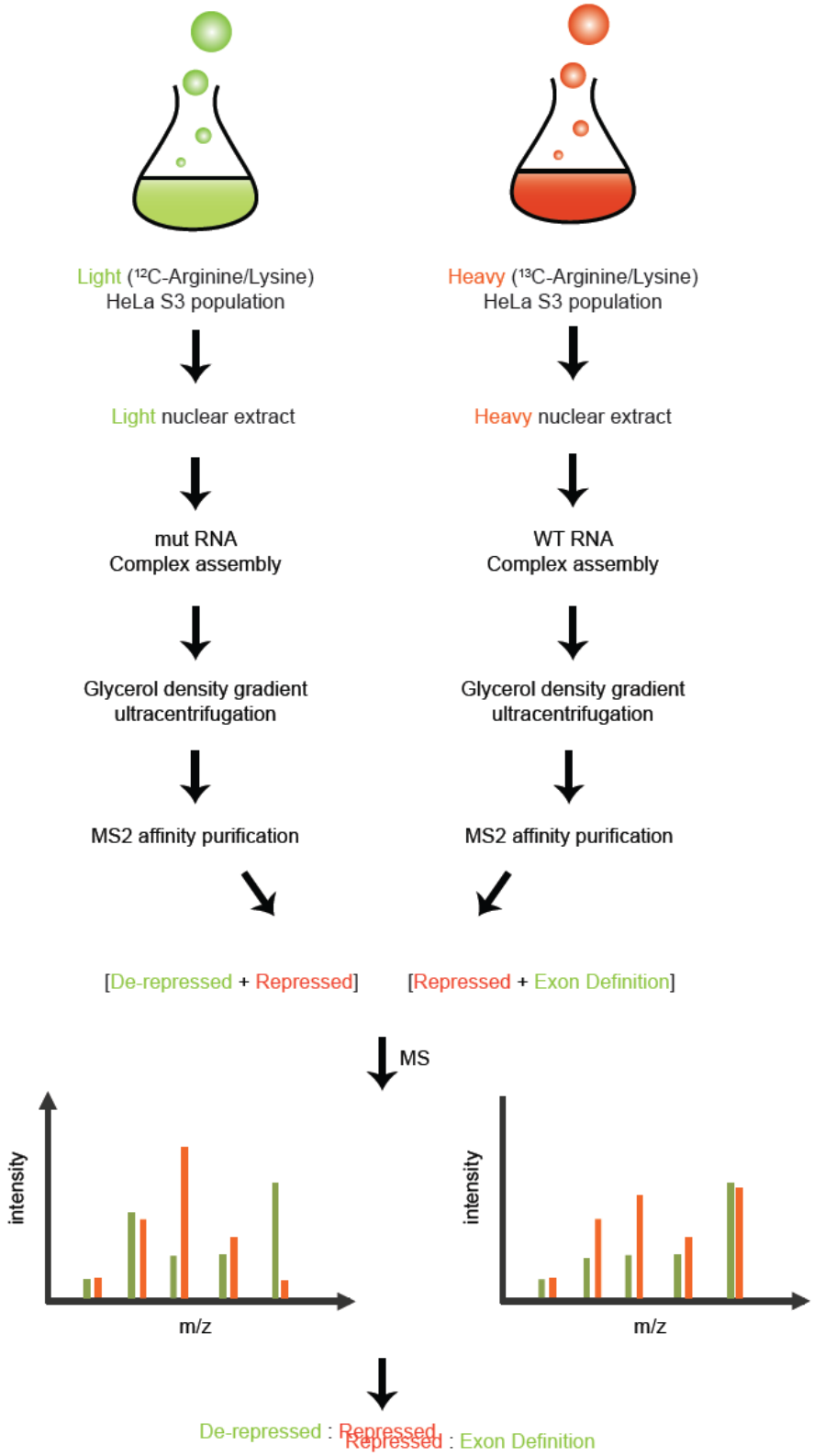
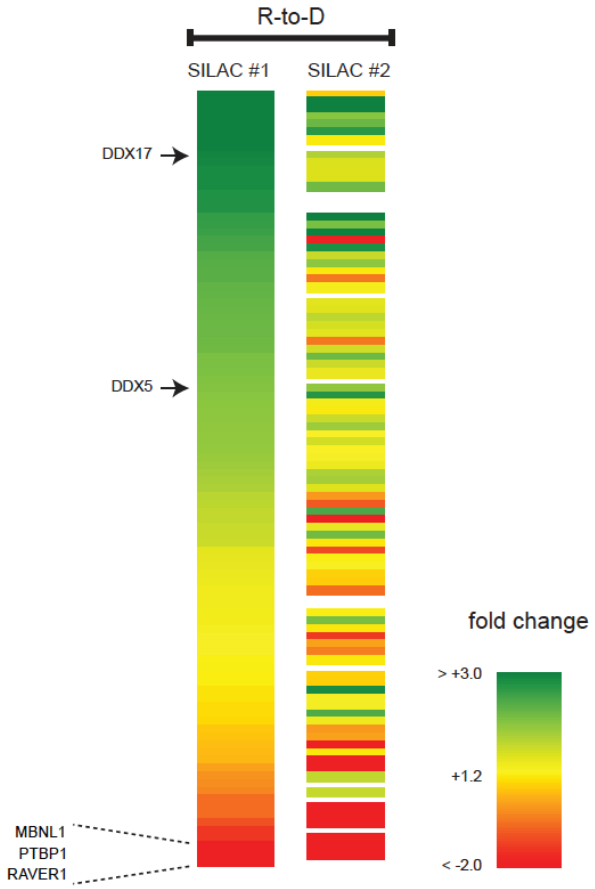

MS2-purified
exon complexes
Salt wash (mM):
80
100
150
PTB-binding site:
WT mut WT mut WT mut
U2AF65
PTBP1
NCBP80

Figure 2-4 Quantitative proteomics reveals PTBP1 co-repressors, its target proteins and molecular assembly of EDC

A



B


R-to-D
SILAC #1
SILAC #2
DDX17
DDX5
MBNL1
PTBP1
RAVER1
fold change
> +3.0
+1.2
< -2.0

C

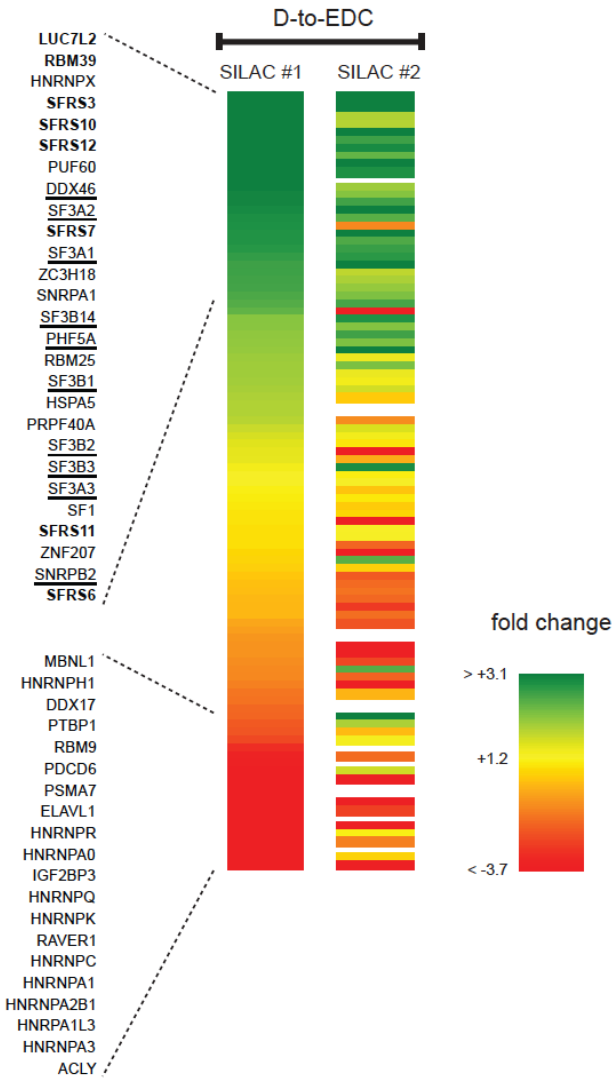

D-to-EDC
SILAC #1
SILAC #2
LUC7L2
RBM39
HNRNPX
SFRS3
SFRS10
SFRS12
PUF60
DDX46
SF3A2
SFRS7
SF3A1
ZC3H18
SNRPA1
SF3B14
PHF5A
RBM25
SF3B1
HSPA5
PRPF40A
SF3B2
SF3B3
SF3A3
SF1
SFRS11
ZNF207
SNRPB2
SFRS6
MBNL1
HNRNPH1
DDX17
PTBP1
RBM9
PDCD6
PSMA7
ELAVL1
HNRNPR
HNRNPA0
IGF2BP3
HNRNPQ
HNRNPK
RAVER1
HNRNPC
HNRNPA1
HNRNPA2B1
HNRPA1L3
HNRNPA3
ACLY
fold change
> +3.1
+1.2
< -3.7

Figure 2-5 Knocking down of RAVER1/MBNL1 and DDX5/17 supports their role as PTBP1 co-repressors and as exon activators, respectively

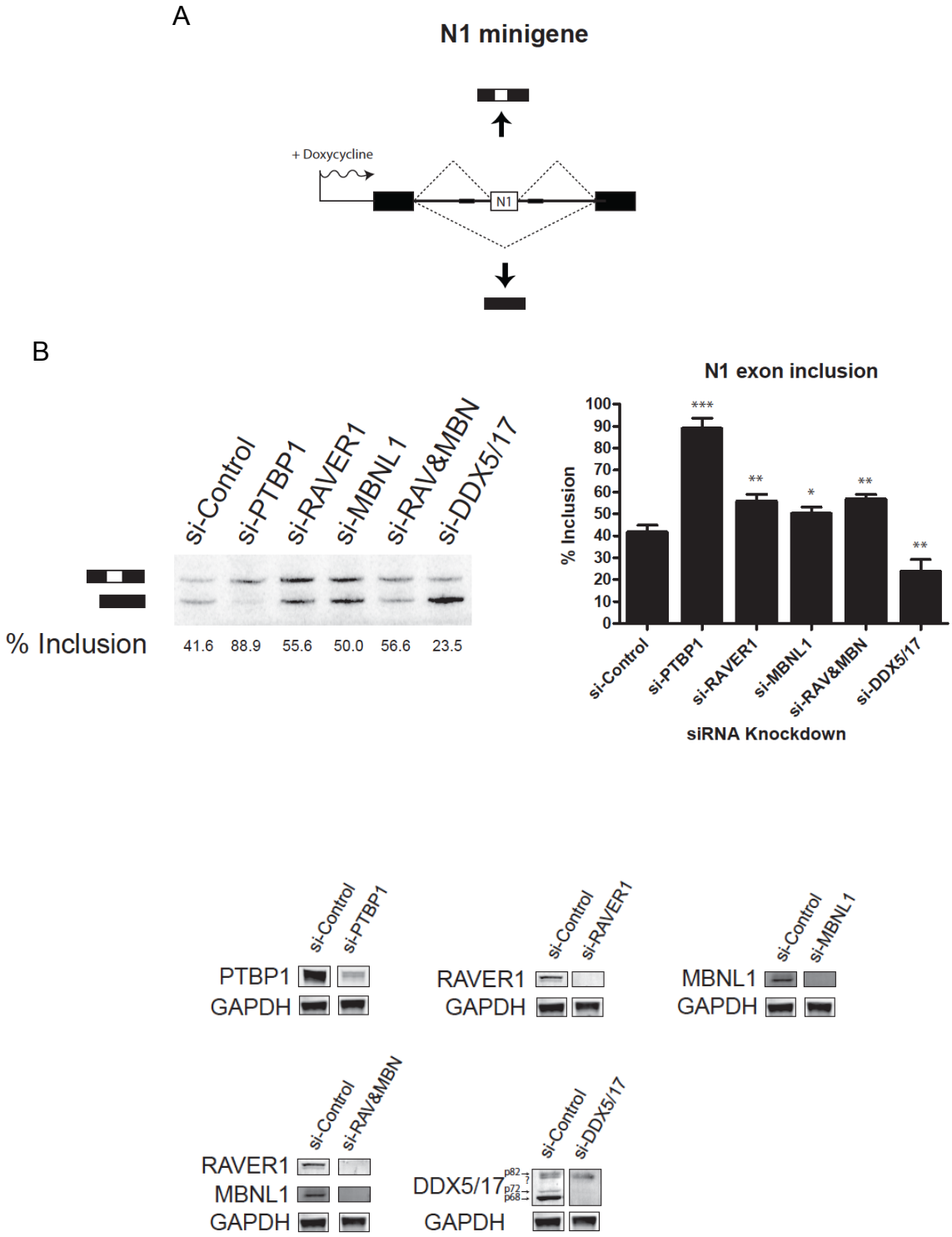
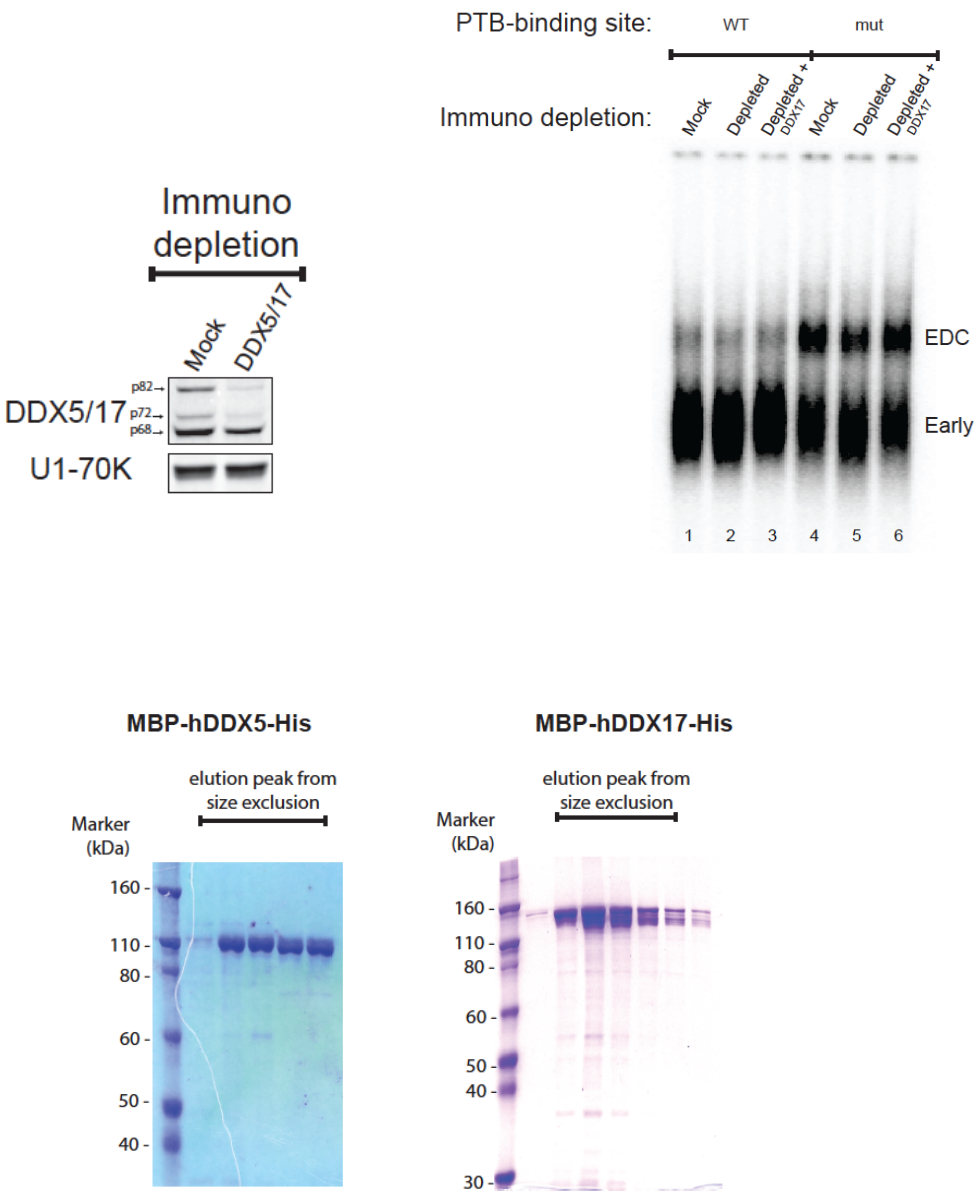
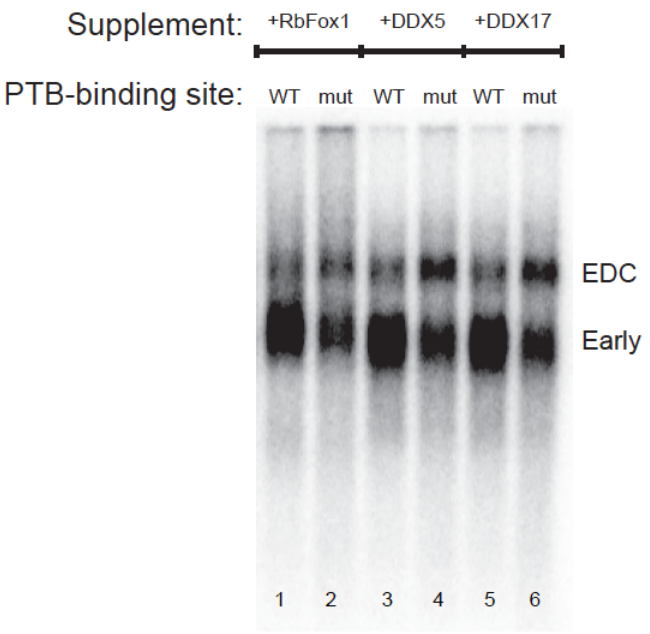


