

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Regulation of Alternative Splicing in Drosophila Melanogaster

Permalink

<https://escholarship.org/uc/item/1cx6d5z9>

Author

Taliaferro, Jefferson Matthew

Publication Date

2012

Peer reviewed|Thesis/dissertation

Regulation of Alternative Splicing in *Drosophila melanogaster*

By

Jefferson Matthew Taliaferro

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Donald C. Rio, Chair

Professor Steven E. Brenner

Professor Kathleen Collins

Professor Britt A. Glaunsinger

Professor Lin He

Fall 2012

ABSTRACT

Regulation of Alternative Splicing in *Drosophila melanogaster*

by

Jefferson Matthew Taliaferro

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Donald C. Rio, Chair

The patterns and mechanisms by which eukaryotic cells regulate the expression of their genetic information are highly complex and intricate. The transmittance of this information from nuclear repository to cytoplasmic translation contains within it several steps, including the selective removal and concomitant joining of pieces of information in a process called alternative splicing. The projects detailed within this document describe the regulation of alternative splicing through the interaction of specific proteins with specific pre-mRNA transcripts.

The Rio lab has studied PSI, a protein involved in the regulation of the P element transposase transcript, for many years. It has since been shown to regulate the splicing of hundreds of other transcripts. The experiments described here look at the organization of PSI and other proteins on the P element transcript by site-specific labeling of the transcript using radioactive ^{32}P . We also investigate two phosphorylation events of PSI, identifying the kinases responsible and demonstrate that these events may change the protein-protein interaction partners of PSI.

It has become increasingly apparent that alternative splicing may not only be regulated by protein/RNA interactions, but also by RNA/RNA interactions. To probe this, we designed experiments to test if some well-known small RNA-associated proteins are regulating alternative splicing. Using splice junction microarrays, we determined that Argonaute-2 (Ago-2) regulated the splicing of over 100 splice junctions, and further experiments using ChIP-seq and mRNA-seq of Ago-2 mutants revealed that Ago-2 also has a role in transcriptional repression, possibly through being incorporated in complexes composed of polycomb-group genes. We also used CLIP-seq to determine the RNA binding profile and preferences of Ago-2 in *Drosophila* tissue culture cells.

Finally, we characterized the functions of a *Drosophila* specific splicing factor called LS2. LS2 is orthologous to the highly conserved splicing factor dU2AF50, but its origin through retroduplication and subsequent divergence to acquire distinct sequence specificity, expression pattern, and function show it to be an interesting case in the evolution of alternative splicing regulation. This may be a mechanism that underlies the

existence of some members of the large families of splicing factors, including hnRNP proteins and SR proteins. That is, by duplicating functional copies of genes, cellular systems create new proteins to tinker with and acquire new functions while keeping the former functionality and stability of the parent protein.

While these projects are essentially independent of each other, they all fall under the umbrella of protein regulation of RNA metabolism and hopefully contribute to a more complete understanding of the regulation of gene expression.

TABLE OF CONTENTS

Acknowledgements	iii
Chapter 1: Introduction	1
Chapter 2: Phosphorylation of PSI	27
Figure 1: Biochemical fractionation and analysis of the PSI kinases	40
Figure 2: Identification of phosphorylation sites by mass spectrometry	42
Figure 3: Biochemical purification of Drosophila casein kinase II	44
Figure 4: Purification of a second PSI kinase activity	46
Figure 5: Protein-protein interactions of PSI	48
Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression	50
Figure 1: RNAi knockdowns in Drosophila S2 cells	67
Figure 2: Analysis of Ago-2 in Drosophila cell culture	69
Figure 3: Analysis of splicing events affected in Ago-2 mutant Drosophila adults	71
Figure 4: mRNA-seq analysis of Ago-2 mutant strains	73
Figure 5: Chromatin immunoprecipitation-sequencing (ChIP-seq) analysis of Drosophila Ago-2	75
Figure 6: Chromatin association of Ago-2 in Drosophila cell culture	77
Figure 7: mRNA-seq analysis of Ago-2 mutant strains	79
Figure 8: Effects of Ago-2 mutations on gene expression	81
Figure 9: CLIP-seq of Ago-2 in Drosophila cell culture	83
Figure 10: CLIP-seq analysis of Drosophila Ago-2	85
Chapter 4: Evolution of a tissue specific splicing network	87
Figure 1: dU2AF ⁵⁰ /LS2 sequence alignment	113
Figure 2: LS2 arose and diverged in function from dU2AF ⁵⁰ in Drosophila	115
Figure 3: dU2AF ⁵⁰ /LS2 splice junction microarray results	117
Figure 4: LS2 has diverged in RNA binding sequence specificity from dU2AF ⁵⁰ and its binding motif is enriched in its target transcripts	119
Figure 5: RNA binding activity of LS2/dU2AF ³⁸ heterodimer	121
Figure 6: LS2 physically interacts with dU2AF ³⁸	123
Figure 7: LS2 and its target transcripts are enriched in testes and regulate testes-related functions	125
Figure 8: Gene ontology networks in LS2, dU2AF50 and dU2AF38 affected transcripts	127
Figure 9: LS2 inhibits splicing of the ftz intron in vitro	129
Figure 10: In vitro splicing assay using LS2 alone	131
Figure 11: In vitro splicing assay using titrated amounts of LS2 and splicing of the G-replacement RNA	133
Figure 12: LS2 acts as a splicing repressor in vivo, is enriched at specific positions in its target transcripts and binds its predicted targets	135

Figure 13: LS2 function is not constrained to splice sites and its affected introns are longer than expected	137
Figure 14: LS2 is expressed only in differentiated cells in the Drosophila testes	139
Figure 15: Recombinant LS2 purification using Poros HS	141
Figure 16: LS2 interacts with SF1	143
Figure 17: LS2 RNAi fly lines	145
Appendix A: Analysis of the P element splicing silencer complex by RNase protection	147
Figure 1: Protection of site-specifically labeled RNA with purified PSI and hrp48	153
Figure 2: Protection of site-specifically labeled RNA with Kc nuclear extract	155
Figure 3: A model for the repression of splicing at the P element third intron	157
Figure 4: Purification of recombinant PSI and hrp48	159

ACKNOWLEDGEMENTS

Of course, no undertaking like this is ever started or completed alone. There are many people that deserve sincere thanks.