Figure 2-6 PTBP1 prevents DDX5/17-mediated efficient formation of EDC

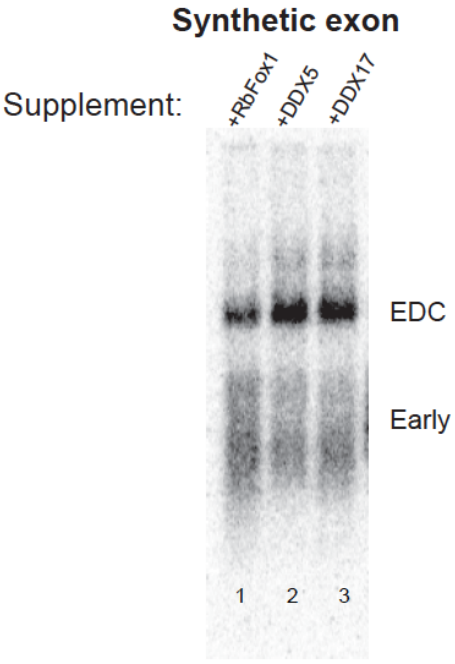
A



B


Supplement:
+RbFox1
+DDX5
+DDX17
PTB-binding site:
WT
mut
WT
mut
WT
mut
EDC
Early
1
2
3
4
5
6

C


Synthetic exon
Supplement:
+RbFox1
+DDX5
+DDX17
EDC
Early
1
2
3

D

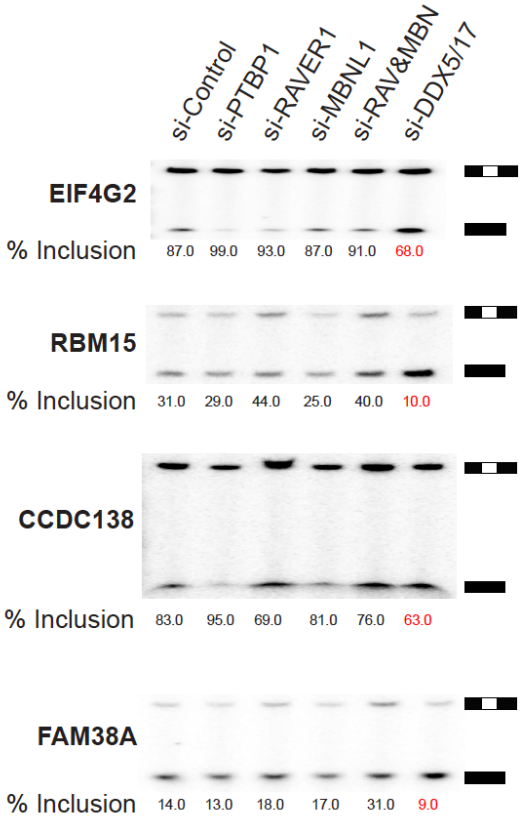

si-Control
si-PTBP1
si-RAVER1
si-MBNL1
si-RAV&MBN
si-DDX5/17
EIF4G2
% Inclusion
87.0
99.0
93.0
87.0
91.0
68.0
RBM15
% Inclusion
31.0
29.0
44.0
25.0
40.0
10.0
CCDC138
% Inclusion
83.0
95.0
69.0
81.0
76.0
63.0
FAM38A
% Inclusion
14.0
13.0
18.0
17.0
31.0
9.0

FIGURE LEGENDS

Figure 2-1 PTBP1 inhibits formation of the Exon Definition Complex. (**A**) A schematic representation of the N1 exon of Src pre-mRNA used in this study. An original upstream PTBP1-binding sites were moved 25-nt upstream of the branch point sequence. A 3' splice site of an Adenovirus major late pre-mRNA (not shown) was added in place of the PTBP1-binding sites to allow U2AF65 binding without competition with PTBP1. Downstream PTBP1-binding sites were kept unchanged. Mutations of the upstream PTBP1-binding sites were previously used in Sharma et al. 2008 and are highlighted in red. Thick lines represent up- and downstream PTBP1-binding sites. (**B**) Native agarose gel electrophoresis of exon complexes assembled on WT and mut N1 exons. ^{32}P uniformly labeled WT and mut RNAs were incubated with HeLa S3 nuclear extract under a standard splicing reaction for the indicated times. The reactions were then analyzed on a 2.5% native agarose gel. (**C**) HeLa S3 nuclear extracts were either mock-depleted or PTBP1-depleted using a monoclonal antibody (BB7). A Western blot on the left shows efficiency of the immune-depletion. U1-70K was used as a loading control for the Western blot. Both HeLa S3 nuclear extracts were then used for assembling exon complexes on ^{32}P uniformly labeled WT and mut N1 exons under a standard splicing reaction for 30 minutes and analyzed on a 2.5% native agarose gel. (**D**) HeLa S3 nuclear extracts were treated with RNAse H and DNA oligos complementary either to GAPDH mRNA (control), U1 snRNA or U2 snRNA in order to degrade the corresponding RNAs. On the left, total RNA from each treated HeLa S3 nuclear extract was analyzed in 8% urea-PAGE and stained with SyberGold to assess efficiency of degradation of U snRNAs. All three HeLa S3 nuclear extracts were then

used for assembling exon complexes on ^{32}P uniformly labeled WT and mut N1 exons under a standard splicing reaction for 30 minutes and analyzed on a 2.5% native agarose gel.

Figure 2-2 PTBP1 repression inhibits reactivity of both 5' and 3' splice sites. (**A**) A schematic representations of trans-splicing reactions. Exon complexes were assembling on either WT or mut N1 exons under a standard splicing reaction for 20 minutes. To test reactivity of 3' splice site of the N1 exon (left), the reactions were mixed with 5' exon complex of Adenovirus major late pre-mRNA that had been assembled separately under the same condition. After mixing, the combined reaction were supplemented with another dose of ATP and Creatine phosphate, and further incubated at 30° C for 3 hours. Total RNA was extracted from the reactions and used for primer extension analysis with 5' ^{32}P labeled primer, which annealed to 3' end of the spliced product. The primer extension products were resolved in 8% urea-PAGE. Spliced and unspliced products were indicated on a gel (**B**). The same experiment was repeated to test reactivity of 5' splice site of the N1 exon; however, 3' exon complex of Adenovirus major late pre-mRNA was use instead. Spliced and unspliced products of 5' splice site test were indicated on a gel (**C**). Identity of the spliced products was confirmed by size and their requirement of splicing after U2 snRNP was degraded.

Figure 2-3 PTBP1 blocks U2 snRNP—but not U2AF65 nor U1 snRNP—from associating with the exon (**A**) U1 snRNA crosslinking to 5' splice site (5'ss). HeLa S3 nuclear extract was treated with RNAse H and either water ("None" treatment) or DNA

oligo complementary U1 snRNA in order to degrade U1 snRNP, according to a protocol used by Merendino et al. 1999. The treated nuclear extracts were then used for assembling exon complexes on both WT and mut N1 exons under a standard splicing reaction for 30 minutes. A crosslinker 4′-aminomethyl trioxsalen (AMT) or water was then added to the reactions. Crosslinking was performed on ice and initiated by irritating 366 nm UV wavelength for 10 minutes, according to a protocol used by Sharma et al. 2008. Total RNA in the reaction was extracted and used for primer extension analysis with a primer annealing to an intron downstream of the N1 exon. Primer extension products were resolved in 8% urea-PAGE. U1 snRNA – 5' splice site crosslink created base adducts in which a Reverse Transcriptase can not extend the primer across. These stopped products were precisely mapped to the N1 5' spliced site and were dependent on the U1 snRNP. Arrows indicate the two crosslinked products detected. Quantification of the major stopped product (bottom band) is shown in the graph on the right (n = 2). (**B**) A schematic presentation of purification of exon complexes. Both WT and mut N1 exons were tagged with 3X MS2 stem-loops and were uniformly labeled with ^{32}P. The RNAs were pre-incubated with a fusion protein of MS2 bacteriophage coat protein and Maltose-binding protein (MS2-MBP) before assembling exon complexes under a standard splicing reaction for 30 minutes. Exon complexes were separated by glycerol density gradient ultracentrifugation. The gradient was fractionated, and peaks were determined by scintillation counting. Peak fractions containing exon complexes were then affinity purified on amylose beads. After washing, bound exon complexes were eluted with maltose for further analyses. (**C**) RNA was extracted from each exon complex and resolved in 8% urea-PAGE. Total

nuclear RNA was used as a marker for U snRNAs. The gel was stained with SyberGold. Intensity of each RNA species was normalized by its size (nt.) in order to calculate its relative molar ratio. The average of two experiments is shown in the graph below. " * " marks uncharacterized band. (**D**) To determine relative amounts of U2AF65 bound to WT and mut RNAs, ^{32}P uniformly labeled MS2 stem-loop-tagged WT and mut N1 exons were pre-incubated with MS2-MBP fusion protein and then used for assembling exon complexes under a standard splicing reaction for 30 minutes. The reactions were then diluted 3-fold and affinity purified on amylose beads. The beads were split into three groups, and then washed with a buffer containing different concentration of potassium glutamate—80, 100 or 150 mM. After washing, exon complexes were eluted with maltose. Proteins were resolved in SDS-PAGE and analyzed for PTBP1, U2AF65 and NCBP80 (loading control) by Western blotting.

Figure 2-4 Quantitative proteomics reveals PTBP1 co-repressors, its target proteins and molecular assembly of EDC. (**A**) A schematic representation of SILAC-MS experiment. ^{32}P uniformly labeled WT and mut N1 exons were used for assembling exon complexes with Heavy and Light HeLa S3 nuclear extracts, respectively. The exon complexes were purified similar to Figure 2-3B. After purification, molar concentration was determined by scintillation counting. One molar amount of the Heavy Repressed exon complex was mixed to one molar amount of either Light De-repressed or Light EDC. Each pair of samples was subjected to MS analyses. For a given protein, two SILAC ratios (Light intensity : Heavy intensity) were obtained—one from each sample. To generate relative abundance of a protein across three different exon

complexes, the Heavy intensity (from Repressed exon complex) was used for normalizing the two SILAC ratios (represented by overlapping two 'Repressed' words at the bottom of the figure). Only proteins in which their abundance could be calculated in all three exon complexes were further analyzed. Their fold changes were calculated during the transition from Repressed to De-repressed (R-to-D) and from De-repressed to EDC (D-to-EDC). (**B**,**C**) Log_2(fold change) was calculated for all proteins and used to generate heatmaps. Proteins with decreasing abundance are in red shades, while those with increasing abundance are in green shades; yellow shades are those in between. SILAC-MS was repeated in a different experiment in which all three exon complexes were assembled only from WT RNA with Light nuclear extract that was either mock- or PTBP1-depleted. Each of the three complexes was purified as above and spiked in with MS2-affinity purified Heavy mixed exon complexes, assembled from WT and mut RNAs. The Heavy mixed exon complex served to normalize SILAC ratios across the three exon complexes. Log_2(fold change) of the same set of proteins from the second SILAC-MS were obtained and arranged according to proteins from the first SILAC-MS. Their heat maps were then generated.

Figure 2-5 Knocking down of RAVER1/MBNL1 and DDX5/17 supports their role as PTBP1 co-repressors and as exon activators, respectively. (**A**) A schematic representation of the N1 minigene reporter used to assay effects of knocking down splicing factors in HeLa cells. N1 exon containing WT PTBP1-binding sites was inserted into an intronic region between constitutive exon 1 and 2 of human beta globin gene (black boxes). The entire construct was inserted into a Flp-In recombination site

of Flp-In T-REx HeLa cells (Life Technology). The reporter was under the control of Tet repressor. Transcription was induced by adding Doxycycline. Thick lines represent up- and downstream PTBP1-binding sites. (**B**) HeLa cells were transfected with corresponding siRNAs for 48 hours before inducing expression of the reporter N1 by adding Doxycycline. Eight hours after induction, total RNA was harvested from cells and used for primer extension assays with a ^{32}P 5' labeled DNA primer annealing to the second exon of the beta globin gene. Products of primer extension were resolved in 4% urea-PAGE. Included and skipped products were indicated on a gel. Intensity of each different spliced product was determined, and then used for calculating percent inclusion for each knockdown. Data from three experiments are shown in the graph on the right. The graph shows the averages and standard deviation (S.D.). p-values were calculated compared to si-Control. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. Efficiency of knocking down each protein was analyzed by Western blotting; GAPDH served as a loading control. Note that the antibody for DDX5/17 recognized three polypeptides—p68 (DDX5), p72 (DDX17), p82 (DDX17). " ? " marks an unidentified protein from monolayer HeLa that cross-reacted with the antibody.

Figure 2-6 PTBP1 prevents DDX5/17-mediated efficient formation of EDC. (A) HeLa S3 nuclear extract was either mock- or DDX5/17-depleted with an antibody. Efficiency of the depletion was analyzed by Western blotting shown on the top left. Each polypeptide was indicated p68 (DDX5), p72 (DDX17), p82 (DDX17). U1-70K was used as a loading control. The treated HeLa S3 nuclear extracts were then used for assembling exon complexes on ^{32}P uniformly labeled WT and mut N1 exons under a

standard splicing reaction at 30°C for 15 minutes. In lane 3 and 6, bacterially purified human DDX17 was added back to the depleted nuclear extract prior to the splicing reaction. The complexes were resolved in 2.5% native agarose gel. Different complexes were indicated. On the bottom, protein-stained SDS-PAGE gels of bacterially purified human DDX5 and DDX17 used in this study were shown. The proteins were fused to Maltose-binding protein (MBP) and 6X His tag on their N and C termini, respectively, and purified on Amylose column followed by Nickel column before size exclusion column. (**B, C**) Splicing-deficient HeLa S3 nuclear extract was supplemented with either bacterially purified RbFox1, DDX5, DDX17 prior to assembling exon complexes on ^{32}P uniformly labeled WT and mut N1 exons (**B**) or a synthetic exon (**C**) under a standard splicing reaction at 30°C for 30 minutes. The complexes were resolved in 2.5% native agarose gel. Different complexes were indicated. (**D**) HeLa cells were transfected with corresponding siRNAs for 56 hours before total RNA was harvested and used for synthesizing random-priming cDNAs. Gene-specific primer pairs targeting flanking exons of alternative exons were labeled with ^{32}P on 5' end (only reverse primers were labeled) and used for PCR. Products of the PCR were resolved in 4% urea-PAGE. Included and skipped products were indicated on a gel. Intensity of the different spliced products were determined, and then used for calculating % inclusion for each knockdown.

CHAPTER THREE

Stem–Loop 4 of U1 snRNA Is Essential for Splicing and Interacts with The U2 snRNP-specific SF3A1 Protein During Spliceosome Assembly

INTRODUCTION

Intron removal is catalyzed by an ~40S RNP complex called the spliceosome, which consists of five small nuclear ribonucleoprotein particles (snRNPs) (U1, U2, U4, U5, and U6) and ~150 auxiliary proteins (120). In vitro, the spliceosome assembles onto an intron through the sequential binding of the snRNPs (121,122). The 5' splice site is recognized by the U1 snRNP through base-pairing between the pre-mRNA and the snRNA. At the 3' end of the intron, the U2 auxiliary factor (U2AF) and splicing factor 1 (SF1) bind to the 3' splice site and branch point sequence, respectively. Initial association of the U2 snRNP can be ATP-independent and has been proposed to occur through interactions with the U2AF65 protein(16,123). This ATP-independent complex is referred to as the E complex. However, the stable association of U2 with the pre-mRNA requires ATP hydrolysis and formation of a base-pairing interaction between U2 snRNA and the branch point sequence with displacement of SF1. This pre-mRNP complex containing both the U1 and U2 snRNPs is called the pre-spliceosomal A complex. Recruitment of the U4/U6.U5 tri-snRNP forms the precatalytic B complex. In another mode of B complex formation, the U4/U6.U5 tri-snRNP binds with the U2 snRNP to the 3' splice site, preceding interactions with the 59 splice site-bound U1 snRNP (30). The B complex formed by either pathway undergoes extensive remodeling in its RNA and protein interactions. These rearrangements include base-pairing of the

U6 snRNA with both the U2 snRNA and the 59 splice site, recruitment of the nineteen complex (NTC), and displacement of the U1 and U4 snRNPs. This forms the activated B (B*) complex in which the first trans-esterification step of splicing occurs to detach the 5' exon and form the intron lariat intermediate. Completion of the first step generates the C complex, which carries out the second catalytic step involving intron lariat detachment and exon ligation. The multiple transitions in spliceosome conformation and assembly are driven by eight evolutionarily conserved DExD/H-type RNA-dependent ATPases/helicases: UAP56, Prp5, Prp28, Brr2, Prp2, Prp16, Prp22, and Prp43 (124,125). These proteins are tightly regulated and thought to provide proofreading functions during spliceosome assembly and splicing catalysis.

The pairing of splice sites is a critical step for accurate spliceosome assembly and splicing catalysis. Analyses of the rate of missplicing indicate that pairing occurs with high fidelity (126). Key questions for understanding the regulation of alternative splicing regard when the decision to pair splice sites is made during assembly and what proofreading mechanisms maintain the fidelity of this pairing. For a simple single intron pre-mRNA studied in vitro in HeLa cell extracts, spatial proximity between the two intron ends is seen prior to A complex formation, but commitment to splice site pairing is thought to occur during the transition to the A complex (9,10,21).

Candidate factors for mediating contact between the 5' and 3' splice site complexes include the DEAD-box proteins Prp5 and UAP56 as well as the SR and SR-related proteins. In *Schizosaccharomyces pombe*, the Prp5 protein has been proposed to bridge the splice site complexes across an intron by forming interactions with an SR-related protein (Rsd1, associated with the U1A protein) and the U2 snRNP-specific

protein SF3B1 (127). In Saccharomyces cerevisiae, Prp5 has been shown to play an important role in maintaining the fidelity of the U2 snRNA/branch point interaction (128). Associations between the U1 snRNP protein U1 70k and the 39 splice site-bound protein U2AF have also been reported. This interaction was proposed to be mediated by the SR protein SC35 in mammals but by a direct contact in yeast (11,81,129,130). In HeLa cell extracts, the SR-related proteins Srm160 and Srm300 have also been shown to bridge the U1 and U2 snRNPs (131,132). Finally, in yeast and humans, the UAP56 protein was shown to facilitate the stable ATP-dependent contact of the U2 snRNP with the branch point sequence, but its role in cross-intron bridging is not known (17,18,133,134). A recent yeast study reported genetic interactions between the U1 snRNA and the branch point-binding complex of Msl5 (branch point-binding protein or SF1 in humans) and Mud2 (U2AF65 in humans) (135). It is not yet clear whether any of the Prp5-, UAP56-, Srm160-, or Srm300-mediated contacts or functions are required and sufficient for splice site pairing or whether the intron-bridging interactions seen in yeast are common to all organisms. To maintain high accuracy of pairing, it is likely that multiple contacts are made between the 5' and 3' splice site complexes and that these contacts change over the splicing cycle.

In previous studies, we analyzed the mechanism of the Src N1 exon splicing repression by the polypyrimidine tract-binding protein 1 (PTBP1) (7,8,85,105,110). We found that the pre-mRNA-bound PTBP1 does not interfere with U1 snRNP binding to the N1 exon 59 splice site (7). In repressing N1 exon splicing, PTBP1 interacts with stem–loop 4 (SL4) of the U1 snRNA and alters the interaction of U1 with the pre-mRNA to prevent formation of a functional spliceosome (110). These results implied that the

PTBP1 interaction with SL4 might block U1 snRNP contacts critical for its further assembly into the spliceosome. In this study, we show that SL4 of U1 snRNA is important for pre-mRNA splicing and identify the U2 snRNP-specific SF3A1 protein as its interacting partner. Our analyses show that the interactions of SL4 in U1 snRNA are required for formation of the prespliceosomal A complex.

MATERIALS AND METHODS

Plasmid constructs and antibodies

The three-exon and two-intron reporter plasmid pDUP51 has been described previously (136,137). In the Dup51p reporter, the wild-type 59 splice site of exon 2 was changed from CAG/GUUGGUAUC to a mutant 5' splice site, AUG/GUGUGUAUC (''/'' is the exon–intron boundary), which had been shown to cause skipping of PCDH15 exon 3 in Usher syndrome (138). The U1 snRNA expression plasmid pNS6U1, originally developed in the laboratory of Nouria Hernandez (University of Lausanne), was a gift from Adrian Krainer (Cold Spring Harbor Laboratory). The constructs expressing mutant U1 snRNAs were generated using PCR mutagenesis and oligonucleotides carrying mutations. The sequences of all oligonucleotides used for mutagenesis are available on request. All mutations were verified by DNA sequencing. The constitutively spliced AdML and BS273 pre-mRNAs were transcribed from linearized plasmids pSPAd and pBS273, respectively. pSPAd contained the first three exons of the AdML with shortened introns (139). After cleavage within the second intron and SP6 transcription, the resulting pre-mRNA contained exon 1, intron 1, exon 2, and a short fragment of intron 2 from the AdML transcription unit. BS273 was similar to BS27 and BS713 as

described in Chou et al (2000) (105). It contained exons 3, N1, and 4 from the Src gene. Exons 3 and N1 were joined to delete intron 3 and PTBP1-binding elements essential to splicing repression. The remaining shortened intron between exons N1 and 4 carried point mutations that improved the 59 and 39 splice sites. As a result, this transcript was not regulated by PTBP1 and spliced efficiently in HeLa nuclear extract (140).

For Figures 6 and 7, anti-SF3A1, SF3A3, U1C, and U1-70K antibodies were raised in rabbits against haptenized peptides from each protein (Thermo Scientific). Peptides included SF3A1 (4654; CGAVIHLALKERGGRKK), SF3A3 (4658; CKKTYEDLKRQGLL), U1C (U1C-CT; PMMVPTRPGMTRPDRC), and U1 70K (U1 70K-CT; YLAPENGYLMEAAPEC). These antibodies were affinity-purified, and their specificity was confirmed by immune-precipitation and immune-blotting. In Figure 5C and Supplemental Figure S5, the antibodies against SF3A1, SF3A2, and SF3A3 were gifts from Angela Kramer (University of Geneva). The antibody against SF3B1 was a gift from Robin Reed (Harvard Medical School). The anti-U2B0 was from Abcam, and the anti-His antibody was from Santa Cruz Biotechnology.

Cell culture, transfection, and primer extension

HeLa cells were cultured in DMEM containing 10% FBS and antibiotics (100 U/mL penicillin, 100 mg/mL streptomycin). For transfection, $1x10^5$ cells per well of a six well plate were transfected with 0.4 mg of Dup51 or Dup51p reporter and 3.6 mg of control (pcDNA3.1) or U1 plasmid using Lipofectamine 2000 according to the manufacturer's protocol. Cells were harvested after 48 h, and RNA was extracted using

Trizol or prepared using the RNAeasy kit (Invitrogen). Primer extension was carried out using 1 mg of RNA and SuperScript II reverse transcriptase. For analyses of the Dup51 and Dup51p transcripts, oligonucleotide Dup3r (59-AACAGCATCAGGAGTGGACA GATCCC-39) was used.

For determining U1-5a snRNA expression, primer extension was carried out with oligonucleotide U17–26 (59-TGGTATCTCCCCTGCCAGGT-39) in the presence of only dATP. To assess the mutant U1 snRNA expression, the ratio of the exogenous mutant U1-5a snRNA to the endogenous wild-type U1 snRNA was calculated from the band intensities (Supplemental **Figure** S1). With the exception of M10f and M10p, this ratio ranged between 0.1 and 0.26 for the U1-5a variants. Endogenous U1 is estimated to be present at 10^6 copies per cell in HeLa cells (Steitz et al. 1988). Thus, the lowest expression of the exogenous U1 was ~10^5 copies per cell, ~100-fold higher than an abundant mRNA. The expression of U1-5aM10f and U1-5aM10p snRNAs was higher and more variable, with the ratios ranging from ~0.6 to 1.2 for different experiments (Supplemental Figure S1). These high accumulation levels did not apparently result in increased activity, as the level of exon 2 inclusion with co-expression of M10f (~82%) (Figure 2B, lane 12) and M10p (~72%) (Figure 2B, lane 15) was still lower than that observed for U1-5a (~94%) (Figure 2B, lane 3). Some other U1 mutants (such as M3, M10d, and M10e) that have expression ratios in the lower range (0.11–0.17) had exon 2 inclusion levels of ~90%. These data indicate that at the observed snRNA expression levels, the exogenous U1-5a snRNAs were not limiting for the splicing of the reporter pre-mRNAs. Changes in the reporter splicing most likely reflect changes in the activity of the snRNAs due to mutations in the SL4 region.

For SILAC experiments, minimum essential medium Eagle (similar to M8028, Sigma) without sodium bicarbonate, glutamate, arginine, and lysine was ordered from AthenaES. 13C-arginine (R6) (catalog no. CLM-2265) and 13C-lysine (K6) (catalog no. CLM-2247) were purchased from Cambridge Isotope Laboratories. The complete medium was reconstituted with 2 mg/mL sodium bicarbonate, 4 mM glutamate, 1X nonessential amino acids (Life Technology), 5% PBS-dialyzed newborn calf serum (Omega Scientific), 1X penicillin–streptomycin (Life Technology), and either ‘‘light’’ (R0K0) or ‘‘heavy’’ (R6K6) arginine and lysine at 12.6 g/L and 7.25 g/L, respectively. HeLa S3 cells were grown in medium containing either R0K0 or R6K6 for five to six doubling cycles to obtain high labeling efficiency (>99%). The cells were harvested in log phase, processed for nuclear extraction, and stored at -80°C as described (85,89).

RNA affinity chromatography

Biotinylated wild-type (Bi-GGGGACUGCGUUCGCGCUUUCCCC) and mutant (Bi-GGAUACUUAUUUCGAUAUUUAUCC) U1-SL4 RNA were custom-synthesized by Thermo Scientific (Figure 3A). Biotinylated U1-SL4M10 (Bi-AUAUACUUAUUUCGA UAUUUAUAU),U1-SL4M10h (Bi-GGGGACUUAUUUCGAAAUUAAAAA), and SL4-Dm(Bi-CGGGAAUGGCGGUUCGCGCCGUCCCG) were custom-synthesized by Integrated DNA Technologies. One-thousand picomoles of RNA were bound to 20-mL of Neutravidin agarose beads (Thermo Scientific) in buffer DG (20 mM HEPES at pH 7.9, 80 mM K-glutamate, 0.1 mM EDTA, 1 mM DTT, 20% glycerol) for 2 h at 4°C. Unbound RNAs were removed by washing the beads with 600 mL of buffer DG four times. Splicing reactions of 600 mL containing 0.4 mM ATP, 2.2 mM $MgCl_2$, 20 mM

creatine phosphate, and 50% of either heavy (for wild type) or light (for mutant) HeLa S3 nuclear extract were mixed with the beads. The reactions were incubated for 90 min at 30°C on a rotating shaker. After the incubation, the supernatant was removed; the beads were washed four times with 600 mL of buffer D containing 2.2 mM $MgCl_2$. The washed beads from the wild-type and mutant reactions were then combined. The bound RNP complexes were eluted by incubating for 30 min at 37°C in 60 mL of RNase solution containing 3 mL of RNase A/T1 cocktail (Life Technology), 10 mM Tris-HCl (pH 7.2), 1 mM $MgCl_2$, and 40 mM NaCl. For salt elution, the combined beads were successively eluted with buffer D-NaCl (using NaCl rather than KCl) containing NaCl at concentrations of 250 mM and 500 mM. All supernatants were recovered and subjected to MS analysis.

Salt-mediated dissociation of the U2 snRNP was carried out by pre-incubating 300 mL of nuclear extract with 0, 250, 500, and 1000 mM NaCl for 20–30 min at 4°C in the presence of 0.4 mM ATP, 2.2 mM $MgCl_2$, and 20 mM creatine phosphate. The reaction mix was then added to 10 mL of NeutrAvidin beads that were pre-bound with 500 nmol of biotinylated wild-type and mutant SL4 RNAs, and incubation was continued for 30 min at room temperature. Beads were washed four times with 300 mL of buffer DG. Total RNA was extracted from the bound complexes using phenol: chloroform (5:1; pH 4.8), precipitated using ethanol, separated on 8% urea-PAGE gels, and visualized using ethidium bromide staining. For protein analysis, the bound complexes were eluted by boiling the beads in 1X SDS-PAGE sample buffer, separated on 10% SDS-PAGE gels, and analyzed by Western blotting.

MS analysis

Protein mixtures were reduced, alkylated, and digested by the sequential addition of lys-C and trypsin proteases as previously described (90). The digested peptide mixture was fractionated online using strong cation exchange and reverse-phase chromatography and eluted directly into a LTQ-Orbitrap mass spectrometer (Thermofisher) (90,91). Tandem MS (MS/MS) spectra were collected and subsequently analyzed using the ProLuCID and DTASelect algorithms (92,93). Database searches were performed considering both the light and heavy versions of each peptide. Protein and peptide identifications were further filtered with a false positive rate of <5% as estimated by a decoy database strategy (94). NSAF values were calculated as described (95). SILAC ratios for each peptide pair were determined using the Census algorithm (96,97).

In vitro splicing and spliceosome assembly

Nuclear extract from HeLa cells was prepared as described previously (85,89). Pre-mRNAs were transcribed in vitro from plasmid pSPAd and pBS273, and splicing was carried out as described (7). The spliceosomal complexes were assembled, separated, and visualized using native agarose gels (7,141).

UV cross-linking and immune-precipitation

The 32P-labeled wild-type and mutant U1-SL4 RNAs were in vitro transcribed from DNA templates that were prepared by annealing DNA oligonucleotides. RNAs were gel-purified and ethanol-precipitated. The SL4 RNAs were incubated at a final

concentration of 300 nM in a reaction containing 2.2 mM $MgCl_2$, 0.4 mM ATP, 20 mM creatine phosphate, 10 U of RNAOUT, and 60% nuclear extract in buffer DG. The reactions were incubated for 30 min at 30°C and UV cross-linked for total energy of 1800 mJ in a Stratalinker. For immune-precipitation, the cross-linked reactions were treated with 0.15% SDS for 2 min at 95°C, diluted 10-fold with 1X PBS, and immune-precipitated using antibodies against U1 70k, SF3A3, and SF3A1 as described (142). The beads were washed four times with 1X PBS. The samples were eluted in SDS-PAGE loading buffer, separated on 8% gel, visualized by PhosphorImager.

For UV cross-linking in spliceosomal complexes, 10 nM uniformly labeled 32P-AdML pre-mRNA was added to a 400-mL splicing reaction and incubated for 20 min at 30°C. The reaction mix was then layered on a 15%–45% glycerol gradient prepared in buffer DG (except that K-glutamate was substituted by K-acetate) and centrifuged at 37,000 rpm for 16 h at 4°C in a SW41 rotor. The gradients were aliquoted into 25 fractions, and the radioactivity was determined by scintillation counting. The positions of the spliceosomal complexes were determined by analyzing the pre-mRNA and snRNAs in the fractions. Total RNA from fractions 13, 14, 18, 19, 23, and 24 was extracted and separated on urea-PAGE gels that were stained with SYBR Gold nucleic acid stain (Life Technologies), which has a lower limit for RNA detection of ~480 pg (143). Band intensities measured by Phosphor Imager and normalized by RNA length were used to calculate ratios of the snRNA bands to the pre-mRNA shown in Supplemental Figure S6B. The peak fractions were pooled, UV cross-linked, and immune-precipitated with anti-SF3A1 antibody as described above. After washing, the beads were treated with SDS-proteinase K for 20 min at 37°C, and total RNA was

extracted, precipitated with ethanol, and labeled with 32P-pCp overnight at 4°C. The samples were separated on 8% urea-PAGE gel and visualized by Phosphor Imager.

RESULTS[i]

SL4 of U1 snRNA is required for splicing in vivo

It was previously shown that the loss of splicing caused by 59 splice site mutations can be suppressed by expression of mutant U1 snRNAs carrying complementary nucleotide changes in their 59 ends (144-146). This suppression assay has allowed detailed analyses of 5' splice site recognition by the U1 snRNP. The assay can also be used to test the function of other regions of the U1 snRNA, such as SL4. By incorporating mutations at additional sites in a suppressor mutant U1 snRNA that activates a mutant 5' splice site in a splicing reporter, the effect of the new mutations can be assessed. We used a three exon–two-intron minigene reporter, Dup51, where exon 2 of the wild-type reporter is included at >90% (**Figure 3-1A** and **1C**, lane 1) (136). We changed the 5' splice site of exon 2 from CAG/GUUGGUAUC to AUG/GUGUGUAUC ("/" is the exon–intron boundary) (**Figure 3-1B**). This mutant splice site causes skipping of the protocadherin 15 (PCDH15) exon 3 in Usher syndrome (138). In the Dup51p reporter, the mutation reduced exon 2 inclusion to ~20% (**Figure 3-1A–C**, lane 2). Similar to the suppression seen in protocadherin, the loss of Dup51p exon 2 splicing could be rescued by co-expression of a U1 snRNA carrying a U>A substitution at the fifth position (U1-5a) but not by co-transfection with pcDNA or co-expression of the wild-type U1 or a U1-5g variant (**Figure 3-1C**, lanes 2–

[i] All supplemental data and materials are not presented here, but they can be found in a published manuscript "Stem–Loop 4 of U1 snRNA Is Essential for Splicing and Interacts with The U2 snRNP-specific SF3A1 Protein During Spliceosome Assembly" (23)

5). Thus, the U>A mutation in the 59 end of U1 snRNA can suppress the 5' splice site mutation in the Dup51p reporter. The co-expression of U1-5a and U1-5g did not alter the inclusion pattern of the wild-type Dup51 reporter (data not shown). The expression of U1 variants in these assays was confirmed by primer extension analysis using a $U1_{7-26}$ oligo in the presence of only dATP (**Figure 3-1D**). This extension reaction yields 22- and 23-nucleotide (nt) products for the endogenous wild-type U1 snRNA but terminates at 21 nt for the U1-5a and U1-5g variants.

The U1 snRNA in higher eukaryotes folds into a structure containing four stem–loops, with SL4 at the 3' terminus downstream from the Sm protein-binding site (Burge et al. 1999). SL4 consists of two G-C-rich stems split by a pyrimidine-rich internal loop and capped by a UUCG tetraloop (**Figure 3-2A**). To examine possible requirements of SL4 in U1 function, we introduced changes in this region of the U1-5a construct and tested their effect on the rescue of the 5' splice site mutation in the Dup51p reporter (**Figure 3-2A**). The pyrimidines in the bulge and the tetraloop were changed to adenosine in mutant U1-5aM3. The G-C base pairs were changed to A-U in either the upper stem (M10b) or the lower stem (M10a) or both stems (M10). In mutants M10c, M10d, and M10e, the strands of the stems were swapped. In M10u, G-C base pairs were changed to C-G base pairs at alternating positions in the lower stem. To disrupt base-pairing in the upper or lower stems or both, G or C residues were changed to adenosines in M10f, M10g, M10h, M10p, M10q, M10r, M10k, M10l, and M10m.

The effect of the SL4 mutations on U1 snRNP activity was assayed by co-transfection of the U1-5a constructs with the Dup51p reporter. The U1-5a variants carrying SL4 mutations were compared with U1-5a carrying the wild-type SL4 for their

ability to restore exon 2 splicing (**Figure 3-2B**, cf. lanes 4–20 and lane 3). SL4 sequence mutations that affect U1 function will compromise the ability of the U1-5a construct to rescue exon 2 inclusion. Many changes in SL4 did not affect U1-5a activity in restoring exon 2 splicing, indicating lack of a significant role for those SL4 nucleotides. Notably, changing the pyrimidine residues in the single-stranded bulge regions and tetraloop to adenosines did not affect the ability of U1 to restore exon 2 splicing (**Figure 3-2A**, M3; **2B**, lane 4). Swapping the two strands of the upper and/or lower stems (M10c, M10d, and M10e) (**Figure 3-2B**, lanes 8–10) also did not substantially affect U1 function. Flipping G-C base pairs at alternate positions in the lower stem (M10u) (**Figure 3-2B**, lane 11) also had minimal effect.

From the suppressor U1 analysis, we found that U1 function was most affected by mutations that alter the strength of base-pairing in the lower stem of SL4. Changing the G-C-rich lower stem to an A-U base-paired stem (M10 and M10a) (**Figure 3-2B**, lanes 5,6) led to loss of exon 2 splicing. G-C-to-A-U changes in the upper stem alone had minimal effect (M10b) (**Figure 3-2B**, lane 7). The largest effects were observed when the majority of base pairs in the lower stem were disrupted (M10g, M10h, M10q, M10r, and M10m) (**Figure 3-2B**, lanes 13,14,16,17,20). This analysis shows that the lower G-C-rich stem of SL4 plays an important role in U1 snRNP function.

The expression of all mutant U1-5a snRNAs was confirmed by primer extension (**Figure 3-2C**). Quantification of the mutant snRNA expression indicated that at the observed levels, the exogenous U1-5a snRNAs were not limiting for the splicing of the reporter pre-mRNA (see the Materials and Methods; Supplemental Figure S1). The loss of DUP51p exon splicing most likely reflects changes in the activity of the snRNAs

due to the mutations in the SL4 region, although other effects cannot be ruled out.