First, I want to thank all the members of the Rio lab for their technical and emotional support. In particular, I want to thank Julie Aspden and Melissa Harrison. I feel like I learned more in four years of daily lunches with you guys than I would have in 10 more years of school. I also want to thank Angela Brooks. Thank you for your patience and availability during my assault on you with computational questions and emails. I also want to thank my undergrad Dhruv Marwha. You really were a pleasure to work with. The questions you asked were always well-formed and thought provoking, but even more than that, just the fact that you cared enough to ask those questions made having you around worth it. Also, the fact that you were an extremely capable protein biochemist and molecular biologist didn't hurt.

Thank you to the members of my committee. During and between our meetings, you kept me focused on the important questions and redirected me when I would go off on a less productive tangent.

Thank you to my aunt and uncle, Jerra and Randy. I can't tell you how nice it was to have a place to go where I could get away from the Bay for a weekend and relax, mentally and physically. You guys made California feel much more like home.

Thank you to Melanie. You kept me motivated after long days in the lab, and seeing a smiling face after those long days made getting up for the next one a lot easier. I can't wait to join you in Boston.

A big, big thanks to Don. In my view, you were the perfect mentor: engaged and present enough to be helpful and give direction, but also hands-off enough to let me explore things on my own. Your door was always open, metaphorically and usually literally. I knew that if I got stuck, I could come to you for a technical hint or a different approach to take. You were never too busy to take a look at a gel or turn around an edit on a paper. You taught me about how to design and conduct an experiment, but also about the culture of science and its people.

Finally, the biggest thanks goes to my family. You guys showed me why education was worthwhile. I could call you guys at any time of day, and sometimes did, and just hearing about what was going on back there was comforting. No matter what happened or where I ended up, I knew you guys would always be there if I needed you. You helped me move out here, and sometimes it feels like you never left, and that's a good thing. Dad and Juliana, I'll read yours if you read mine. Mom, I guess you have to read all three.

Chapter 1: Introduction

Introduction

Pre-messenger RNA (pre-mRNA) splicing is a central process in RNA metabolism. During this process, intervening sequences called introns are removed from the pre-mRNA transcript, and the remaining introns are joined, or spliced, together. This process requires the specific and dynamic interaction of a range of cellular proteins and RNA-protein complexes and is highly regulated such that many different combinations of exons can be produced, allowing the cell to respond to its environment. These activities give the cell great flexibility to change the informational output of the genome in response to a variety of stimuli, from common processes like the cell cycle (Moore et al. 2010) to complete developmental programs (Barberan-Soler and Zahler 2008). Aberrant pre-mRNA splicing has also been linked to many diseases (Pagani and Baralle 2004; Cooper et al. 2009).

Basics of splicing

Each important region in the nuclear precursor RNA (pre-mRNA) has specific yet somewhat degenerate sequences that facilitate intron recognition but which still allow for either subsequent intron splicing or for a high degree of regulation via alternative splicing patterns. The three key sites for recognition of an intron are the 5' splice site, the branchpoint, and 3' splice site. Although there is a high degree of degeneracy, the consensus sites for 5' splice sites and branchpoints are AG/GUAAG and YURAY, respectively. Between the branchpoint and the 3' splice site lies a long stretch of pyrimidine residues. The length of tract varies but is often between 10 and 30 nt. Both the length and strength, that is, the total number of pyrimidines and the number of consecutive pyrimidines, can affect the efficiency with which a given intron is spliced. The 3' splice site is information-poor by comparison and is defined by essentially only the terminal AG, although proximity to a well-defined polypyrimidine tract plays a major role in 3' splice site definition. Thus, the 3' splice site consensus sequence is a run of pyrimidine residues followed by the intron-terminal AG dinucleotide. Often, the first nucleotide of the downstream exon is a G.

Chemically, splicing proceeds as a pair of S_N2 nucleophilic attack phosphodiester bond transesterification reactions (Green 1986; Padgett et al. 1986; Maschhoff and Padgett 1993; Moore and Sharp 1993). In both reactions, the nucleophile is a hydroxyl group from the ribose ring of a nucleoside in the RNA chain that attacks a phosphodiester bond elsewhere in the pre-mRNA. First, the 2' hydroxyl of the branchpoint adenosine attacks the phosphodiester bond at the 5' splice site, liberating the 5' exon and resulting in an intermediate in which the intron, in the form of a lariat, remains attached to the 3' exon via an unusual 2'-5' phosphodiester bond. This reaction intermediate is resolved in the second transesterification reaction in which the newly liberated 3' hydroxyl in the 5' exon attacks the phosphodiester bond at the 3' splice site, yielding two joined exons and releasing the intron RNA as a lariat (Ruskin et al. 1984).

Spliceosome assembly and catalysis

Although the reactions described above are thought to be mostly catalyzed by RNA, efficient and accurate intron splicing requires the concerted action of a number of small nuclear ribonucleoproteins (snRNPs), most notably the abundant U1, U2, U4, U5, and U6 snRNPs (Black et al. 1985) which assemble into a large catalytic RNP termed the spliceosome. There are many snRNPs in the nucleus, but the majority of pre-mRNA introns (of the GT-AG type) are spliced through the action of U1, U2, U4, U5 and U6. Each of these snRNPs consists of a common core set of proteins called the Sm proteins, a short RNA with a trimethylated guanosine cap, as well as a few snRNP-specific proteins (Will and Luhrmann 2011).

Additionally, there is a second, minor spliceosome that, although it is functionally and mechanistically analogous to the major spliceosome, splices introns that have very tightly constrained 5' splice sites and branchpoint sequences and also lack polypyrimidine tracts. These so-called AT-AC introns are spliced by spliceosomes in which U1 and U2 snRNPs are replaced by U11 and U12 snRNPs, respectively, and U4 and U6 snRNPs are replaced by U4atac and U6atac snRNPs, respectively (Patel and Steitz 2003).

Spliceosomes assemble in a stepwise, ordered manner (figure 1). The large complexes formed at each step are stable enough to be studied *in vitro* and have been named, in order of assembly, as E, A, B, B* and C. The first snRNP to bind to a nascent pre-mRNA is U1 (Ruskin et al. 1984; Hoskins et al. 2011). The U1/pre-mRNA interaction constitutes the E complex. U1 is able to recognize 5' splice sites through base-pairing interactions between the 5' end of its snRNA and the 5' splice site consensus sequence and is able to bind 5' splice sites in an ATP-independent manner (Zhuang and Weiner 1986). As is the case for other spliceosome/pre-mRNA interactions, 5' splice sites that more closely match the consensus sequence and thus form more stable base pairing interactions are more tightly associated with U1 and therefore more likely to be committed to the splicing pathway (Mount 1982).