U1-SL4 from other species is functional in human cells

A comparison of U1-SL4 sequences from humans, *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Schizosaccharomyces pombe* shows significant differences in their sequences and structures (Supplemental Figure S2A). Human SL4 shares 68%, 59%, and 19% sequence identity with the *D. melanogaster*, *C. elegans*, and *S. pombe* RNAs, respectively, although all of these SL4 sequences can be folded into stem–loop structures (Supplemental Figure S2B). In *C. elegans* and *S. pombe*, the stem lacks the pyrimidine-containing internal loop, and in Drosophila melanogaster, there is a single bulged adenosine. In *S. pombe*, the loop is bigger than the tetraloop found in other organisms, and the stem is shorter. In all species, the lower portion of the terminal stem is highly G-C-rich. Note that the budding yeast *S. cerevisiae* does not have a clearly homologous terminal SL4 structure. However, the *S. cerevisiae* U1 snRNA is unusually long (568 nts) and may have an equivalent structure internal to its normal position.

To further investigate the sequence requirements for SL4 in U1 function, we made U1-5aM10s and U1-5aM10t constructs that eliminate the internal loop separating the upper and lower stems (Supplemental Figure S1B). We also made chimeric U1-5a constructs carrying the *D. melanogaster* (Dm), *C. elegans* (Ce), and *S. pombe* (Sp) sequences in place of the human SL4. All of these constructs were active in the U1 complementation assay (Supplemental Figure S2C, lanes 4–8). The activity of the U1-5aSL4Sp construct showed that the size of the SL4 loop is not critical for U1 function.

The G-C base pairs at the base of SL4 that were found to be important for the human U1 function are present in all of these constructs. Taken together, these experiments indicate that the G-C base pairs in the lower stem of SL4 play an important role in U1 snRNP function.

Free U1-SL4 inhibits pre-mRNA splicing in vitro by blocking formation of the pre-spliceosomal A complex

To assess whether the SL4 engaged in interactions essential for splicing, we examined the effect of free U1-SL4 on pre-mRNA splicing and spliceosomal complex assembly in vitro. Short 24-nt RNAs containing just the terminal U1 hairpin were transcribed in vitro (**Figure 3-3A**). HeLa nuclear extract active for in vitro splicing was pre-incubated with increasing concentrations of free wild-type and mutant SL4 RNAs at 4°C. Uniformly 32P-labeled Adenovirus Major Late transcription unit (AdML) pre-mRNA was then added, and incubation was continued at 30°C. The wild-type SL4 RNA inhibited pre-mRNA splicing with a half maximal inhibitory concentration (IC50) of ~3 mM (**Figure 3-3A**, lanes 2–7 and **3B**). In contrast, the mutant SL4 RNA did not have a significant effect on splicing activity (**Figure 3-3A**, lanes 8–13 and **3B**). A similar analysis of another constitutively spliced pre-mRNA (BS273 derived from Src) showed inhibition by wild-type SL4 with an IC50 of ~6 mM and, again, no inhibition by the mutant RNA (Supplemental Figure S3A). The SL4M mutation used in this in vitro analysis was also confirmed to reduce U1 activity in the U1 suppression assay, with an expression level comparable with the U1-5a snRNA (Supplemental Figure S4A, B). At the IC50 value, the concentration of U1-SL4 in the splicing reaction is ~10-fold higher

than the U1 snRNP concentration, which we estimated to be ~0.12 mM (data not shown) and has been reported to be in the range of 0.2–0.3 mM in similar HeLa nuclear extracts (147).

Splicing complex analysis on the AdML pre-mRNA using native agarose gels showed formation of the ATP-dependent spliceosomal A, B, and C complexes in the control reactions where no SL4 RNA was added (**Figure 3-3C**, lanes 1–4). In the presence of the wild-type (**Figure 3-3C**, lanes 5–8) but not the mutant (**Figure 3-3C**, lanes 9–12) SL4 RNA, there was a dramatic decrease in all spliceosomal complexes. Analyses in the absence of ATP showed equal formation of the ATP-independent E complex in the presence of either wild-type or mutant U1-SL4 (**Figure 3-3D**). Similarly, complex analysis on the BS273 transcript showed loss of A, B, and C complexes in the presence of the wild-type SL4 but not the mutant (Supplemental Figure S3C, lanes 1–3). Again, no effect on E complex assembly was seen in the presence of the wild-type or mutant RNA (Supplemental Figure S3C, lanes 4–6). These splicing complex analyses indicate that the excess of free U1-SL4 inhibits splicing by interfering with the transition from the E complex to the A complex.

The human U1-SL4 interacts with the SF3A1 protein

The free U1-SL4 presumably inhibits splicing and spliceosome assembly by competing for interactions of the endogenous U1 snRNA. To identify U1-SL4-interacting partners, we used a combination of stable isotope labeling by amino acids in cell culture (SILAC), RNA affinity chromatography, and mass spectrometry (MS). SILAC coupled with MS is a powerful unbiased quantitative proteomics approach that

was previously applied to the identification of specific interactions of proteins with RNA (148,149). An advantage of SILAC is that it can detect specific interactions in the presence of abundant nonspecific background binding (150,151). Therefore, mild incubation and wash buffer conditions can be used to preserve low-affinity but specific interactions such as individual contacts that may occur in the spliceosome.

Figure 4A shows a diagram of the SILAC-based RNA pull-down strategy (148). Nuclear extracts were prepared from cultures of HeLa S3 cells that were metabolically labeled in medium containing either $^{13}C6$ (heavy) or $^{12}C6$ (light) arginine and lysine. Wild-type and mutant biotinylated U1-SL4 RNAs (**Figure 3-3A**) were pre-bound to Neutravidin beads and then incubated in heavy and light extracts, respectively. Unbound proteins were washed away under the same mild buffer conditions (buffer DG; 80 mM K-glutamate). After washing, the beads containing the wild-type and mutant SL4 complexes were combined, and the bound proteins were eluted by RNase A/T1 digestion. Proteins were digested with proteases (trypsin and lys-C) and subjected to MS analysis using multidimensional protein identification technology (MudPIT). SILAC ratios were calculated from the intensities of the heavy and light (wild-type:mutant) peptides using Census, a program for proteomic quantification (**Figure 3-4B**; Supplemental Table S1) (96,97). These ratios are close to one for nonspecific proteins that bound equally well to the wild-type and mutant RNAs. For proteins enriched in the wild-type RNA eluate, the SILAC ratio is >1, whereas proteins that bound more strongly to the mutant RNA yield a ratio of <1. Normalized spectral abundance factor (NSAF) values were also calculated for each peptide as an alternative for identifying proteins enriched in either the wild-type or mutant RNP complexes or proteins present in both

complexes (Supplemental Table S2). The NSAF values for each protein are calculated from the total number of spectrum-matching peptides from the protein (spectrum counts) that are then normalized for protein length (95).

Multiple U2 snRNP proteins had SILAC ratios >1, indicating enhanced binding to the wild-type SL4 over the mutant RNA (**Figure 3-4B**). Strikingly, the NSAF analysis also showed that all of these U2 snRNP-specific proteins were detected only in the wild-type eluate and not in the mutant eluate (Supplemental Table S2). Proteins from the U1 snRNP and U4/U6.U5 tri-snRNP did not bind to either SL4 RNA. Other proteins with high SILAC ratios included certain hnRNP and SR proteins (Supplemental Table S1). Some of these are known to be U2 snRNP-associated, which could lead to their enrichment in the wild-type SL4 complexes. In a previous study, we showed that PTBP1 binds the internal loop of SL4, which is present in both wild-type and mutant SL4 RNAs (110). Consistent with this, PTBP1 bound to both SL4 RNAs (Supplemental Table S2). Although their SILAC ratios were high, the NSAF analysis indicated that many of the hnRNP proteins were present in both wild-type and mutant SL4 eluates (Supplemental Table S2). hnRNP and other proteins that interact with SL4 are interesting for their possible involvement in splicing regulation. Such proteins may contact SL4 to alter spliceosome assembly, as seen with PTBP1. However, in this study, our focus was on the contacts of SL4 during general spliceosome assembly, and thus we were particularly interested in the SL4 interactions with the U2 snRNP proteins.

Immunoblot analysis confirmed the presence of the core U2 protein U2B", the SF3A complex proteins A1 (also known as SF3a120) and A3 (also known as SF3a60), and the SF3B complex protein SF3B1 (also known as SF3b155) in the wild-type SL4

complexes but not in the mutant (Supplemental Figure S5; data not shown). MS and immunoblot analyses after high-salt elution also showed binding of U2 snRNP proteins to the wild-type SL4 but not the mutant (data not shown). The binding of multiple U2-specific proteins to the wild-type U1-SL4 complex indicated the binding of the whole U2 snRNP. It was not clear which U2 protein was making a direct contact with U1-SL4 during affinity purification.

The mature U2 snRNP is a 17S complex that consists of the core U2 particle and the SF3A and SF3B complexes. It was reported previously that the association of the SF3A and SF3B complexes with the core U2 snRNP is salt-sensitive (152,153). Incubation with ~250 mM NaCl dissociates the SF3A complex from the core U2 snRNP. The SF3B complex is released from the core snRNP at 500 mM (**Figure 3-5A**). To determine which U2 snRNP-specific protein might directly interact with U1-SL4, we took advantage of the salt-sensitive property of the U2 snRNP. Extracts were pre-incubated with increasing concentrations of NaCl and then used for RNA affinity purification using biotinylated wild-type and mutant U1-SL4 RNAs (**Figure 3-3A**). The bound complexes were analyzed for snRNAs and U2 snRNP-specific proteins. The snRNA analysis showed binding of the U2 snRNA to the wild-type U1-SL4 in extract under a standard splicing reaction (**Figure 3-5B**, lane 2). Pre-incubation of the extract with 250 mM or higher salt concentrations led to loss of U2 snRNA binding (**Figure 3-5B**, lanes 3–5). The U2 snRNA did not bind to the mutant SL4 complex in any condition (**Figure 3-5B**, lanes 6–9).

As seen previously, immunoblot analysis of proteins bound to wild-type U1-SL4 in splicing conditions confirmed the binding of U2 snRNP proteins SF3A1, SF3A2,

SF3A3, SF3B1, and U2B" (**Figure 3-5C**, lane 2). In contrast, the U1 snRNP-specific protein U1C did not bind to SL4. After pre-incubation of the extract with NaCl, binding of SF3A2 (66 kDa), SF3B1 and the core U2 protein U2B" were lost (**Figure 3-5C**, lanes 3–5). Interestingly, although the amounts of SF3A1 (120 kDa) and SF3A3 (60 kDa) proteins bound to the wild-type U1-SL4 decreased after salt pre-incubation (**Figure 3-5C**, lanes 2–5), the interaction of these proteins was more resistant to NaCl, and substantial amounts of each of these proteins remained bound at higher salt concentrations. The mutant SL4 did not bind SF3B1, U2B", SF3A1, and SF3A2 (**Figure 3-5C**, lanes 6–9). The SF3A3 protein did bind to the mutant SL4 but at lower levels than to wild-type. These analyses indicate that the U2 snRNP associates with the wild-type U1-SL4. Upon high-salt pre-incubation, although the SF3A complex dissociates from the core U2 snRNP, the SF3A1 and SF3A3 proteins can still bind to SL4, indicating that one of these proteins is likely to directly interact with U1-SL4.

To identify the SF3A complex protein that directly contacts U1-SL4, we raised antibodies to SF3A1 and SF3A3 that were used in UV cross-linking and immune-precipitation experiments. Uniformly 32P-labeled wild-type and mutant U1-SL4 RNAs were incubated in HeLa nuclear extract. After incubation, the reactions were UV cross-linked, denatured by treatment with 0.15% SDS, and then immune-precipitated using antibodies against U1 70k, SF3A3, and SF3A1 proteins as described (142). SDS-PAGE analysis showed cross-linking of an ~120-kDa and an ~55-kDa protein to the wild-type SL4 (**Figure 3-6A**, lane 1). The ~55-kDa protein cross-links to both the wild-type and mutant SL4 RNAs and is presumably PTBP1 (see below). The 120-kDa protein did not cross-link efficiently to the mutant SL4 RNA (**Figure 3-6A**, lanes 5–8). Most

interestingly, the anti-SF3A1 antibody (**Figure 3-6A**, lane 3), but not the U1 70k (**Figure 3-6A**, lane 2) or SF3A3 (**Figure 3-6A**, lane 4) antibody, significantly enriched the ~120-kDa protein. Thus, the binding of U2 snRNP to the wild-type U1-SL4 likely occurs through direct contact with the SF3A1 protein.

The interaction between U1 snRNA and SF3A1 occurs in pre-spliceosomal complexes

We next examined whether the interaction between SF3A1 protein and U1 snRNA occurred in spliceosomal complexes. For this, uniformly 32P-labeled AdML pre-mRNA was incubated in HeLa nuclear extract in the presence of ATP. The reactions were fractionated on glycerol density gradients, and the pre-mRNA in each fraction was measured by scintillation counting (Supplemental Figure S6A). Total RNA in peak 1 (21S; fractions 13 and 14), peak 2 (30S; fractions 18 and 19), and peak 3 (40S; fractions 23 and 24) was extracted and analyzed for the presence of pre-mRNA and spliceosomal snRNAs, and the ratio of each snRNA to the pre-mRNA was calculated from the fluorescent band intensities as described in the Materials and Methods (Supplemental Figure S6B). In peaks 1 and 2, the ratio of the U1 and U2 snRNAs to the pre-mRNA was ≥1, and that of the tri-snRNP snRNAs was low. In peak 3 fractions, the amount of the U5 and U6 snRNAs was substantially increased, while the U1 snRNA was lower than in peaks 1 and 2. U4 snRNA was also higher in peak 3 but not as abundant as the U2, U5, and U6 snRNAs that comprise the active spliceosome. Under these centrifugation conditions, the free U1 and U2 snRNPs peak in fractions 7 and 10, respectively (data not shown). Peak 1 and peak 2 contain both the U1 and U2 snRNPs

and are the expected sizes of the pre-spliceosomal E and A complexes. Peak 3 is expected to be comprised of the higher-order spliceosomal B, B*, and C complexes, in keeping with the presence of the U2, U5, and U6 snRNPs.

Since free U1-SL4 blocked assembly prior to the A complex, we looked for an interaction between SF3A1 and U1 snRNA in the pre-spliceosomal complexes. For this, fractions from peak 1 and peak 2 were pooled separately, UV cross-linked, and immune-precipitated using an anti-SF3A1 antibody that was pre-bound to γ-bind beads. The bound complexes were digested with SDS/proteinase K, and the total RNA was extracted, labeled with 32P-pCp, and analyzed for the presence of pre-mRNA and snRNAs (**Figure 3-6B**). Equivalent gradient fractions from control reactions lacking pre-mRNA (**Figure 3-6B**, lanes 2–4) were treated identically to the fractions from pre-mRNA-containing reactions (**Figure 3-6B**, lanes 5–7).

We found that after cross-linking, U1 snRNA was immune-precipitated with anti-SF3A1 antibody from peak 1 (**Figure 3-6B**, lane 7) but not from peak 2 (data not shown) fractions. An anti-His tag antibody did not immune-precipitate any of the spliceosomal snRNAs (**Figure 3-6B**, lanes 3,6). Importantly, the anti-SF3A1 immuno-precipitate from corresponding fractions of the control reaction lacking pre-mRNA yielded fourfold less U1 snRNA relative to U2 (**Figure 3-6B**, cf. lanes 7 and 4). The identity of the cross-linked snRNAs was confirmed by RNase H digestion in the presence of snRNA complementary oligonucleotides (**Figure 3-6B**, lanes 9,10). In addition to U1 snRNA, we found that U4 snRNA weakly cross-linked and was immune-precipitated with the anti-SF3A1 antibody (**Figure 3-6B**, cf. lanes 8 and 10). These results confirm that the U1 snRNA within the snRNP and not just the free SL4 interacts

with SF3A1 and that this interaction occurs on the pre-mRNA but not between the free snRNPs. We did not observe substantial cross-linking of the pre-mRNA to SF3A1 as was reported previously (16). In the earlier study, the pre-mRNA was specifically labeled in the region upstream of the branch point, and the cross-links were detected by label transfer to the protein in the pre-spliceosomal A complex. Gozani et al. (1996) found that the cross-linking of SF3A1 to the pre-mRNA was much weaker than that seen for other SF3A and SF3B proteins. Our different observation may result from different sensitivities of the assays and the fact that we are analyzing early complexes prior to the binding of U2 to the branch point.

We also did not observe substantial cross-linking between the SF3A1 protein and the U2 snRNA. This is in agreement with a previous report in which the U2 snRNA was shown to interact with the SF3A3 but not the A1 and A2 proteins (154).

The SL4 mutations that affect U1 function also affect SF3A1 protein binding

We next wanted to examine the effect of SL4 mutations that decrease U1 snRNP activity in the complementation assay on the binding of the SF3A1 protein. We measured binding of proteins to wild-type and mutant SL4 RNAs using RNA affinity chromatography and immunoblot analysis. Synthetic biotinylated SL4 RNAs—including the wild-type, the M10 and M10h mutants, and the *D. melanogaste*r SL4 (**Figure 3-2**; Supplemental Figure S2)—were incubated under splicing conditions in HeLa nuclear extract, and the bound proteins were isolated on Neutravidin beads as described above. The binding of SF3A1 and PTBP1 was assessed by immunoblot (**Figure 3-7A**). The ratio of protein bound to each SL4 variant to that bound to the wild-type human SL4 was

determined. Interestingly, the SF3A1 protein showed substantially lower binding to the mutant M10 and M10h RNAs but bound at higher levels to the *D. melanogaster* SL4 (**Figure 3-7B**, cf. lanes 3,4 and lane 5). Thus, the binding of SF3A1 correlated with the presence of the G-C-rich stem structures and not the presence of the bulged nucleotides, as was seen for U1 activity in the in vivo assays. In contrast, PTBP1 bound well to both the wild-type and M10 mutant RNAs (**Figure 3-7B**, lanes 2,3). Both of these RNAs contain the bulged pyrimidines that were shown previously to bind PTBP1 (110). The M10h mutant, which contains these nucleotides but not in a bulged structure, showed moderate binding to PTBP1 (**Figure 3-7B**, lane 4). In contrast, PTBP1 bound very poorly to the SL4-Dm RNA that lacks the pyrimidine bulge (**Figure 3-7B**, lane 5). Thus, these two proteins recognize different determinants on the SL4. The U1 70K protein did not bind to any of the SL4 RNAs (**Figure 3-7A**). The loss of function of the mutant U1-5a snRNAs in vivo correlates well with the loss of SF3A1 binding in vitro.

DISCUSSION

SL4 of the U1 snRNA is a conserved structural feature whose function in splicing has not previously been examined. SL4 is not required for 3' end formation of the U1 snRNA or for SMN complex binding during snRNP biogenesis and is not bound by any of the U1-specific proteins (155-159). Crystal and cryo-electron microscopy (cryo-EM) structures of purified or in vitro assembled U1 snRNPs show that SL4 is well separated from the first three stem–loops of U1 snRNA by the Sm ring (160-162). These as well as biochemical studies indicate that in the free snRNP, SL4 contacts the B and D2

subunits of the Sm core (162,163).

We identified a role for SL4 in pre-mRNA splicing and found that the G-C-rich lower stem is critical for U1 snRNP activity in vivo. In vitro, free U1-SL4 inhibits pre-mRNA splicing by blocking formation of the pre-spliceosomal A complex on the pre-mRNA. In these extracts, we identified the U2 snRNP protein SF3A1 as an interacting partner of U1-SL4. We found that this interaction takes place on the pre-mRNA and does not occur between the free snRNPs. The lack of interaction between the free snRNPs may indicate that a change in U1 conformation occurs upon binding the pre-mRNA that increases the accessibility of SL4. The contacts between SL4 and the Sm proteins B and D2 (162,163) may prevent di-snRNP formation and then be altered upon U1 snRNP binding to the pre-mRNA. It will be interesting to examine the accessibility of these U1 nucleotides in different states of U1 snRNP assembly into the spliceosome.

We propose that the interaction between SL4 in the U1 snRNA and the SF3A1 is one of the molecular bridges that help to pair the 59 and 39 splice sites during early spliceosome assembly (**Figure 3-8**). For the simple, single intron pre-mRNA studied here, the U1 snRNP and the U2 snRNP form their initial contacts with the pre-mRNA at opposite ends of the intron. The 5' and 3' splice sites are thought to be brought into proximity (10–20 Å) early in mammalian spliceosome assembly (9,10). Recent fluorescence resonance energy transfer (FRET) studies in yeast indicate that the 5' splice site and branch point are brought into proximity only after A complex formation and coincident with the exit of U1 from the complex (164). This interaction within the mature spliceosome presumably reflects the assembly of the active site and an even closer juxtaposition than occurs in the pre-spliceosome. It is not clear how closely the 5'

splice site and branch point approach each other in forming the U1/SF3A contact, but the U1/SF3A interaction must be transient, as the contact will be disrupted during the release of U1 snRNP from the mature spliceosome.

We observed SF3A1–U1 snRNA cross-linking in early pre-spliceosomes but cannot yet define when the interaction is lost. Peak 2 in our gradients, which is the size of complex A, did not show cross-linking. However, the yield of this complex may be too low to detect the interaction. It will be interesting in the future to characterize the early assembly steps defined by the presence of the SF3A1–U1 interaction. In the E complex, the U1 snRNP is bound to the 5' splice, and the U2 snRNP has been shown to be loosely associated via interactions with the U2AF65 protein (123). In the A complex, U2 becomes stably bound to the branch point sequence by base-pairing interactions. This does not appear to occur in the presence of excess U1-SL4. It will be interesting to purify complexes assembled in the presence of ATP but arrested with excess SL4 to assess the presence of the U1 snRNP and factors that might be important for the E-to-A transition.

A variety of protein–protein interactions have been suggested to bridge the 5' and 3' splice sites in cross-intron complexes, including contacts mediated by the Prp5 protein, SR and SR-like proteins, and other factors (81,127,132). Since mutations in SL4 do not completely abolish U1 function in vivo, additional interactions likely contribute to splice site pairing in vivo, but the precise roles of the other identified bridging contacts in early spliceosome assembly are not yet clear. In contrast to the in vivo case of partial splicing inhibition by the SL4 mutations within the suppressor U1 snRNA, the striking inhibition of the in vitro splicing reaction by the SL4 competitor may

indicate blockage of additional contacts between the U1 and U2 snRNPs. The inhibition of in vitro splicing and pre-spliceosomal A complex formation by excess U1-SL4 supports a major role for this contact in splicing.

The pairing of splice sites across the intron and the formation of the A complex are appealing steps for the regulation of alternative pre-mRNA splicing. We previously found that the regulator PTBP1 contacts SL4 within an exon complex repressed for splicing (110). Interestingly, the PTBP1-interacting sites on SL4 and the regions required for splicing are non-overlapping. PTBP1 binds to the internal loop, whereas the lower stem is required for splicing function. PTBP1 binding to SL4 could block the SF3A1 interaction to prevent further spliceosome assembly. Interestingly, we found that several other hnRNP proteins interact with the wild-type U1-SL4. Investigations of these proteins may show that SL4 in the U1 snRNA is a common contact point for regulators of alternative splicing.

The evolutionarily conserved heterotrimeric SF3A complex is essential for the assembly of a functional U2 snRNP and for pre-mRNA splicing (152,153,165,166). In yeast and humans, a trimer is formed from binding of SF3A3 and SF3A2 proteins to SF3A1, but the two smaller subunits do not interact with each other (153,165,167). In yeast, an SF3A3/A1 dimer interacts with the U2 snRNA SL2 (U2-SL2), which toggles between SL2a and SL2c conformations during the splicing cycle (168,169). This SF3A3/A1 dimer binds to SL2a but not SL2c through the first zinc finger domain of the yeast SF3A3 (170). The lack of SL2c binding has been postulated to be the mechanism for the release of SF3A from the spliceosome before the first catalytic step of splicing in yeast (170,171). Interestingly, the SL2a-interacting ZnF domain is not conserved in

human SF3A3, although the overall structure and charge distribution of the protein appear to be similar in the two species. In addition, recent chemical probing of RNA structure in human spliceosomes did not detect the mutually exclusive U2 snRNA conformations (172). This indicates that some differences likely exist in the molecular interactions and dynamics of yeast and human spliceosome assembly. Future studies will focus on defining more of these molecular contacts between the 5' and 3' splice site complexes during splice site pairing, commitment to splicing, and the transition to active spliceosomal complexes in metazoans.