At the other end of the intron, the 3' splice site is first recognized by a trio of non-snRNP proteins. The polypyrimidine tract and 3' terminal AG dinucleotide are initially recognized by the heterodimeric U2 snRNP-associated factor (U2AF) (Ruskin et al. 1988). U2AF is highly conserved from fission yeast (*S. pombe*) to humans and contains an N-terminal RS dipeptide repeat domain and three RRM-type RNA binding domains, two of which contact the polypyrimidine tract and one which interacts with SF1. The large subunit of U2AF (U2AF⁶⁵ in humans, dU2AF⁵⁰ in *Drosophila*) recognizes the polypyrimidine tract (Singh et al. 1995). In any given intron, the length and base composition of the polypyrimidine tract influences the affinity of U2AF for the pre-mRNA and thus the efficiency with which an intron is spliced. The large and small subunits of

Chapter 1: Introduction

U2AF are held together by reciprocal “tongue-in-groove” interactions between a tryptophan residue on one subunit and a hydrophobic pocket on the other subunit (Zamore and Green 1989; Rudner et al. 1998b; Kielkopf et al. 2001). The U2AF small subunit (U2AF³⁵ in humans, dU2AF³⁸ in *Drosophila*) is then positioned to contact the terminal intron AG dinucleotide (MacMorris et al. 1999; Merendino et al. 1999; Wu et al. 1999). High-affinity binding of U2AF to the 3' splice site requires both U2AF subunits and the presence of at least one of the RS domains in one of the subunits (Rudner et al. 1998a), and in the case of U2AF⁶⁵, the RS domain may assist in binding to the branchpoint (Valcarcel et al. 1996). RS dipeptide repeat domains are commonly found in splicing factors and are so named for the large number of arginine and serine residues they contain. Because of the large positive charge contributed by the arginine residues, these domains are thought to be useful for promoting high affinity, but non-specific, protein-RNA interactions. RS domains have also been implicated in protein-protein interactions (Wu and Maniatis 1993).

Recognition of the 3' splice site is completed by the binding of SF1 to the branch point sequence (Berglund et al. 1998b). SF1 is a small single KH-domain RNA binding protein that directly binds the branchpoint and defines the region of the pre-mRNA around the adenosine that will be the nucleophile in the first transesterification reaction. SF1 is stabilized at the branch point through a largely electrostatic protein-protein interaction with a helix in the third RRM of the large subunit of U2AF (Berglund et al. 1998a; Selenko et al. 2003).

Following the complete recognition of the 3' splice site and branchpoint by U2AF and SF1, respectively, U2 snRNP is recruited to the branch point sequence in a reaction that requires ATP (Konarska and Sharp 1986). This first ATP-dependent step commits that intron to the splicing pathway. The U2 snRNP-branchpoint interaction involves extensive basepairing between U2 snRNA and the branch point sequence. Binding of U2 snRNP to the branchpoint leads to the displacement of SF1. The role of SF1, however, is performed by SF3b in its binding to the branchpoint and in its binding to RRM3 of the large subunit of U2AF (Gozani et al. 1998; Will et al. 2001). Upon binding of U2 snRNP to the branchpoint, the spliceosome has transitioned into what is known as the A complex (Figure 1).

Conceptually, there are then two ways in which U1 and U2 could interact to form a functional complex that would then continue spliceosome assembly. In a process called intron definition, the U1 snRNP complex bound to a 5' splice site can interact with a U2 snRNP bound to a downstream 3' splice site such that the intron is now bridged between the two snRNPs. Alternatively, the U1 snRNP could interact with a U2 snRNP bound at a 3' splice site that is upstream of where the U1-5' splice site complex is. This process now results in the spanning of an exon between the two snRNPs and is called exon definition.

Chapter 1: Introduction

Most metazoan transcripts have relatively long introns and relatively short exons, although this difference is much less pronounced in *Drosophila* transcripts. Thus, it may be simpler and cleaner to first define an exon from the sea of intron sequences in large pre-mRNAs rather than the reverse. Exon definition is therefore thought in metazoans to be the mode by which most splicing events are first organized (Robberson et al. 1990; Hoffman and Grabowski 1992; Reed 2000; Schneider et al. 2010). At some point in the reaction pathway, however, the chemistry of the splicing reactions requires a reorganization from exon definition to intron definition. This can be a key point for regulation, as has recently been shown for the splicing factor PTB, which for some transcripts inhibits the switch from an exon-defined complex to an intron-defined complex (Sharma et al. 2008).

After the pre-mRNA substrate has entered into the A complex, U5 and U4/U6 snRNPs join to form the B complex. The U4/U6 snRNP complex, however, is in a catalytically inactive state, with extensive basepairing between the U4 and U6 snRNAs (Bringmann et al. 1984). During activation of the spliceosome for catalysis, a substantial rearrangement of the initial U4-U6 RNA-RNA basepairing interactions occurs. U1 is removed from the pre-mRNA, and in its place, U6 basepairs with consensus sequences at the 5' splice site. Additional RNA-RNA rearrangements occur to break the basepairing between the U4 and U6 snRNPs. This allows U4 to be released and frees the U6 snRNA to adopt structures important for generating the catalytically active site of the spliceosome (Yean et al. 2000) and to make basepairing interactions with U2 snRNA (Hausner et al. 1990; Madhani and Guthrie 1992; Sun and Manley 1995). This complex is then known as B*, or the activated spliceosome. C complex is then formed after the first transesterification reaction. After the second catalytic step, U2, U5 and U6 are released in a complex with the intron lariat and the process can begin again (Figure 1). Although the protein components of the spliceosome are important for efficient catalysis, it remains the role of the snRNAs to perform the actual catalytic actions, as protein-free complexes of U2 and U6 snRNAs are active for splicing catalysis (Valadkhan and Manley 2001).

Chapter 1: Introduction

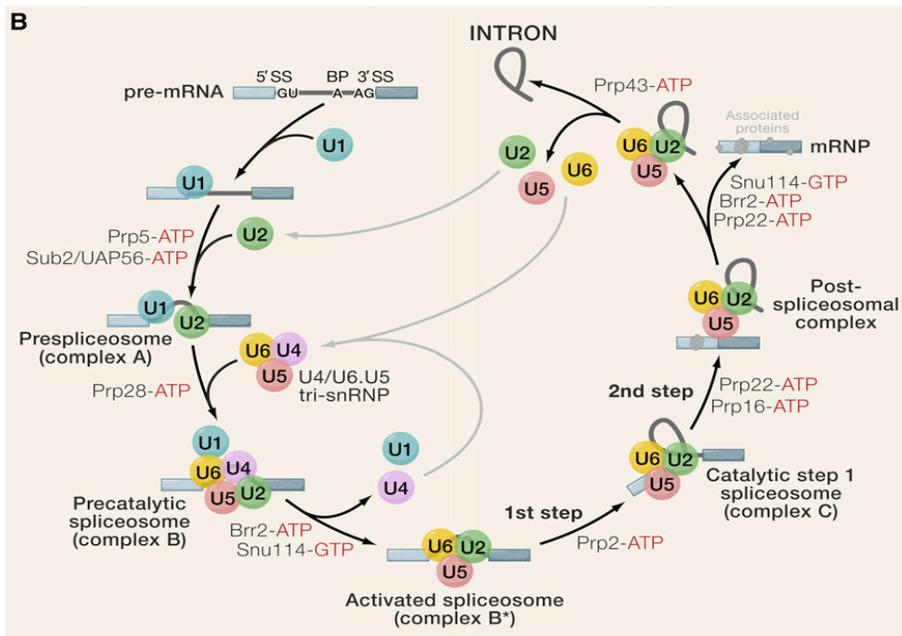


Figure 1. Spliceosome assembly and recycling. Wahl MC, Will CL, and Luhrmann R. (2009) *Cell*. **136**: 701-18.

RNA-binding proteins

The majority of RNA-binding proteins interact with RNA through the action of one of two types of RNA-binding domains: the RNA recognition motif (RRM, also known as the RNA-binding domain (RBD) and the RNP domain) and the hnRNP K homology (KH) domain (Perez-Canadillas and Varani 2001). Any given RNA-binding protein may have anywhere from one to several of these domains that may independently or cooperatively bind RNA (Glisovic et al. 2008). In proteins that contain multiple RNA-binding domains, these are often separated by short, flexible linkers that may become more ordered upon RNA binding to promote the formation of a stable relative interdomain orientation (Crowder et al. 1999; Inoue et al. 2000).

RRM domains, which are the most common type of RNA binding domain, have been analyzed structurally in complex with RNA ligands (Lee et al. 1994; Oubridge et al. 1994; Sickmier et al. 2006). In all cases, there is a similar protein secondary structure arrangement: beta-alpha-beta-beta-alpha-beta (b-a-b-b-a-b). These fold into a sheet of four beta strands supported in the back by the two alpha helices. Normally, RNA binding occurs by the draping of the RNA strand across the beta sheet, with sequence specific contacts made across the sheet and in the loops connecting the strands and helices (Sickmier et al. 2006). Many RRM domains can have intermolecular protein-protein interaction functions, as is the case for RRM3 in U2AF⁶⁵ and its interaction with SF1 (Selenko et al. 2003). Some RRM domains also have intramolecular protein-protein interaction functions, as is the case with RRM1 and RRM2 in U2AF⁶⁵ where the

Chapter 1: Introduction

presence or absence of this interaction modulates the entire protein's affinity for RNA (Mackereth et al. 2011).

Another common type of RNA binding domain is the hnRNP K-homology, or KH domain. Although there are two types of KH domains, they differ only in the linear arrangement of secondary structure elements and upon tertiary folding produce approximately the same organization of beta strands and alpha helices: A three-stranded beta sheet supported by three alpha helices (Perez-Canadillas and Varani 2001). Binding of RNA occurs in a cleft between two of these helices in a conserved GXXG motif (Lewis et al. 2000).

Double-stranded RNA binding domains (dsRBD), such as those in proteins like the Argonaute and Dicer families, often display little sequence specificity to their RNA targets. As with the other domains, the dsRBD structurally consists of a beta sheet supported in the back by two helices. However, from the structures known so far, most of the RNA binding occurs through loops between two strands in the sheet and with one of the helices (Ryter and Schultz 1998; Ramos et al. 2000).

Alternative Splicing

While almost all introns are spliced using the same machinery and mechanism, splicing some exons into the final mRNA or skipping of other exons allows the cell to generate many different mRNAs from a single pre-mRNA. This flexibility is more common in larger genomes, such as the human genome, which typically contain 3-4 alternatively spliced isoforms for a given gene (Wang et al. 2008). The splicing of one transcript in several different ways allows the generation of the vast proteomic diversity observed in metazoans from a comparatively small number of genes (Nilsen and Graveley 2010). These alternatively spliced transcripts are often restricted to particular cell types or tissues and encode proteins that are critical to proper cell or tissue function (Xu et al. 2002; Wang et al. 2008). In general, the complexity of alternative splicing seems to increase with organismal complexity from yeast to *Drosophila* to mammals. Alternative splicing may therefore be a key way in which higher organisms are able to meet the demands of many different cell types, developmental programs or stresses and facilitate the development of highly elaborate gene expression programs while maintaining a relatively similarly sized genome.

The extent to which alternative splicing is used in mammalian genes is quite large, with current estimates at over 95% of all transcripts in the human genome (Wang et al. 2008). However, this is likely an underestimate due to the inefficiency of sequencing low abundance mRNA isoforms and the inability to sequence the RNA content of mixed cell types in different organs, developmental lineages, or at different developmental timepoints.

Chapter 1: Introduction

There are several different types or patterns of alternative splicing events, including the use of alternative 5' and 3' splice sites, the retention of introns, and the use or exclusion of cassette exons (Black 2003). A summary of the different types of alternative splicing is presented in Figure 2.

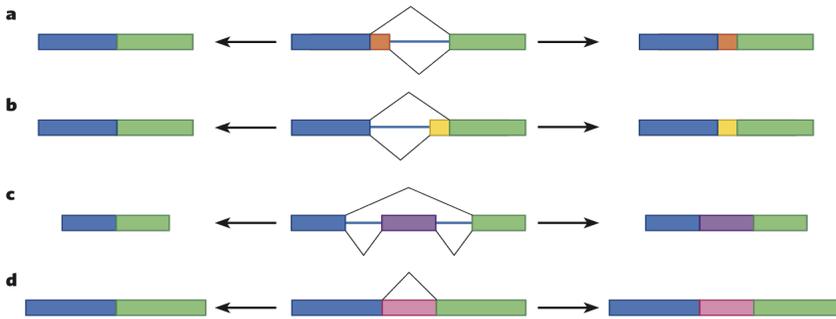


Figure 2. Modes of alternative splicing. a) Alternative 5' splice site. b) Alternative 3' splice site. c) Cassette exon. d) Retained intron. Nilsen T and Graveley BR. (2010) *Nature*. **463**: 457-63.