FIGURES

Figure 3-1 Suppressor U1 snRNAs can rescue splicing

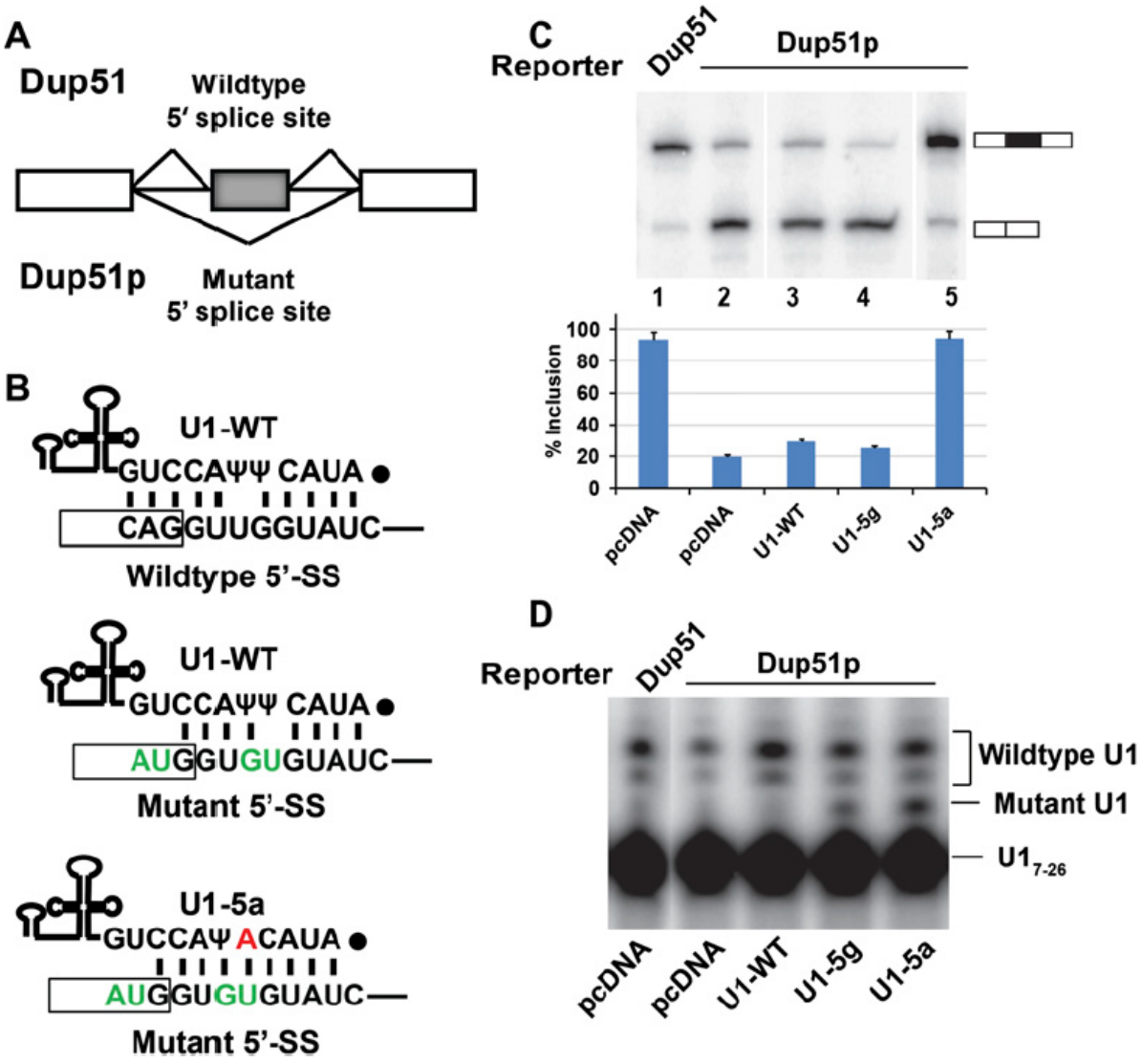


Figure 3-2 SL4 of U1 snRNA is important for U1 function

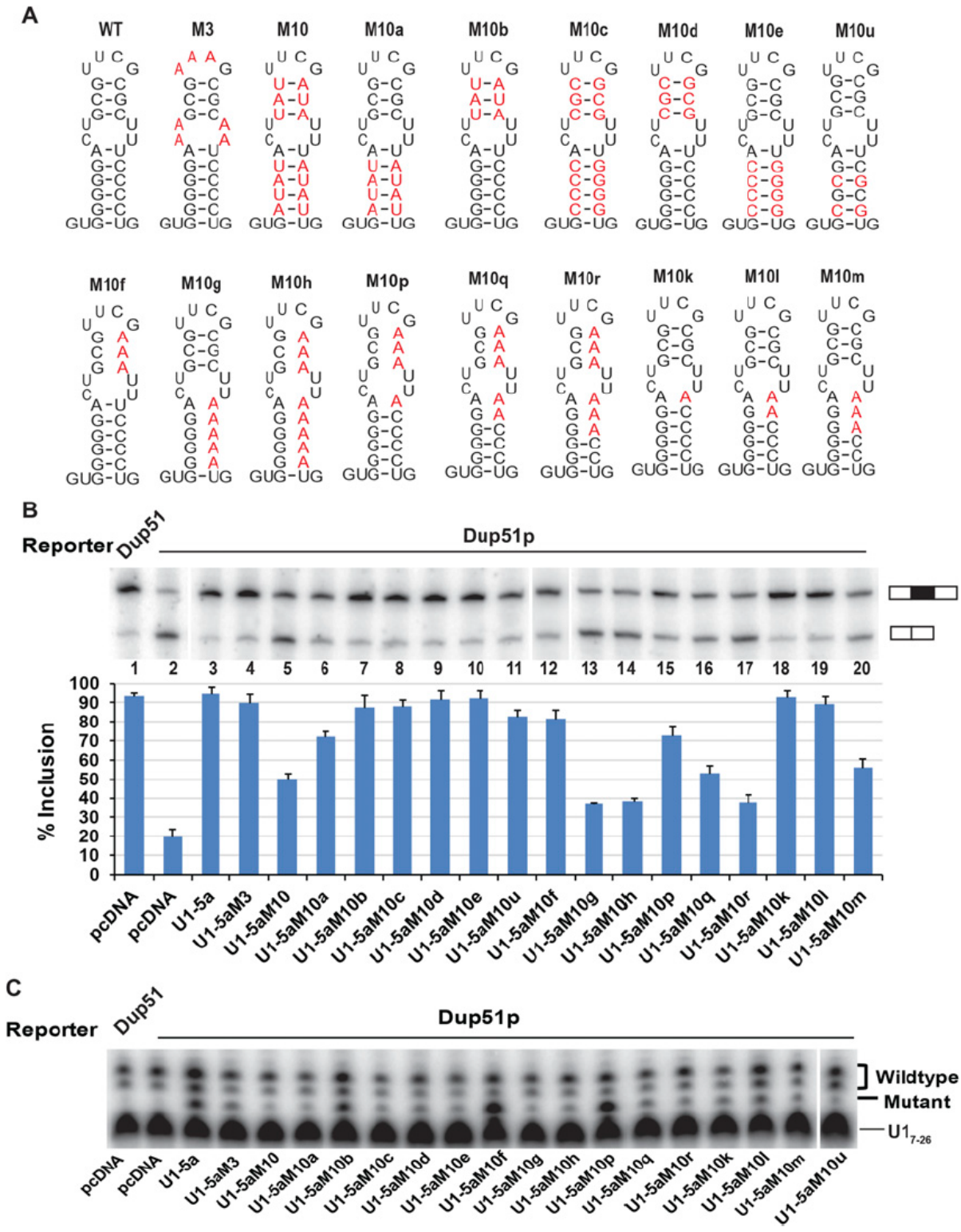


Figure 3-3 Free U1-SL4 inhibits pre-mRNA splicing in vitro

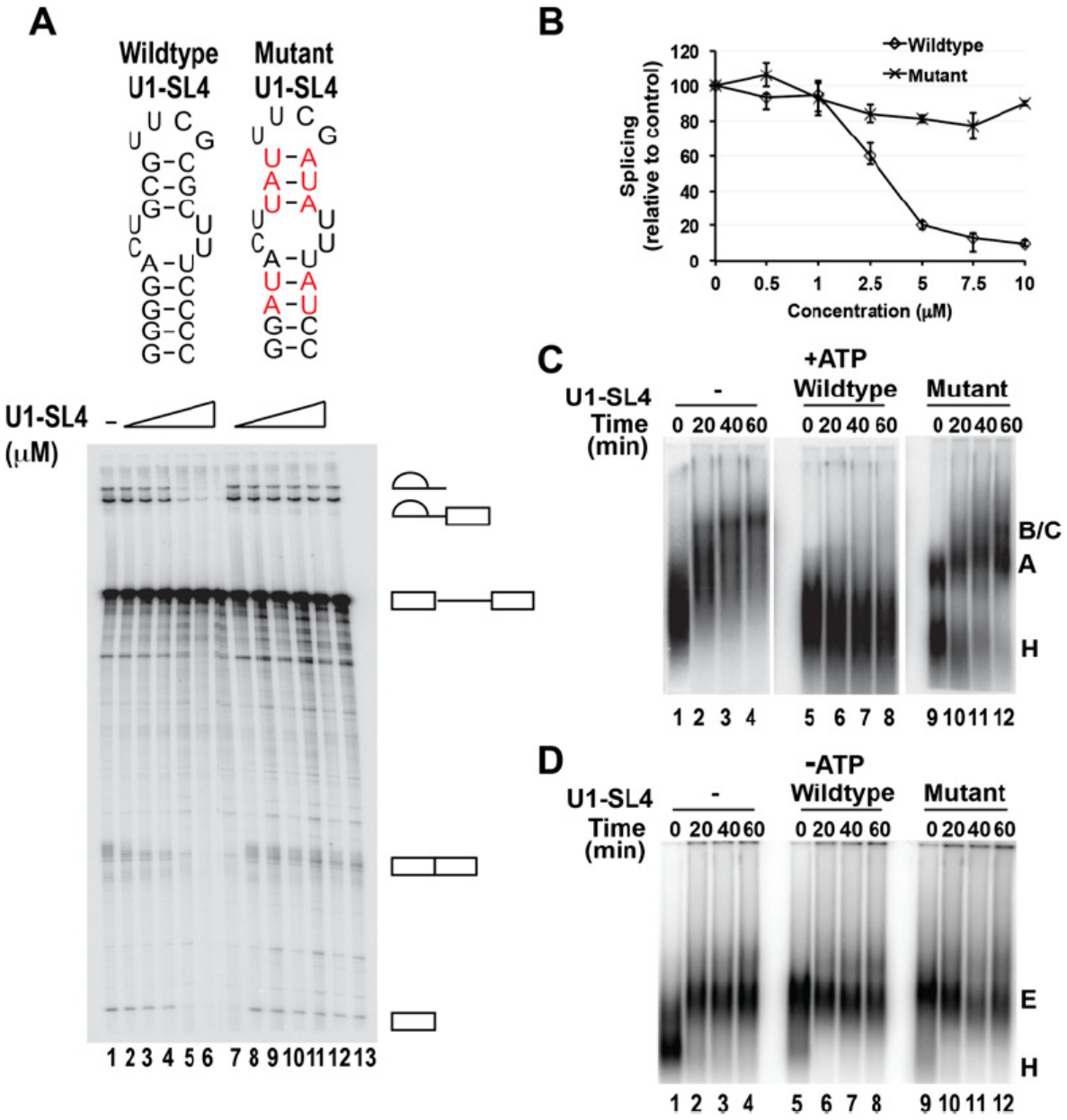


Figure 3-4 Identification of U1-SL4-interacting proteins

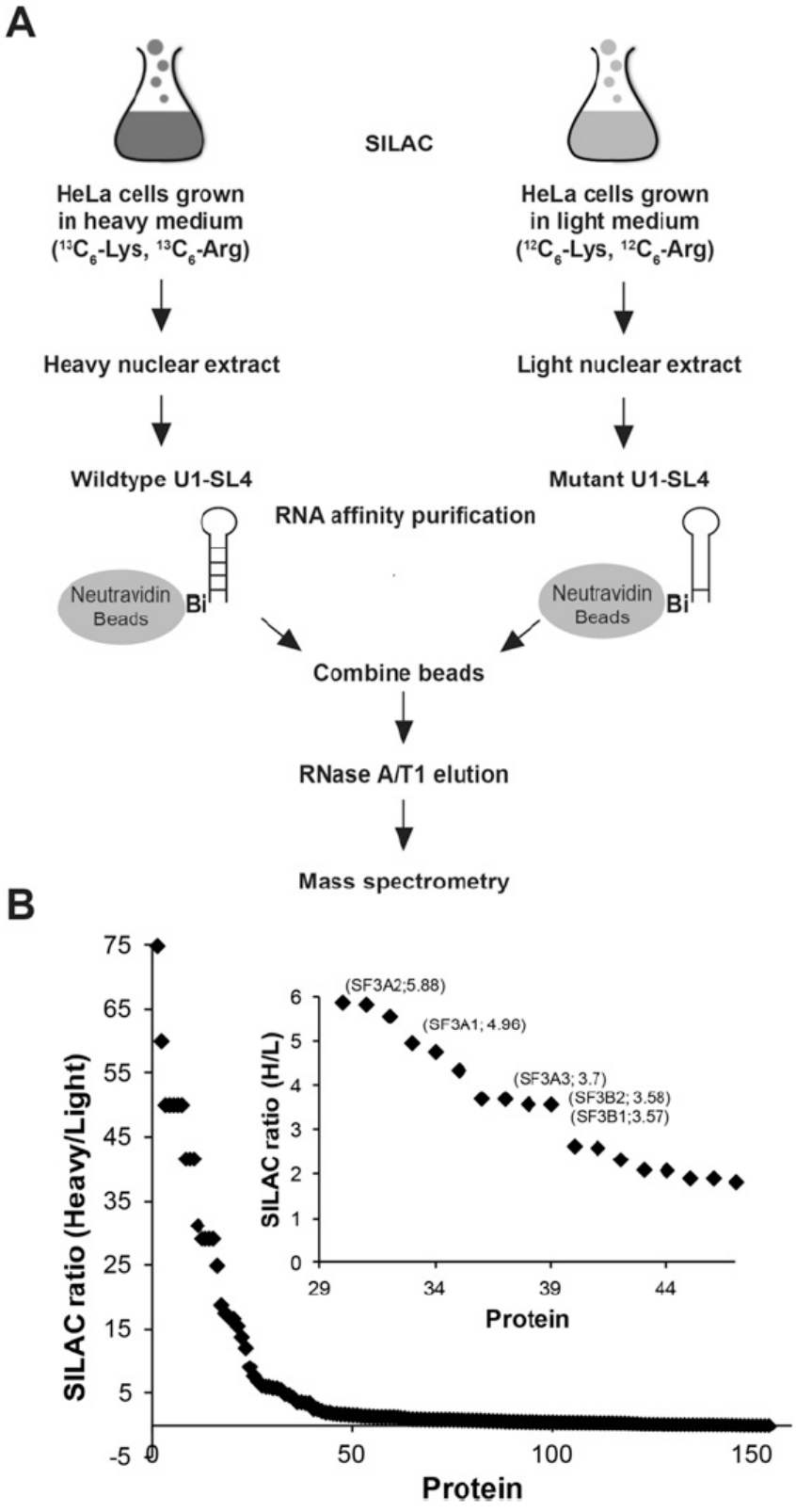


Figure 3-5 SF3A complex proteins interact with the wild-type U1-SL4

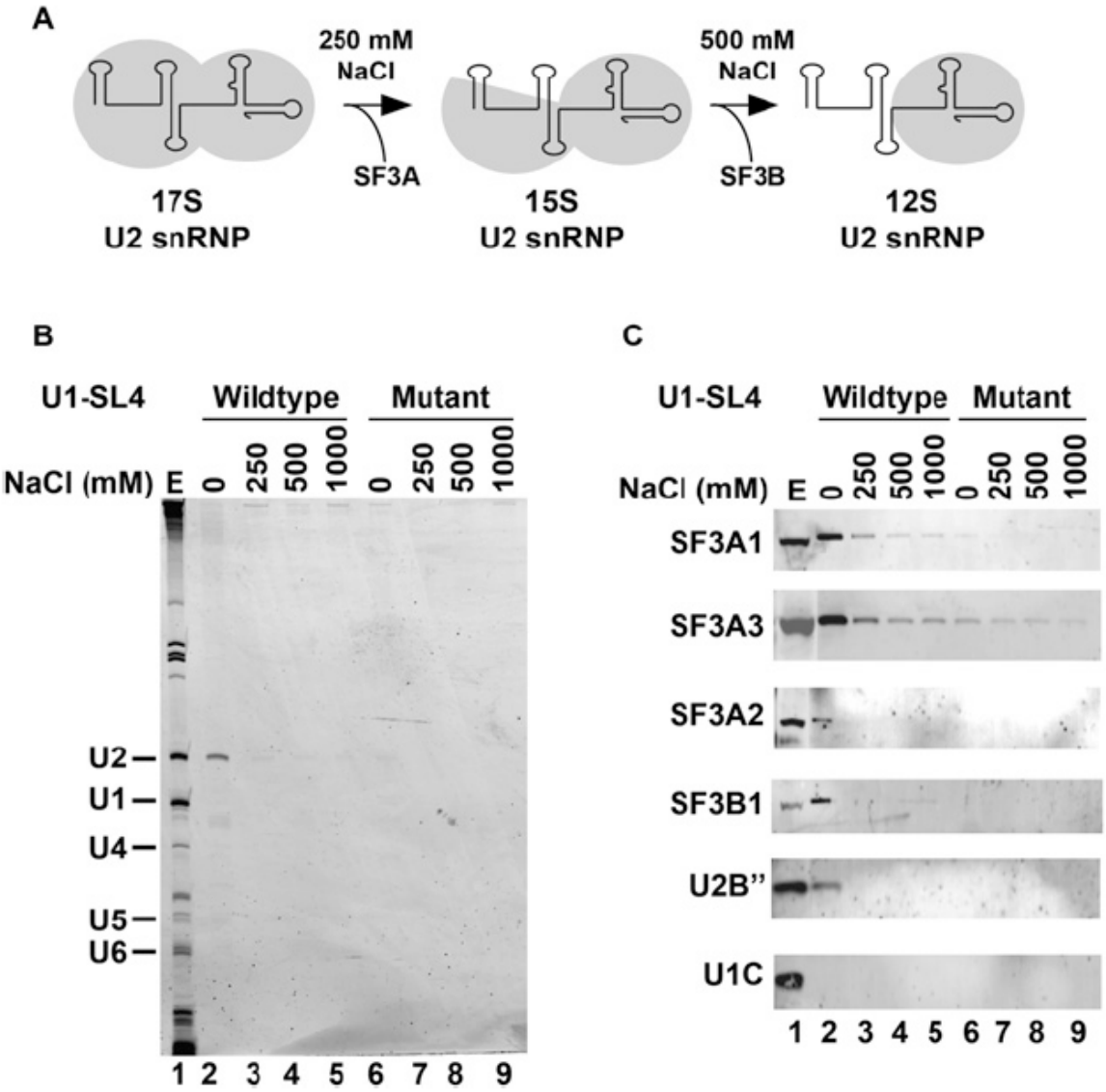


Figure 3-6 Interaction between SF3A1 and U1 snRNA occurs in pre-spliceosomal complexes

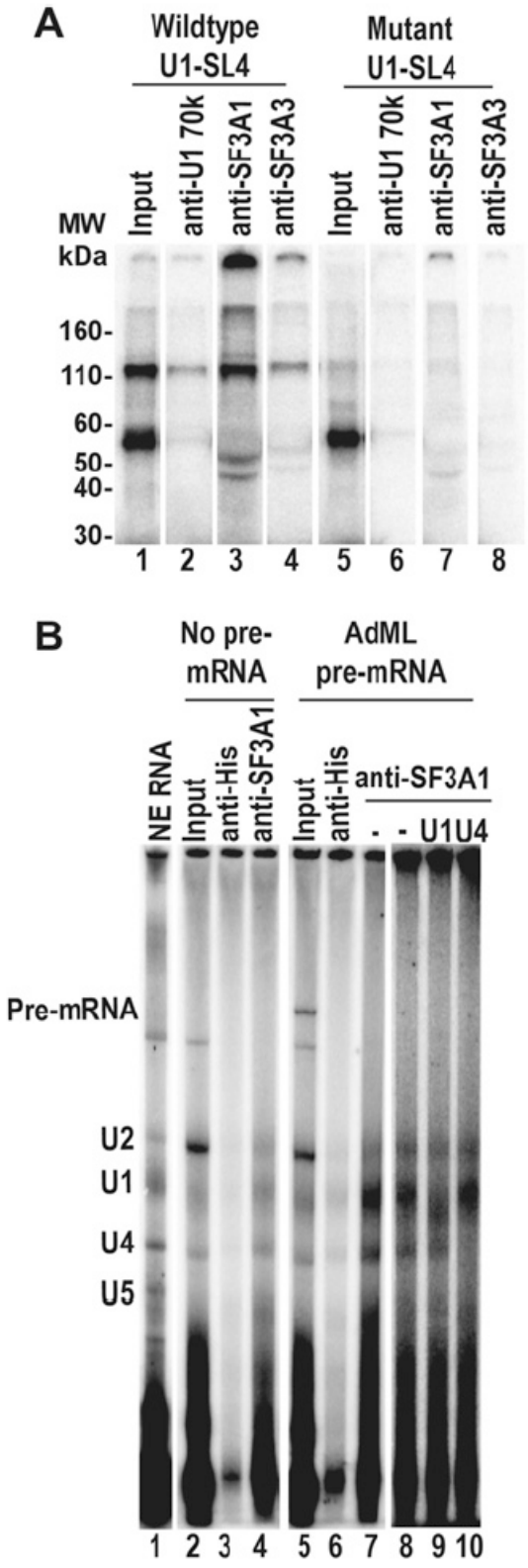


Figure 3-7 SL4 mutations that affect function also affect binding of the SF3A1 protein

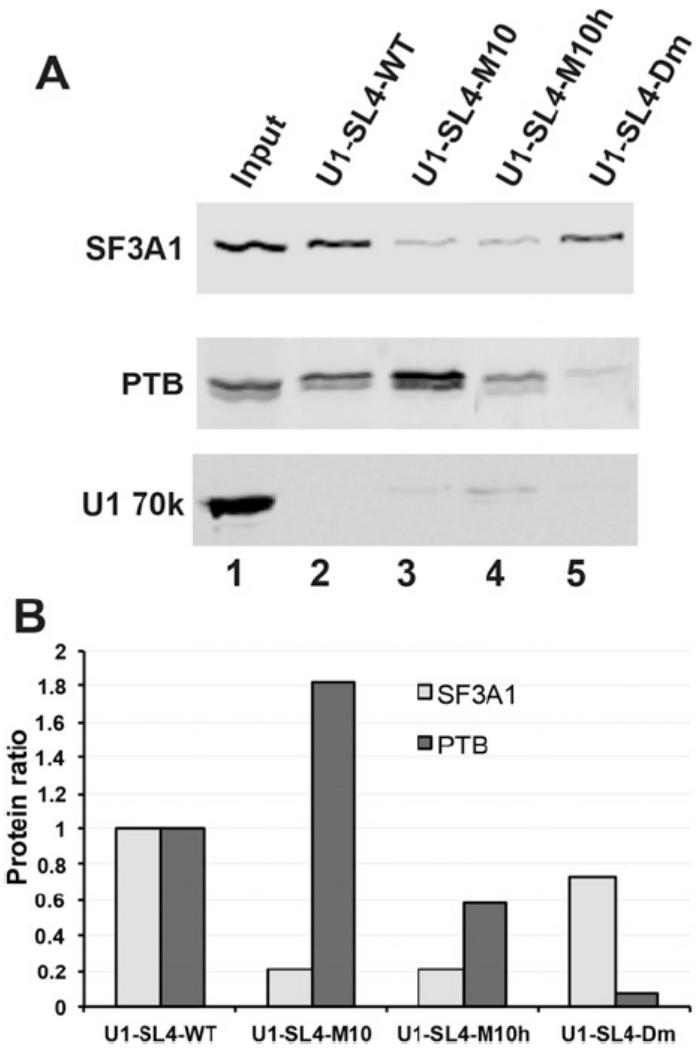


Figure 3-8 Model for the role of SL4 of U1 snRNA in splicing assembly

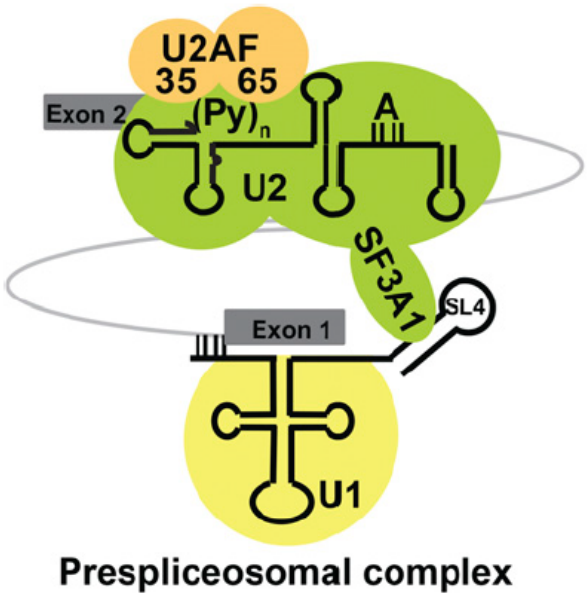


FIGURE LEGENDS

Figure 3-1 Suppressor U1 snRNAs can rescue splicing. (**A**) Schematic representation of three-exon/two intron Dup51 and Dup51p reporters. (**B**) Base-pairing of the wild-type and mutant 59 splice sites with the 59 end of U1 and U1-5a snRNAs. (**C**) Primer extension analysis of the Dup51minigenes after co-transfection with control pcDNA or U1 expression plasmids. The mRNA products are indicated at the right and quantified in the graph below. (**D**) Primer extension analysis with oligonucleotide $U1_{7–26}$ showing expression of the wild-type U1 and variant U1-5a and U1-5g snRNAs.

Figure 3-2 SL4 of U1 snRNA is important for U1 function. (**A**) Schematic of the secondary structures of SL4 region of the wild-type and mutant U1 snRNAs. Structures predicted to have the lowest DG are shown (173). (**B**) Primer extension analysis of the Dup51 minigene transcripts after co-transfection with control or U1 plasmids. The mRNA products are indicated at the right and quantified in the graph below. (**C**) Primer extension analysis with oligonucleotide $U1_{7–26}$ showing expression of the endogenous wild-type U1 and variant U1-5a snRNAs.

Figure 3-3 Free U1-SL4 inhibits pre-mRNA splicing in vitro. (**A**) In vitro splicing of the AdML transcript in HeLa nuclear extract in the absence (lane 1) or presence of 0.5, 1.0, 2.5, 5.0, 7.5, and 10 mM wildtype (lanes 2–7) and mutant (lanes 8–13) short U1-SL4 RNA competitors. Prior to addition of the pre-mRNA, the splicing reactions were pre-incubated with the SL4 RNAs for 20 min at 4°C. The RNA splicing products and intermediates are diagrammed at the right. (**B**) Splicing activity is plotted as a function

of SL4 concentration (in micromolar). Splicing was measured as the percent intensity of the intron lariat product in the presence of wild-type (A, lanes 2–7) and mutant (lanes 8–13) SL4 RNAs relative to the control reaction lacking SL4 (lane 1). (**C**) Analysis of ATP-dependent spliceosomal complexes in the presence of the short U1-SL4 RNA competitors. Transcripts were incubated in HeLa nuclear extract under splicing conditions with ATP in the absence (lanes 1–4) or presence of wild-type (lanes 5–8) or mutant (lanes 9–12) SL4 for the indicated times and separated by 2% native agarose gel. (**D**) Analysis of ATP-independent spliceosomal complexes in the presence of U1-SL4 RNA competitors. Transcripts were incubated in HeLa nuclear extract under splicing conditions without ATP in the absence (lanes 1–4) or presence of wild-type (lanes 5–8) or mutant (lanes 9–12) SL4 RNA for the indicated times and separated by 1.5% native agarose gel. Positions of the H, E, A, B, and C complexes are indicated.

Figure 3-4 Identification of U1-SL4-interacting proteins. (**A**) Protocol for SILAC, RNA affinity purification, and MS. (**B**) Graph of SILAC ratios versus protein. The inset shows the ratios for U2 snRNP-specific proteins.

Figure 3-5 SF3A complex proteins interact with the wild-type U1-SL4. (**A**) Salt-dependent disassembly of the U2 snRNP. (**B**) snRNA analysis of the complexes that bind to the biotinylated wild-type and mutant U1-SL4 RNAs after pre-incubation of HeLa nuclear extract with 0, 250, 500, and 1000 mM NaCl. Positions of the spliceosomal snRNAs are indicated at the left. (**C**) Western analyses of the proteins present in the wild-type and mutant U1-SL4 RNA complexes.

Figure 3-6 Interaction between SF3A1 and U1 snRNA occurs in pre-spliceosomal complexes. UV cross-linking and immune-precipitation identified a direct contact between the wild-type SL4 RNA and SF3A1 protein. (**A**) 32P-labeled wild-type and mutant SL4 RNAs were incubated in HeLa nuclear extracts in splicing conditions. After incubation, the reactions were UV cross-linked and immune-precipitated with antibodies against U1 70k (lanes 2,6), SF3A1 (lanes 3,7), and SF3A3 (lanes 4,8) proteins, followed by analysis by SDS-PAGE. (**B**) Splicing reactions containing ATP in the presence (lanes 5–7) or absence (lanes 2–4) of AdML premRNA were fractionated on glycerol density gradients (see Supplemental Figure S6). Gradient fractions from the 21S peak containing the pre-mRNA and U1 and U2 snRNA were pooled, UV cross-linked, and immune-precipitated with anti-SF3A1 (lanes 7–10) or anti-His (lane 6) antibodies. Equivalent fractions from gradients lacking pre-mRNA were also immune-precipitated with anti-SF3A1 (lane 4) or anti-His (lane 3) antibodies. Total RNA from the fractions (shown in lanes 2,5) and the immune-precipitated complexes was extracted and analyzed. (Lanes 8–10) The identity of the U1 and U4 snRNAs was confirmed by RNase H cleavage in the presence of U1 oligos ($U1_{1–15}$ and $U1_{64–75}$) and U4 oligos ($U4_{2–16}$ and $U4_{66–85}$). (Lane 1) Total RNA from nuclear extract was used as a marker for the U snRNAs. The positions of the U snRNAs and pre-mRNA are indicated.