In the most general sense, all alternative splicing decisions come down to the choice of one splice site over another, i.e. competition. Many of these choices are thought to be kinetic in nature. For example, in some systems, the 5' splice site that most efficiently recruits U1 snRNP is the site that is most likely to be used in a reaction (Hicks et al. 2010). These efficiencies can be modulated by several different types of protein factors. These RNA binding proteins include the SR and hnRNP families of proteins (Smith and Valcarcel 2000; Shepard and Hertel 2009). Previously, it was thought that most SR proteins act at splicing enhancers while most hnRNP proteins act as splicing repressors, but recently it has been seen that both families of proteins can perform both functions (Blanchette et al. 2009). Moreover, cis-acting pre-mRNA splicing regulatory elements, termed splicing enhancers and silencers, have been identified in many pre-mRNAs (Smith and Valcarcel 2000; Black 2003; Blencowe 2006; Wang and Burge 2008). Additionally, RNA structure can play important roles in determining the availability of splice sites as has recently been shown for an *in vitro* selected splice site and for the DSCAM locus (Graveley 2005; Yu et al. 2008; McManus and Graveley 2011).

Most alternative splicing factors influence the kinetics of splicing by assisting in or preventing the recruitment of spliceosomal proteins (Black 2003; Yu et al. 2008[Nilsen, 2010 #22]). Some of these mechanisms, such as the splicing of the *Drosophila P* transposable element third intron and its associated splicing silencer, have been worked out in detail (Siebel et al. 1992; Siebel et al. 1994; Siebel et al. 1995; Adams et al. 1997).

The *Drosophila P* transposable element contains four exons and three introns and can encode an active transposase. When all three introns are spliced out, which occurs only in germline cells, the full-length active transposase protein is produced. This

Chapter 1: Introduction

happens to a small extent in germline cells. In somatic cells, the third intron is fully retained (Laski et al. 1986; Siebel et al. 1992), encoding an mRNA that contains an in-frame stop codon, leading to the expression of a truncated protein that represses P element transposase activity. This splicing inhibition happens predominantly through the actions of the splicing factors PSI (P element somatic inhibitor)(Siebel et al. 1994) and hrp48 (Siebel et al. 1995), which interact with an exonic splicing silencer (ESE) located in the 5' exon adjacent to the third P element intron. PSI, through its AB domain, interacts with the 70K protein of U1 snRNP (Labourier et al. 2001) and recruits it to a pseudo-5' splice site that is upstream of the legitimate 5' splice site in the ESE of the third intron (Siebel et al. 1994; Siebel et al. 1995; Adams et al. 1997). Hrp48 is also a member of this U1-recruiting complex (Siebel et al. 1994). By recruiting U1 to this pseudo-5' splice site, the complex assembled at the ESE may be sterically block the availability of the legitimate 5' splice site to another U1 molecule, thereby preventing the splicing of the third intron. Other repressors, such as mammalian hnRNPA1, can bind on either side of an exon to be repressed and then interact, looping out the repressed exon and bringing the two splice sites to be used into close proximity (Blanchette and Chabot 1999; Martinez-Contreras et al. 2006). hnRNPA1 can also bind to high affinity sites on a target exon leading to exon coating and splicing inhibition (Mayeda and Krainer 1992; Okunola and Krainer 2009).

Splicing activators, on the other hand, recruit spliceosomal proteins to the transcript and facilitate their binding of the important recognition sequences in the intron. As with splicing repressors, the RNA sequences that activators bind (enhancers) can be located in the exon (ESEs) or in the intron (ISEs). Some proteins, such as Nova and PTB, act as both repressors and activators, depending on their locations relative to the alternative splice junctions in question (Ule et al. 2006; Xue et al. 2009; Zhang et al. 2010). The positions of specific sequence RNA regulatory motifs is, therefore, just as important as the presence of the motifs themselves.

Beyond sequences in the RNA, it is also becoming apparent that chromatin structure can also play a role in alternative splicing through modulation of the rate of RNA polymerase II transcription. This can be most easily rationalized by again considering splice site choice in kinetic terms. Chromatin structure and histone marks have been shown to influence the speed with which Pol II transcribes through a chromatin region (Li et al. 2007). Since most splicing occurs cotranscriptionally (Wuarin and Schibler 1994; Khodor et al. 2011), many splice sites are recognized more or less as soon as they are transcribed. Splice sites in areas that are transcribed rapidly may have little opportunity to make a kinetic decision about which site to use (Kornbliht 2006). Conversely, if a weak site is transcribed before a strong site in a transcriptionally slow region, the weak site may have ample time to recruit spliceosomal proteins before the transcriptional appearance of its kinetically stronger competitor site.

Chromatin-mediated alternative splicing may also be much more direct. Certain splicing factors may be recruited to specific histone marks either directly or through histone-

Chapter 1: Introduction

binding factors (Luco et al. 2010) or the interaction of transcriptional proteins with the splicing machinery (Sims et al. 2007). This would allow efficient binding of these proteins to their preferred RNA sequences very soon after they were transcribed, again shifting the kinetics of the process one way or another. Finally, this link between chromatin and alternative splicing has been taken one step further to include siRNA-induced changes in chromatin structure as an initiating force in chromatin-mediated alternative splicing patterns (Allo et al. 2009; Ameyar-Zazoua et al. 2012).

Genomic approaches to studying alternative splicing

Due to the large number of transcripts produced in a typical eukaryotic cell, it certainly cannot be the case that the splicing of each transcript is regulated by an independent, dedicated factor. Splicing factors, therefore, must regulate sets of transcripts, and often do this in a way that produces informational outputs across the transcriptome in a coordinated and synergistic manner (Blanchette et al. 2005; Huelga et al. 2012). It follows, then, that the best way to characterize the effect of a splicing factor is to take into account its influences on a global, genome-wide scale.

Generally, there are two separate questions to be answered when using genome-wide approaches to probe alternative splicing. First, what are the effects on global isoform population (and specific splicing events affected) due to the contribution of a particular factor? Secondly, what transcripts, and specifically, what sequences within those transcripts, are that factor binding? Initially in the early to mid 2000's, both issues were investigated using microarrays. Rough ideas about changes in alternative splicing could be gleaned from microarrays in which the probes were complimentary to exon junctions. Expression of the splicing factor of interest could be reduced using RNAi, and the actions of the factor could then be inferred by analyzing the resulting changes in probe intensities on the array (Johnson et al. 2003; Blanchette et al. 2005; Ule et al. 2005). Although these arrays provided a starting point for a large scale analysis of splicing, their limitations, especially in their scope, were considerable.

The original arrays designed by the Rio lab contained 44,000 probes monitoring over 9400 alternatively spliced junctions and 10,600 constitutively spliced junctions. The remaining probes hybridized to constitutive exons and were used to control for the level of transcript expression.

Later arrays carried more probes that interrogated a larger percentage of the transcriptome, and therefore a larger percentage of potential alternative splicing events (Blanchette et al. 2009). The next generation of these arrays increased the transcriptome coverage substantially. These arrays were used for the genome-wide studies of LS2 and Ago-2 that will be discussed later in this thesis and contained 500,000 probes. These probes monitored the splicing of 11,368 of the 13,472 *Drosophila melanogaster* genes, with 48,550 exon-exon junctions assayed with 3

Chapter 1: Introduction

probes each. The majority of the unassayed genes did not contain introns. Additionally, 242,020 isothermal exonic probes, spaced approximately every 100 nucleotides, were used to control for changes in transcript expression level. Changes in splicing are called for those junctions whose probes change in a statistically significantly fashion relative to the overall change in transcript abundance, as measured by hybridization to the exonic probes.

Although the *Drosophila* genome is very well-annotated, the design of probes for use on all microarrays are inherently limited by the need for a prior knowledge of sequences to be analyzed. New unannotated spliced isoforms and transcribed regions are therefore not detected on these arrays. Additionally, since fluorescent spot intensity at each probe is dependent on nucleic acid hybridization, differential efficiencies of that hybridization must be taken into account. For these reasons, most current studies make use of high-throughput next generation cDNA sequencing (Wang et al. 2009).

Millions of RNA sequence reads can be generated from a small amount of starting material, and with a well-annotated reference genome these can be mapped back with the precision necessary to observe and quantitate differences in splice isoform population. The functions of many different splicing factors have now been characterized using this method: RNA interference knockdown or mutant animals, high-throughput RNA sequencing, mapping sequence reads and determining changes in isoform abundance (Chen and Manley 2009; Licatalosi and Darnell 2010).

Although RNA interference followed by splice junction analysis can be informative, demonstrating that such effects are directly due to the factor of interest requires at the minimum showing an interaction between the protein and its putative pre-mRNA target. High-throughput approaches for determining RNA-protein interactions have followed much the same trajectory as those for splice junction analysis. Originally, RNA binding proteins present in ribonucleoprotein complexes (RNPs) were immunoprecipitated and the co-precipitating RNA was analyzed using tiling arrays similar to those used for chromatin immunoprecipitation followed by tiling array analysis (RIP-CHIP)(Keene et al. 2006). However, with advances in next generation high-throughput sequencing, the popularity of arrays has decreased.

By generating covalently crosslinked chemical bonds between RNAs and proteins, purification strategies can become much more stringent. While most strategies for purifying chromatin-bound proteins rely on formaldehyde as the chemical crosslinking reagent, most strategies for purifying RNA-bound proteins rely on photochemical crosslinking with shortwave UV light. Early studies used UV crosslinking of mRNA-protein complexes in cells to purify and identify the constellation of heterogeneous nuclear ribonucleoproteins (hnRNP proteins) bound to polyA⁺ mRNA (Pinol-Roma et al. 1988). Such a strategy usually involves immunoprecipitation of the protein after UV crosslinking and is thus called CLIP (crosslinking followed by immunoprecipitation) (Ule et al. 2003). Because UV radiation will only form crosslinks over distances of a few

Chapter 1: Introduction

angstroms, such interactions can be relatively safely assumed to be direct (Schoemaker and Schimmel 1974).

Following immunoprecipitation and stringent washing, the RNA is trimmed using nucleases to leave only a small “footprint” of interaction. These small (~30-50nt) RNAs are then isolated, deproteinized, reverse-transcribed, and sequenced (Licatalosi et al. 2008; Witten and Ule 2011). The original CLIP method involved cloning and sequencing of individual cDNA fragments (Ule et al. 2003; Ule et al. 2005). More recent studies have coupled the cDNA population analysis to high-throughput sequencing, termed HITS-CLIP (Licatalosi et al. 2008) or CLIP-seq (Sanford et al. 2009; Yeo et al. 2009). This new method allows hundred of thousands to millions of “CLIP tags” to be mapped and quantitated giving rise to position-specific RNA maps (Licatalosi and Darnell 2010; Witten and Ule 2011).

These methods have now been used on several proteins, including several splicing factors (Xue et al. 2009; Leung et al. 2011; Wang et al. 2012; Xiao et al. 2012). Interesting features about the combinatorial control of splicing through the interaction of several factors with the same pre-mRNA targets are also coming to light (Huelga et al. 2012).

Using *Drosophila* to study alternative splicing

As with other biological questions, *Drosophila* is an excellent system in which to study alternative splicing. Perhaps one of the most famous examples are alternative splicing regulation involves sexual differentiation and behavior in *Drosophila* (Black 2003; Demir and Dickson 2005). The extent to which alternative splicing occurs in *Drosophila* is lower than in mammalian systems. However, because the *Drosophila* genome is approximately 20 times smaller than the human genome, and because the *Drosophila* genome is extremely well annotated, alternative splicing patterns in *Drosophila* can often be easier to analyze.

The genetic and molecular genetic tools available in *Drosophila* also make it an excellent model organism to study alternative splicing. Many mutant or insertion strains are available that contain transposon insertions throughout the genome at precise locations (Parks et al. 2004; Thibault et al. 2004). These can then be used directly, if the transposon interrupts the gene of interest, or by further manipulations with pairs of transposons to delete genes or entire genomic regions following recombination.

In addition, new collections of *Drosophila* strains are being generated that express short hairpin RNAs (shRNA) against the gene of interest (Ni et al. 2008; Haley et al. 2010; Ni et al. 2011). As of 2012, strains exist expressing shRNA hairpins against almost every annotated gene. However, the promoters and enhancers of these shRNAs can determine whether that shRNA is expressed efficiently in somatic cells, germline cells,

Chapter 1: Introduction

or both. Currently, strains do not exist with shRNAs directed against every gene that can be expressed in every tissue type. However, the construction of new lines to fill a particular need is not difficult and only takes a few months.

Splice site silencing mechanism at the P element third intron

As previously stated, the mechanism of splice site silencing at the P element third intron has been studied in detail. There are still, however, several unanswered questions.

First, what proteins are involved in the silencing? A former postdoc in the Rio lab, Jiro Yasuhara, studied this question using a combination of biochemistry and functional assays in S2 cells. Using biotinylated P element silencer transcripts, he biochemically purified in vitro-assembled splicing complexes and identified several proteins that stably bound transcripts with the wildtype silencer but that bound transcripts with a mutated silencer region much less efficiently. These proteins included the previously identified PSI and hrp48 proteins, but also included several new factors such as Sqd/hrp40, hrp38, hrp36, and PABPC1. These proteins were then tested for their ability to mediate silencing using an RNAi-based splicing reporter assay in S2 cells. RNA interference knockdown of hrp36, hrp38, hrp48 and PABP each caused a decrease in the level of silencing at the third intron, implying a functional involvement in P element splicing control in vivo.