Figure 3-7 SL4 mutations that affect function also affect binding of the SF3A1 protein. (**A**) Immunoblot of the SF3A1, PTBP1, and U1 70k proteins bound to immobilized SL4 RNAs after incubation in HeLa nuclear extract under splicing

conditions. SL4 variants included the wild-type human sequence, the M10 and M10h mutants, and the *D. melanogaster* SL4 (Dm). The input lane contains 1.5% of the total reaction. Lanes 2–5 contain 20% of the protein eluted from each RNA. (**B**) Quantification of the amount of bound SF3A1 and PTBP1 proteins expressed as the ratio of bound protein to that bound by the wild-type SL4.

Figure 3-8 Model for the role of SL4 of U1 snRNA in splicing assembly. Interaction of the SL4 with the U2 snRNP-specific SF3A1 protein occurs during pairing of splice sites prior to the formation of the pre-spliceosomal A complex. The presence of free U1-SL4 blocks this interaction and subsequent A complex formation. U2 is shown engaged at the branch point, as seen in the pre-spliceosomal A complex.

CHAPTER FOUR

Concluding Remarks

Together with transcriptional regulation, alternative pre-mRNA splicing plays an essential role in generating unique gene expression profiles that are specific to tissues, developmental stages and environmental reponses. Alternative pre-mRNA splicing is primarily regulated by splicing factors that bind to regulatory elements on pre-mRNAs. Once bound, these factors can either promote or disrupt recognition of splice sites to determine whether or not they will be used by splicing machinery. Given a complexity of vertebrate genome, splicing factors have evolved various mechanisms to accomplish such goal.

In this dissertation, we aimed to characterize how a splicing factor PTBP1 impose its repressive effects when it binds to distal sites flanking an alternative exon. Using a modified version of an alternative N1 exon as a model, we found that the process of exon definition is disrupted when PTBP1 is bound, resulting in a failure to recruit U2 snRNP to the exon. Quantitative characterization of exon-bound proteins in the presence or absence of PTBP1 using SILAC-MS, revealed that PTBP1 can modulate proteome of the exon. We also showed that PTBP1 cooperates with other proteins during the repression. This resulted in blockage of two RNA helicases, DDX5 and DDX17, which function to promote exon definition. It is likely that the failure of the PTBP1-bound exon to recruit U2 snRNP, in part, stems from less association of DDX5 and DDX17 RNA helicases.

From the mass spectrometry analyses we also uncovered dynamic assembly of

the exon definition complex. We found that hnRNP proteins, in general, are removed during the assembly. This coincides with a significant increase in binding of SR protein to the N1 exon. It is likely that the ATP-dependent RNA helicases DDX5 and DDX17, which we identified being associated with the exon, may facilitate these processes, because they have the potential ability to remodel the exon by removing hnRNP proteins and thus promoting binding of SR proteins. These ATP-dependent remodeling events might represent the first ATP reactions before U2 snRNP binding. Still, the contribution of the two RNA helicases in these remodeling processes needs to be determined.

An increase in binding of SR proteins during the process of exon definition is intriguing. It points toward unappreciated interactions inside the exon definition complex. These interactions are likely to be important for exon definition as we demonstrated that having U1 snRNP and U2AF65 alone are not sufficient to recruit U2 snRNP to the exon. Future experiments are needed to identify those interactions. This could be done by applying protein-protein crosslinkers on the purified EDC and analyzed cross-linked peptides by mass spectrometry.

Interestingly, we found in Chapter three that stem-loop 4 of U1 snRNA is very important for U1 snRNP to interact with U2 snRNP on intron-defined pre-mRNA; thus it is important for pre-mRNA splicing. However, some intriguing questions to pursue further are 1) whether or not stem-loop 4 is important for U1 and U2 snRNPs to interact across an exon and 2) if so, whether or not this interaction is one of the interactions disrupted by PTBP1. If the stem-loop 4 and U2 snRNP interacted across an exon, this interaction may be an essential toggle switch that controls the transition from exon

definition to intron definition during pre-spliceosomal complex assembly.

As exon definition is the major pathway for recognizing splice sites in vertebrates, it is becoming clear that an ultimate goal of splicing repression is to disrupt this process. Given divergent sequences of splice sites, exonic and intronic sequences in vertebrates, every exon is expected to have a different splice site strength, sets of SR proteins, hnRNP proteins and other RNA binding factors. Employing a quantitative proteomic approach on various model alternative exons under normal and repressive conditions will lead to an understanding of the steps during assembly of the EDC, revealing factors necessary for the process or factors needed for repression. We showed in Chapter two how this approach could be used to gain insights into complex assembly of the EDC and the repression mechanism by PTBP1. With an ensemble of data from different systems, we may be able to reveal a recurring principle of how splicing repressors accomplish their function.

BIBLIOGRAPHY

1. Jurica M, Moore M. Pre-mRNA Splicing:: Awash in a Sea of Proteins. Molecular cell. 2003.

2. Wahl M, Will C, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009.

3. Galej WP, Oubridge C, Newman AJ, Nagai K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature. 2013 Jan 23.

4. Fica SM, Tuttle N, Novak T, Li N-S, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA. RNA catalyses nuclear pre-mRNA splicing. Nature. 2013 Nov 14;503(7475):229–34.

5. Izquierdo JM, Majós N, Bonnal S, Martínez C, Castelo R, Guigó R, Bilbao D, Valcárcel J. Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. Molecular cell. 2005 Aug 19;19(4):475–84.

6. House AE, Lynch KW. An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. Nat Struct Mol Biol. 2006 Oct;13(10):937–44.

7. Sharma S, Falick AM, Black DL. Polypyrimidine Tract Binding Protein Blocks the 5′ Splice Site-Dependent Assembly of U2AF and the Prespliceosomal E Complex. Molecular cell. 2005 Aug;19(4):485–96. PMCID: PMC1635971

8. Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. Nat Struct Mol Biol. 2008 Feb;15(2):183–91. PMCID: PMC2546704

9. Kent OA, MacMillan AM. Early organization of pre-mRNA during spliceosome assembly. Nat Struct Biol. 2002 Aug 1;9(8):576–81.

10. Dönmez G, Hartmuth K, Kastner B, Will CL, Lührmann R. The 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. Molecular cell. 2007 Feb 9;25(3):399–411.

11. Wu JY, Maniatis T. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell [Internet]. 1993 Dec;75(6):1061–70. Retrieved from: http://www.sciencedirect.com/science/article/pii/009286749390316I#

12. Zuo P, Maniatis T. The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes &

Development. 1996 Jun 1;10(11):1356–68.

13. Kohtz JD, Jamison SF, Will CL, Zuo P, Lührmann R, Garcia-Blanco MA, Manley JL. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature. 1994 Mar 10;368(6467):119–24.

14. Michaud S, Reed R. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes & Development. 1991 Dec;5(12B):2534–46.

15. Zamore PD, Patton JG, Green MR. Cloning and domain structure of the mammalian splicing factor U2AF. Nature. 1992 Feb 13;355(6361):609–14.

16. Gozani O, Potashkin J, Reed R. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. Mol Cell Biol. 1998 Aug;18(8):4752–60. PMCID: PMC109061

17. Kistler AL, Guthrie C. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. Genes & Development. 2001 Jan 1;15(1):42–9. PMCID: PMC312603

18. Fleckner J, Zhang M, Valcárcel J, Green MR. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. Genes & Development. 1997 Jul 15;11(14):1864–72.

19. O'Day CL, Dalbadie-McFarland G, Abelson J. The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. J Biol Chem. 1996 Dec 27;271(52):33261–7.

20. Jarmoskaite I, Russell R. RNA helicase proteins as chaperones and remodelers. Annu Rev Biochem. 2014;83:697–725. PMCID: PMC4143424

21. Lim SR, Hertel KJ. Commitment to splice site pairing coincides with A complex formation. Molecular cell. 2004 Aug 13;15(3):477–83.

22. Xu Y-Z, Newnham CM, Kameoka S, Huang T, Konarska MM, Query CC. Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. EMBO J. 2004 Jan 28;23(2):376–85.

23. Sharma S, Wongpalee SP, Vashisht A, Wohlschlegel JA, Black DL. Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. Genes & Development. 2014 Nov 15;28(22):2518–31. PMCID: PMC4233244

24. Talerico M, Berget SM. Intron definition in splicing of small Drosophila introns. Mol Cell Biol. 1994 May;14(5):3434–45. PMCID: PMC358708

25. Sakharkar MK, Perumal BS, Sakharkar KR, Kangueane P. An analysis on gene

architecture in human and mouse genomes. In Silico Biol. (Gedrukt). 2005;5(4):347–65.

26. Berget SM. Exon Recognition in Vertebrate Splicing. Journal of Biological Chemistry. 1995 Feb 10;270(6):2411–4.

27. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nature Reviews Molecular Cell Biology. 2009 Nov;10(11):741–54. PMCID: PMC2958924

28. Romfo CM, Alvarez CJ, van Heeckeren WJ, Webb CJ, Wise JA. Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe. Mol Cell Biol. 2000 Nov;20(21):7955–70. PMCID: PMC86406

29. Robberson BL, Cote GJ, Berget SM. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol Cell Biol. 1990 Jan;10(1):84–94. PMCID: PMC360715

30. Schneider M, Will CL, Anokhina M, Tazi J, Urlaub H, Lührmann R. Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. Molecular cell. 2010 Apr 23;38(2):223–35.

31. Hoffman BE, Grabowski PJ. U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. Genes & Development. 1992 Dec;6(12B):2554–68.

32. Bourgeois CF, Popielarz M, Hildwein G, Stevenin J. Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5“ or 3” splice site activation. Mol Cell Biol. 1999 Nov;19(11):7347–56. PMCID: PMC84728

33. Furuyama S, Bruzik JP. Multiple roles for SR proteins in trans splicing. Mol Cell Biol. 2002 Aug;22(15):5337–46. PMCID: PMC133944

34. De Conti L, Baralle M, Buratti E. Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip Rev RNA. 2012 Oct 8.

35. Martinez-Contreras R, Cloutier P, Shkreta L, Fisette J-F, Revil T, Chabot B. hnRNP proteins and splicing control. Adv. Exp. Med. Biol. 2007;623:123–47.

36. Izquierdo JM. Hu antigen R (HuR) functions as an alternative pre-mRNA splicing regulator of Fas apoptosis-promoting receptor on exon definition. J Biol Chem. 2008 Jul 4;283(27):19077–84.

37. Bonnal S, Martínez C, Förch P, Bachi A, Wilm M, Valcárcel J. RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. Molecular cell. 2008 Oct 10;32(1):81–95.

38. Sanford JR, Gray NK, Beckmann K, Cáceres JF. A novel role for shuttling SR proteins in mRNA translation. Genes & Development. 2004 Apr 1;18(7):755–68. PMCID: PMC387416

39. Huang Y, Gattoni R, Stévenin J, Steitz JA. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. Molecular cell. 2003 Mar;11(3):837–43.

40. Zhang Z, Krainer AR. Involvement of SR proteins in mRNA surveillance. Molecular cell. 2004 Nov 19;16(4):597–607.

41. Pandit S, Zhou Y, Shiue L, Coutinho-Mansfield G, Li H, Qiu J, Huang J, Yeo GW, Ares M, Fu X-D. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. Molecular cell. 2013 Apr 25;50(2):223–35. PMCID: PMC3640356

42. Busch A, Hertel KJ. Evolution of SR protein and hnRNP splicing regulatory factors. Wiley Interdiscip Rev RNA. 2012 Jan;3(1):1–12. PMCID: PMC3235224

43. Wang J, Smith PJ, Krainer AR, Zhang MQ. Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. Nucleic Acids Res. 2005;33(16):5053–62. PMCID: PMC1201331

44. Fu XD, Maniatis T. The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. Proc Natl Acad Sci USA. 1992 Mar 1;89(5):1725–9. PMCID: PMC48525

45. Feng Y, Chen M, Manley JL. Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. Nat Struct Mol Biol. 2008 Oct;15(10):1040–8. PMCID: PMC2668916

46. Staknis D, Reed R. SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol Cell Biol. 1994 Nov;14(11):7670–82. PMCID: PMC359303

47. Shen H, Green MR. RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. Genes & Development. 2006 Jul 1;20(13):1755–65. PMCID: PMC1522072

48. Choi YD, Dreyfuss G. Isolation of the heterogeneous nuclear RNA-ribonucleoprotein complex (hnRNP): a unique supramolecular assembly. Proc Natl Acad Sci USA. 1984 Dec;81(23):7471–5. PMCID: PMC392168

49. Chou MY, Rooke N, Turck CW, Black DL. hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. Mol Cell Biol. 1999 Jan;19(1):69–77. PMCID: PMC83866

50. Dreyfuss G, Matunis MJ, Piñol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem. 1993;62:289–321.

51. Rooke N, Markovtsov V, Cagavi E, Black DL. Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1. Mol Cell Biol. 2003 Mar;23(6):1874–84. PMCID: PMC149473

52. Okunola HL, Krainer AR. Cooperative-binding and splicing-repressive properties of hnRNP A1. Mol Cell Biol. 2009 Oct;29(20):5620–31. PMCID: PMC2756886

53. Kashima T, Manley JL. A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. Nature genetics. 2003 Aug;34(4):460–3.

54. Zhou H-L, Lou H. Repression of prespliceosome complex formation at two distinct steps by Fox-1/Fox-2 proteins. Mol Cell Biol. 2008 Sep;28(17):5507–16. PMCID: PMC2519742

55. Crawford JB, Patton JG. Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. Mol Cell Biol. 2006 Dec;26(23):8791–802. PMCID: PMC1636816

56. Expert-Bezançon A, Sureau A, Durosay P, Salesse R, Groeneveld H, Lecaer JP, Marie J. hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B. J Biol Chem. 2004 Sep 10;279(37):38249–59.

57. Keppetipola N, Sharma S, Li Q, Black DL. Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. Crit. Rev. Biochem. Mol. Biol. 2012 Jul;47(4):360–78. PMCID: PMC3422667

58. Garcia-Blanco MA, Jamison SF, Sharp PA. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. Genes & Development. 1989 Dec;3(12A):1874–86.

59. Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmady B, Pitsch S, Black DL, Allain FH-T. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. Science. 2005 Sep 23;309(5743):2054–7.

60. Spellman R, Smith CWJ. Novel modes of splicing repression by PTB. Trends Biochem Sci. 2006 Feb 1;31(2):73–6.

61. Lamichhane R, Daubner GM, Thomas-Crusells J, Auweter SD, Manatschal C, Austin KS, Valniuk O, Allain FH-T, Rueda D. RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. Proc Natl Acad Sci USA. 2010 Mar 2;107(9):4105–10.

62. Xue Y, Ouyang K, Huang J, Zhou Y, Ouyang H, Li H, Wang G, Wu Q, Wei C, Bi

Y, Jiang L, Cai Z, Sun H, Zhang K, Zhang Y, Chen J, Fu X-D. Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. Cell. 2013 Jan 17;152(1-2):82–96. PMCID: PMC3552026

63. Spellman R, Llorian M, Smith CWJ. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. Molecular cell. 2007 Aug 3;27(3):420–34. PMCID: PMC1940037

64. Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon Y-S, Zhang C, Yeo G, Black DL, Sun H, Fu X-D, Zhang Y. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. Molecular cell. 2009 Dec 25;36(6):996–1006. PMCID: PMC2807993

65. Llorian M, Schwartz S, Clark TA, Hollander D, Tan L-Y, Spellman R, Gordon A, Schweitzer AC, la Grange de P, Ast G, Smith CWJ. Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. Nat Struct Mol Biol. 2010 Sep;17(9):1114–23. PMCID: PMC2933513

66. Singh R, Valcárcel J, Green MR. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. Science. 1995 May 26;268(5214):1173–6.

67. Saulière J, Sureau A, Expert-Bezançon A, Marie J. The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. Mol Cell Biol. 2006 Dec;26(23):8755–69. PMCID: PMC1636812

68. Matlin AJ, Southby J, Gooding C, Smith CWJ. Repression of alpha-actinin SM exon splicing by assisted binding of PTB to the polypyrimidine tract. RNA. 2007 Aug;13(8):1214–23. PMCID: PMC1924891

69. Shen H, Green MR. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. Molecular cell. 2004 Nov 5;16(3):363–73.

70. Bielli P, Bordi M, Biasio VD, Sette C. Regulation of BCL-X splicing reveals a role for the polypyrimidine tract binding protein (PTBP1/hnRNP I) in alternative 5' splice site selection. Nucleic Acids Res. 2014 Oct 7.

71. Rideau AP, Gooding C, Simpson PJ, Monie TP, Lorenz M, Hüttelmaier S, Singer RH, Matthews S, Curry S, Smith CWJ. A peptide motif in Raver1 mediates splicing repression by interaction with the PTB RRM2 domain. Nat Struct Mol Biol. 2006 Sep;13(9):839–48.

72. Joshi A, Coelho MB, Kotik-Kogan O, Simpson PJ, Matthews SJ, Smith CWJ, Curry S. Crystallographic analysis of polypyrimidine tract-binding protein-raver1 interactions involved in regulation of alternative splicing. Structure. 2011 Dec

7;19(12):1816–25.

73. Gromak N, Rideau A, Southby J, Scadden ADJ, Gooding C, Hüttelmaier S, Singer RH, Smith CWJ. The PTB interacting protein raver1 regulates alpha-tropomyosin alternative splicing. EMBO J. 2003 Dec 1;22(23):6356–64. PMCID: PMC291850

74. Gooding C, Edge C, Lorenz M, Coelho MB, Winters M, Kaminski CF, Cherny D, Eperon IC, Smith CWJ. MBNL1 and PTB cooperate to repress splicing of Tpm1 exon 3. Nucleic Acids Res. 2013 May;41(9):4765–82. PMCID: PMC3643581

75. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature genetics. 2008 Dec;40(12):1413–5.

76. Black DL. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem. 2003;72:291–336.

77. Venables JP, Klinck R, Koh C, Gervais-Bird J, Bramard A, Inkel L, Durand M, Couture S, Froehlich U, Lapointe E, Lucier J-F, Thibault P, Rancourt C, Tremblay K, Prinos P, Chabot B, Elela SA. Cancer-associated regulation of alternative splicing. Nat Struct Mol Biol. 2009 Jun;16(6):670–6.