Second, how are these proteins organized on the P element transcript to form a repressive complex? To address this question, a nuclease protection assay using a site-specifically labeled P element silencer RNA was used. These RNAs contain a single radioactive phosphate at a known position. In contrast to standard nuclease protection assays in which the entire RNA may be labeled, the use of site-specifically labeled RNA allows much more precision in the determination of binding sites (Maroney et al. 2000). The binding sites for PSI and hrp48 on the P element silencer RNA have been determined using this method (Appendix A), and they agree with those derived from previous data (Siebel et al. 1992; Siebel et al. 1994). A preliminary model has been outlined in which the interaction of U1 snRNP with the silencer complex at the pseudo-5' splice site is necessary to create a complex that is large enough to sterically block the interaction of another U1 snRNP molecule with the accurate third intron 5' splice site.

Phosphorylation of PSI

Previous data produced by our lab and whole proteome phosphopeptide mapping (Zhai et al. 2008) had shown that PSI was phosphorylated at a minimum of two positions: serine 42 and serine 61. These phosphorylation sites were identified by purification of an epitope-tagged PSI transgene produced in S2 cells followed by mass spectrometry.

Chapter 1: Introduction

However, the protein kinase or kinases responsible for these phosphorylations and their function were unknown.

To identify the kinases responsible, we have biochemically purified the kinase activities associated with PSI phosphorylation from S2 cell nuclear extract. After several biochemical fractionation steps and identification of proteins by mass spectrometry, we have identified casein kinase II as the protein kinase that phosphorylates serine 61. We have confirmed this finding using serine to alanine mutants of PSI. We have also identified tousel-like kinase (tlk) as the protein kinase likely responsible for phosphorylation of serine 42. These phosphorylation events seem to modulate the protein-protein interaction partners of PSI, as indicated by pulldown experiments with mutant PSI proteins.

Endogenous modulation of splice site use by small RNAs and associated proteins

Small RNAs are well-known modulators of mRNA levels for many genes. Their biogenesis begins with a double-stranded RNA precursor. MicroRNAs are first trimmed by Drosha. Both classes of small RNAs, microRNAs and siRNAs, are then processed, usually by a Dicer protein, to between 20 and 25 nt and loaded into an effector molecule, using an Argonaute (Ago) family member, where it can then be used as the basis for recognition of specific transcripts through basepairing interactions (Peters and Meister 2007). Once targeted by an Argonaute protein loaded with a small RNA, these transcripts are then subject to a variety of different fates in the cytoplasm, including degradation and translational repression (Carthew and Sontheimer 2009). Additionally, Ago proteins can localize to chromatin where they can direct localized changes in histone marks and chromatin structure (Allo et al. 2009). Ago proteins can also influence heterochromatin formation in *S. pombe* (Verdel et al. 2004) and the establishment of chromatin marks in mammalian cells (Ameyar-Zazoua et al. 2012).

There is evidence that changes in chromatin structure may indirectly influence splicing (Luco et al. 2010). Certain chromatin marks are enriched in exons over introns (Andersson et al. 2009; Schwartz et al. 2009), indicating a link between chromatin structure and splicing. Chromatin structure has also been known to modulate the rate of transcription through a particular genomic region (Li et al. 2007), and changes in the rate of transcription can then result in splicing changes (Kornbliht 2006; Munoz et al. 2009).

Alternatively, changes in histone marks could more directly lead to splicing changes through the recruitment of alternative splicing factors. This was recently shown to be the case for trimethylated H3K36 marks and their recruitment of PTB via direct binding of modified chromatin by an adapter protein (Luco et al. 2010).

Chapter 1: Introduction

In *Drosophila*, there are two small RNA systems that operate in parallel: the microRNA (miRNA) and small interfering RNA (siRNA) pathways. Although there is some crosstalk between these two systems, they generally have dedicated components and lead to two distinct outcomes for the targeted transcript (Forstemann et al. 2007).

Although some microRNAs can arise from introns, most are produced from distinct microRNA loci and arise via processing of hairpin RNA intermediates by Dicer enzymes. A primary microRNA is transcribed in the nucleus and forms a secondary structure that is then processed first by the nuclear enzyme Drosha to remove the tails from the stem part of the structure. This processing reaction yields a pre-microRNA which is then exported to the cytoplasm. Next, the loop and part of the stem of the pre-microRNA is then removed by the enzyme Dicer-1 and the mature microRNA is exported to the cytoplasm where it is loaded into Ago1 complexes (Carthew and Sontheimer 2009).

siRNAs can originate from both exogenous and endogenous double-stranded RNA sources. Exogenous sources of siRNAs include viral replication intermediates and double-stranded RNAs used in RNA interference. Endogenous siRNAs come from many different places in the genome and arise through a handful of different mechanisms. Convergent transcription events can lead to RNAs produced from both DNA strands at a single locus which can then anneal and enter the siRNA processing pathway. Separate transcripts from distantly located genes can also have complementary regions that anneal to form double-stranded RNA (Czech et al. 2008; Okamura et al. 2008; Okamura and Lai 2008). Once annealed, these RNAs are then processed by Dicer-2 and loaded into Ago-2 complexes (Kawamura et al. 2008).

The principal difference between miRNAs and siRNAs is their complementarity. MicroRNAs tend to not be perfectly basepaired, either in the pre-microRNA precursor or, with the exception of the seed sequence, upon hybridization to the target mRNA. siRNAs, on the other hand, tend to be perfectly basepaired across their entire length, due to their method of production.

It is the nature of this RNA-RNA complementarity that is the basis for the sorting of these different small RNAs into their respective Argonaute protein complexes. Processed double-stranded RNAs that contain mismatches are more likely to be loaded into Ago-1, while those that are perfectly basepaired are more likely to be loaded into Ago-2 (Tomari et al. 2007). The nucleolytic activity of Ago-1 is quite weak, and therefore Ago-1 is thought to be more involved in translational repression rather than degradation (Forstemann et al. 2007). The catalytic activity of Ago-2 is much higher than Ago-1, and it is thought to be more involved in degradative mechanisms, in accordance with its role in viral defense (Forstemann et al. 2007).