78. David CJ, Manley JL. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes & Development. 2010 Nov 1;24(21):2343–64. PMCID: PMC2964746

79. Dreyfuss G, Kim VN, Kataoka N. Messenger-RNA-binding proteins and the messages they carry. Nature Reviews Molecular Cell Biology. 2002 Mar;3(3):195–205.

80. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nature Reviews Genetics. 2010 May;11(5):345–55.

81. Boukis LA, Liu N, Furuyama S, Bruzik JP. Ser/Arg-rich protein-mediated communication between U1 and U2 small nuclear ribonucleoprotein particles. J Biol Chem. 2004;279(28):29647–53.

82. Zheng S, Gray EE, Chawla G, Porse BT, O'Dell TJ, Black DL. PSD-95 is post-transcriptionally repressed during early neural development by PTBP1 and PTBP2. Nat. Neurosci. 2012 Mar;15(3):381–8–S1. PMCID: PMC3288398

83. Auweter S, Allain F. Structure-function relationships of the polypyrimidine tract binding protein. Cellular and molecular life sciences. 2008.

84. Wagner EJ, Garcia-Blanco MA. Polypyrimidine tract binding protein antagonizes exon definition. Mol Cell Biol. 2001 May 1;21(10):3281–8.

85. Black D. Activation of c-src neuron-specific splicing by an unusual RNA element in vivo and in vitro. Cell. 1992.

86. Carey MF, Peterson CL, Smale ST. The primer extension assay. Cold Spring Harb Protoc. 2013 Feb;2013(2):164–73.

87. Black DL, Chabot B, Steitz JA. U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. Cell. 1985 Oct;42(3):737–50.

88. Merendino L, Guth S, Bilbao D, Martínez C, Valcárcel J. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. Nature. 1999 Dec 16;402(6763):838–41.

89. Dignam JD. Preparation of extracts from higher eukaryotes. Meth Enzymol. 1990;182:194–203.

90. Kaiser P, Wohlschlegel J. Identification of ubiquitination sites and determination of ubiquitin-chain architectures by mass spectrometry. Meth Enzymol. 2005;399:266–77.

91. Wohlschlegel JA. Identification of SUMO-conjugated proteins and their SUMO attachment sites using proteomic mass spectrometry. Methods Mol. Biol. 2009;497:33–49.

92. Tabb DL, McDonald WH, Yates JR. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J. Proteome Res. 2002 Jan;1(1):21–6. PMCID: PMC2811961

93. Xu T, Park SK, Venable JD, Wohlschlegel JA, Diedrich JK, Cociorva D, Lu B, Liao L, Hewel J, Han X, Wong CCL, Fonslow B, Delahunty C, Gao Y, Shah H, Yates JR. ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. J Proteomics. 2015 Jul 11.

94. Elias JE, Gygi SP. Target-decoy search strategy for mass spectrometry-based proteomics. Methods Mol. Biol. 2010;604:55–71. PMCID: PMC2922680

95. Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, Workman JL, Washburn MP. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. Methods. 2006 Dec;40(4):303–11. PMCID: PMC1815300

96. Park SK, Venable JD, Xu T, Yates JR. A quantitative analysis software tool for mass spectrometry-based proteomics. Nat Methods. 2008 Apr;5(4):319–22. PMCID: PMC3509211

97. Park SK, Yates JR. Census for proteome quantification. Curr Protoc Bioinformatics. 2010 Mar;Chapter 13:Unit13.12.1–11.

98. Chen H, Li Y, Zhang J, Ran Y, Wei J, Yang Y, Shu H-B. RAVER1 is a coactivator of MDA5-mediated cellular antiviral response. Journal of Molecular Cell Biology. 2013 Apr;5(2):111–9.

99. Ho TH, Charlet-B N, Poulos MG, Singh G, Swanson MS, Cooper TA. Muscleblind proteins regulate alternative splicing. EMBO J. 2004 Aug 4;23(15):3103–12. PMCID: PMC514918

100. Dardenne E, Pierredon S, Driouch K, Gratadou L, Lacroix-Triki M, Espinoza MP, Zonta E, Germann S, Mortada H, Villemin J-P, Dutertre M, Lidereau R, Vagner S, Auboeuf D. Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness. Nat Struct Mol Biol. 2012 Nov;19(11):1139–46.

101. Jurica MS, Moore MJ. Capturing splicing complexes to study structure and mechanism. Methods. 2002 Nov;28(3):336–45.

102. Amir-Ahmady B, Boutz PL, Markovtsov V, Phillips ML, Black DL. Exon repression by polypyrimidine tract binding protein. RNA. 2005 May;11(5):699–716. PMCID: PMC1370756

103. Keenan S, Lewis PA, Wetherill SJ, Dunning CJR, Evans GJO. The N2-Src neuronal splice variant of C-Src has altered SH3 domain ligand specificity and a higher constitutive activity than N1-Src. FEBS letters. 2015 Jul 8;589(15):1995–2000.

104. Chan RC, Black DL. The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. Mol Cell Biol. 1997 Aug 1;17(8):4667–76.

105. Chou MY, Underwood JG, Nikolic J, Luu MH, Black DL. Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. Molecular cell. 2000 Jun;5(6):949–57.

106. Nilsen TW. Detecting RNA-RNA interactions using psoralen derivatives. Cold Spring Harb Protoc. 2014 Sep;2014(9):996–1000.

107. Rahman MA, Masuda A, Ohe K, Ito M, Hutchinson DO, Mayeda A, Engel AG, Ohno K. HnRNP L and hnRNP LL antagonistically modulate PTB-mediated splicing suppression of CHRNA1 pre-mRNA. Sci Rep. 2013;3:2931. PMCID: PMC3796306

108. Martins de Araújo M, Bonnal S, Hastings ML, Krainer AR, Valcárcel J. Differential 3' splice site recognition of SMN1 and SMN2 transcripts by U2AF and U2 snRNP. RNA. 2009 Apr;15(4):515–23. PMCID: PMC2661831

109. Arias MA, Lubkin A, Chasin LA. Splicing of designer exons informs a biophysical model for exon definition. RNA. 2014 Dec 9.

110. Sharma S, Maris C, Allain FH-T, Black DL. U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression. Molecular cell. 2011 Mar 4;41(5):579–88. PMCID: PMC3931528

111. Michaud S, Reed R. A functional association between the 5“ and 3” splice site is established in the earliest prespliceosome complex (E) in mammals. Genes & Development [Internet]. 1993 Jun;7(6):1008–20. Retrieved from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=8504926&retmode=ref&cmd=prlinks

112. Zhu J, Mayeda A, Krainer AR. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Molecular cell. 2001 Dec;8(6):1351–61.

113. Singh G, Pratt G, Yeo GW, Moore MJ. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu Rev Biochem. 2015 Jun 2;84:325–54.

114. Hartmuth K, Urlaub H, Vornlocher H-P, Will CL, Gentzel M, Wilm M, Lührmann R. Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc Natl Acad Sci USA. 2002 Dec 24;99(26):16719–24. PMCID: PMC139210

115. Kar A, Fushimi K, Zhou X, Ray P, Shi C, Chen X, Liu Z, Chen S, Wu JY. RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. Mol Cell Biol. 2011 May;31(9):1812–21. PMCID: PMC3133221

116. Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY. Systematic discovery of Xist RNA binding proteins. Cell. 2015 Apr 9;161(2):404–16. PMCID: PMC4425988

117. Camats M, Guil S, Kokolo M, Bach-Elias M. P68 RNA helicase (DDX5) alters activity of cis- and trans-acting factors of the alternative splicing of H-Ras. PloS one. 2008;3(8):e2926. PMCID: PMC2491553

118. Guil S, Gattoni R, Carrascal M, Abián J, Stévenin J, Bach-Elias M. Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. Mol Cell Biol. 2003 Apr;23(8):2927–41. PMCID: PMC152554

119. Dardenne E, Polay Espinoza M, Fattet L, Germann S, Lambert M-P, Neil H, Zonta E, Mortada H, Gratadou L, Deygas M, Chakrama FZ, Samaan S, Desmet F-O, Tranchevent L-C, Dutertre M, Rimokh R, Bourgeois CF, Auboeuf D. RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. Cell Rep. 2014 Jun 26;7(6):1900–13.

120. Will CL, Lührmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol. 2011 Jul;3(7). PMCID: PMC3119917

121. Behzadnia N, Hartmuth K, Will CL, Lührmann R. Functional spliceosomal A complexes can be assembled in vitro in the absence of a penta-snRNP. RNA. 2006 Sep;12(9):1738–46. PMCID: PMC1557700

122. Hoskins A, Friedman L, Gallagher S. Ordered and Dynamic Assembly of Single Spliceosomes. Science. 2011.

123. Das R, Zhou Z, Reed R. Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. Molecular cell. 2000 May;5(5):779–87.

124. Cordin O, Hahn D, Beggs JD. Structure, function and regulation of spliceosomal RNA helicases. Curr Opin Cell Biol. 2012 Jun;24(3):431–8.

125. Semlow DR, Staley JP. Staying on message: ensuring fidelity in pre-mRNA splicing. Trends Biochem Sci. 2012 Jul;37(7):263–73. PMCID: PMC3735133

126. Fox-Walsh KL, Hertel KJ. Splice-site pairing is an intrinsically high fidelity process. Proc Natl Acad Sci USA. 2009 Feb 10;106(6):1766–71. PMCID: PMC2644112

127. Shao W, Kim H-S, Cao Y, Xu Y-Z, Query CC. A U1-U2 snRNP interaction network during intron definition. Mol Cell Biol. 2012 Jan;32(2):470–8. PMCID: PMC3255776

128. Xu Y-Z, Query CC. Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. Molecular cell. 2007 Dec 14;28(5):838–49. PMCID: PMC2246091

129. Abovich N, Rosbash M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell. 1997 May 2;89(3):403–12.

130. Reed R. Mechanisms of fidelity in pre-mRNA splicing. Curr Opin Cell Biol. 2000 Jun;12(3):340–5.

131. Eldridge AG, Li Y, Sharp PA, Blencowe BJ. The SRm160/300 splicing coactivator is required for exon-enhancer function. Proc Natl Acad Sci USA. 1999 May 25;96(11):6125–30. PMCID: PMC26846

132. Blencowe BJ, Baurén G, Eldridge AG, Issner R, Nickerson JA, Rosonina E, Sharp PA. The SRm160/300 splicing coactivator subunits. RNA. 2000 Jan;6(1):111–20. PMCID: PMC1369899

133. Libri D, Graziani N, Saguez C, Boulay J. Multiple roles for the yeast SUB2/yUAP56 gene in splicing. Genes & Development. 2001 Jan 1;15(1):36–41. PMCID: PMC312604

134. Shen H, Zheng X, Shen J, Zhang L, Zhao R, Green MR. Distinct activities of the

DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. Genes & Development. 2008 Jul 1;22(13):1796–803. PMCID: PMC2492666

135. Schwer B, Chang J, Shuman S. Structure-function analysis of the 5' end of yeast U1 snRNA highlights genetic interactions with the Msl5*Mud2 branchpoint-binding complex and other spliceosome assembly factors. Nucleic Acids Res. 2013 Aug;41(15):7485–500. PMCID: PMC3753624

136. Dominski Z, Kole R. Selection of splice sites in pre-mRNAs with short internal exons. Mol Cell Biol. 1991 Dec;11(12):6075–83. PMCID: PMC361780

137. Modafferi EF, Black DL. A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon. Mol Cell Biol. 1997 Nov;17(11):6537–45. PMCID: PMC232507

138. Le Guédard-Méreuze S, Vaché C, Molinari N, Vaudaine J, Claustres M, Roux A-F, Tuffery-Giraud S. Sequence contexts that determine the pathogenicity of base substitutions at position +3 of donor splice-sites. Hum. Mutat. 2009 Sep;30(9):1329–39.

139. Solnick D. Alternative splicing caused by RNA secondary structure. Cell. 1985.

140. Markovtsov V, Nikolic J, Goldman J, Turck C, Chou M, Black D. Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. Mol Cell Biol. 2000;20(20):7463.

141. Das BK, Xia L, Palandjian L, Gozani O, Chyung Y, Reed R. Characterization of a protein complex containing spliceosomal proteins SAPs 49, 130, 145, and 155. Mol Cell Biol. 1999 Oct;19(10):6796–802. PMCID: PMC84676

142. Will CL, Schneider C, MacMillan AM, Katopodis NF, Neubauer G, Wilm M, Lührmann R, Query CC. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. EMBO J. 2001 Aug 15;20(16):4536–46. PMCID: PMC125580

143. Tuma RS, Beaudet MP, Jin X, Jones LJ, Cheung CY, Yue S, Singer VL. Characterization of SYBR Gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. Anal. Biochem. 1999 Mar 15;268(2):278–88.

144. Zhuang Y, Weiner AM. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell. 1986 Sep 12;46(6):827–35.

145. Roca X, Krainer AR. A splicing component adapted to gene silencing. Nature biotechnology. 2009 Mar;:250–1.

146. Roca X, Akerman M, Gaus H, Berdeja A, Bennett CF, Krainer AR. Widespread

recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. Genes & Development. 2012 May 15;26(10):1098–109. PMCID: PMC3360564

147. Hall KB, Konarska MM. The 5' splice site consensus RNA oligonucleotide induces assembly of U2/U4/U5/U6 small nuclear ribonucleoprotein complexes. Proc Natl Acad Sci USA. 1992 Nov 15;89(22):10969–73. PMCID: PMC50464

148. Butter F, Scheibe M, Mörl M, Mann M. Unbiased RNA–protein interaction screen by quantitative proteomics. Proceedings of the …. 2009.

149. Lee N, Pimienta G, Steitz JA. AUF1/hnRNP D is a novel protein partner of the EBER1 noncoding RNA of Epstein-Barr virus. RNA. 2012 Nov;18(11):2073–82. PMCID: PMC3479396

150. Ibarrola N, Kalume DE, Grønborg M, Iwahori A, Pandey A. A proteomic approach for quantitation of phosphorylation using stable isotope labeling in cell culture. Analytical Chemistry. 2003 Nov 15;75(22):6043–9.

151. Mann M. Functional and quantitative proteomics using SILAC. Nature Reviews Molecular Cell Biology. 2006 Dec;7(12):952–8.

152. Behrens SE, Tyc K, Kastner B, Reichelt J, Lührmann R. Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. Mol Cell Biol. 1993 Jan;13(1):307–19. PMCID: PMC358910

153. Brosi R, Gröning K, Behrens SE, Lührmann R, Krämer A. Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. Science. 1993 Oct 1;262(5130):102–5.

154. Dybkov O, Will CL, Deckert J, Behzadnia N, Hartmuth K, Lührmann R. U2 snRNA-protein contacts in purified human 17S U2 snRNPs and in spliceosomal A and B complexes. Mol Cell Biol. 2006 Apr;26(7):2803–16. PMCID: PMC1430325

155. Hernandez N. Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. EMBO J. 1985 Jul;4(7):1827–37. PMCID: PMC554424

156. Ach RA, Weiner AM. The highly conserved U small nuclear RNA 3'-end formation signal is quite tolerant to mutation. Mol Cell Biol. 1987 Jun;7(6):2070–9. PMCID: PMC365327

157. Pellizzoni L, Yong J, Dreyfuss G. Essential role for the SMN complex in the specificity of snRNP assembly. Science. 2002 Nov 29;298(5599):1775–9.

158. Yong J, Pellizzoni L, Dreyfuss G. Sequence-specific interaction of U1 snRNA

with the SMN complex. EMBO J. 2002 Mar 1;21(5):1188–96. PMCID: PMC125911

159. Chen J, Wagner EJ. snRNA 3' end formation: the dawn of the Integrator complex. Biochem Soc Trans. 2010 Aug;38(4):1082–7. PMCID: PMC3969742

160. Stark H, Dube P, Lührmann R, Kastner B. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature. 2001 Jan 25;409(6819):539–42.

161. Pomeranz K, Oubridge C, Leung A, Li J, Nagai K. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature. 2009.

162. Weber G, Trowitzsch S, Kastner B, Lührmann R, Wahl MC. Functional organization of the Sm core in the crystal structure of human U1 snRNP. EMBO J. 2010 Dec 15;29(24):4172–84. PMCID: PMC3018796

163. McConnell TS, Lokken RP, Steitz JA. Assembly of the U1 snRNP involves interactions with the backbone of the terminal stem of U1 snRNA. RNA. 2003 Feb;9(2):193–201. PMCID: PMC1370385

164. Crawford DJ, Hoskins AA, Friedman LJ, Gelles J, Moore MJ. Single-molecule colocalization FRET evidence that spliceosome activation precedes stable approach of 5' splice site and branch site. Proc Natl Acad Sci USA. 2013 Apr 23;110(17):6783–8. PMCID: PMC3637735

165. Nesic D, Krämer A. Domains in human splicing factors SF3a60 and SF3a66 required for binding to SF3a120, assembly of the 17S U2 snRNP, and prespliceosome formation. Mol Cell Biol. 2001 Oct;21(19):6406–17. PMCID: PMC99788

166. Tanackovic G, Krämer A. Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo. Molecular biology of the cell. 2005 Mar;16(3):1366–77. PMCID: PMC551499

167. Chiara MD, Champion-Arnaud P, Buvoli M, Nadal-Ginard B, Reed R. Specific protein-protein interactions between the essential mammalian spliceosome-associated proteins SAP 61 and SAP 114. Proc Natl Acad Sci USA. 1994 Jul 5;91(14):6403–7. PMCID: PMC44210

168. Hilliker AK, Mefford MA, Staley JP. U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. Genes & Development. 2007 Apr 1;21(7):821–34. PMCID: PMC1838533

169. Perriman RJ, Ares M. Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. Genes & Development. 2007 Apr 1;21(7):811–20. PMCID: PMC1838532

170. Lin P-C, Xu R-M. Structure and assembly of the SF3a splicing factor complex of U2 snRNP. EMBO J. 2012 Mar 21;31(6):1579–90. PMCID: PMC3321192

171. Lardelli RM, Thompson JX, Yates JR, Stevens SW. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA. 2010 Mar;16(3):516–28. PMCID: PMC2822917

172. Anokhina M, Bessonov S, Miao Z, Westhof E, Hartmuth K, Lührmann R. RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. EMBO J. 2013 Oct 30;32(21):2804–18. PMCID: PMC3817461

173. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003 Jul 1;31(13):3406–15. PMCID: PMC169194