Previous studies had linked siRNA-induced heterochromatin formation with changes in splicing patterns (Allo et al. 2009). However, these studies had looked at changes at only a few loci. To understand these mechanisms on a global scale, we used RNA

Chapter 1: Introduction

interference knockdowns of the *Drosophila* Dicers and Argonautes followed by splice junction microarrays to determine genome-wide which splice junctions in the *Drosophila* transcriptome were sensitive to the levels of these factors. We detected minimal splicing changes in response to Dcr-1, Dcr-2, and Ago1 knockdown, but over 100 changes in response to the Ago-2 knockdown. We then attempted to correlate these changes in splicing with the direct binding of Ago-2, both at the chromatin level using ChIP-seq and at the RNA level using RIP-RT-PCR and CLIP-seq. We observe chromatin loci that are bound by Ago-2 and polycomb group proteins. Transcription at these loci is downregulated. This population of bound genes is distinct, however, from transcripts that we observe bound by Ago-2. These results are discussed in detail in Chapter 3.

Drosophila orthologs of U2AF and the evolution of a tissue-specific splicing network

The recognition and determination of 3' splice sites is primarily carried out by U2-associated factor (U2AF) (Ruskin et al. 1988; Zamore and Green 1989). The essential, highly conserved U2AF general splicing factor is a heterodimer composed of large (U2AF⁶⁵ in humans, dU2AF⁵⁰ in *Drosophila*) and small (U2AF³⁵ in humans, dU2AF³⁸ in *Drosophila*) subunits that promotes spliceosome assembly (Ruskin et al. 1988). U2AF is conserved among all eukaryotic species, from *S. pombe* to humans. The large subunit of U2AF recognizes the polypyrimidine tract at the 3' end of the intron (Zamore and Green 1989; Kanaar et al. 1993) while the small subunit interacts with the intron-terminal AG dinucleotide (Merendino et al. 1999; Wu et al. 1999; Zorio and Blumenthal 1999). The large subunit additionally cooperates with the branch point binding protein SF1 through interactions in its C-terminal pseudo-RNA recognition motif (Kent et al. 2003; Selenko et al. 2003).

Following assembly of these RNA-protein complexes, the 3' end of the intron is then competent for interaction with U2 snRNP. U2AF therefore functions to promote spliceosome assembly.

Much work has been done concerning the evolutionary conservation of the *cis*-acting RNA regulatory sequence elements. Many RNA sequence elements are widely conserved even across vast evolutionary distances and often lead to similar splicing patterns in the orthologous transcripts (Brooks et al. 2011). However, little is understood about how related family members of the RNA-binding proteins that mediate these splicing effects arise and diverge to acquire distinct and diverse functions (Baek and Green 2005; Akerman et al. 2009). These distinct functions allow evolutionarily related proteins to form regulatory networks, with each family member controlling the splicing of specific transcripts through the recognition of specific sequence motifs.

Chapter 1: Introduction

We identified and characterized the appearance and evolutionary divergence of a *Drosophila* splicing factor that we termed LS2 (Large Subunit 2, also known as CG3162). LS2 arose from a retroduplicated copy of the highly conserved, positively acting dU2AF⁵⁰ and has diverged sufficiently from dU2AF⁵⁰ such that it is highly specialized in its specificity, function, and expression (Taliaferro et al. 2011). LS2 expression is restricted to the differentiated cells in *Drosophila* testes and may play a role in the restriction of splicing that is seen upon cell differentiation in the testes (Gan et al. 2010). These results are discussed in detail in Chapter 4.

REFERENCES

- Adams MD, Tarrn RS, Rio DC. 1997. The alternative splicing factor PSI regulates P-element third intron splicing in vivo. *Genes Dev* **11**: 129-138.
- Akerman M, David-Eden H, Pinter RY, Mandel-Gutfreund Y. 2009. A computational approach for genome-wide mapping of splicing factor binding sites. *Genome Biol* **10**: R30.
- Allo M, Buggiano V, Fededa JP, Petrillo E, Schor I, de la Mata M, Agirre E, Plass M, Eyras E, Elela SA et al. 2009. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol* **16**: 717-724.
- Ameyar-Zazoua M, Rachez C, Souidi M, Robin P, Fritsch L, Young R, Morozova N, Fenouil R, Descostes N, Andrau JC et al. 2012. Argonaute proteins couple chromatin silencing to alternative splicing. *Nat Struct Mol Biol* **19**: 998-1004.
- Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J. 2009. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res* **19**: 1732-1741.
- Baek D, Green P. 2005. Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. *Proc Natl Acad Sci U S A* **102**: 12813-12818.
- Barberan-Soler S, Zahler AM. 2008. Alternative splicing regulation during *C. elegans* development: splicing factors as regulated targets. *PLoS Genet* **4**: e1000001.
- Berglund JA, Abovich N, Rosbash M. 1998a. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev* **12**: 858-867.

Chapter 1: Introduction

- Berglund JA, Fleming ML, Rosbash M. 1998b. The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA* **4**: 998-1006.
- Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**: 291-336.
- Black DL, Chabot B, Steitz JA. 1985. U2 as Well as U1 Small Nuclear Ribonucleoproteins Are Involved in Pre-Messenger Rna Splicing. *Cell* **42**: 737-750.
- Blanchette M, Chabot B. 1999. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J* **18**: 1939-1952.
- Blanchette M, Green RE, Brenner SE, Rio DC. 2005. Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. *Genes Dev* **19**: 1306-1314.
- Blanchette M, Green RE, MacArthur S, Brooks AN, Brenner SE, Eisen MB, Rio DC. 2009. Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. *Mol Cell* **33**: 438-449.
- Blencowe BJ. 2006. Alternative splicing: new insights from global analyses. *Cell* **126**: 37-47.
- Bringmann P, Appel B, Rinke J, Reuter R, Theissen H, Luhrmann R. 1984. Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J* **3**: 1357-1363.
- Brooks AN, Yang L, Duff MO, Hansen KD, Park JW, Dudoit S, Brenner SE, Graveley BR. 2011. Conservation of an RNA regulatory map between Drosophila and mammals. *Genome Res* **21**: 193-202.
- Carthew RW, Sontheimer EJ. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**: 642-655.
- Chen M, Manley JL. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**: 741-754.
- Cooper TA, Wan L, Dreyfuss G. 2009. RNA and disease. *Cell* **136**: 777-793.
- Crowder SM, Kanaar R, Rio DC, Alber T. 1999. Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal. *Proc Natl Acad Sci U S A* **96**: 4892-4897.

Chapter 1: Introduction

- Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R et al. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453**: 798-802.
- Demir E, Dickson BJ. 2005. fruitless splicing specifies male courtship behavior in *Drosophila*. *Cell* **121**: 785-794.
- Forstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD. 2007. *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* **130**: 287-297.
- Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K, Chen X. 2010. Dynamic regulation of alternative splicing and chromatin structure in *Drosophila* gonads revealed by RNA-seq. *Cell Res* **20**: 763-783.
- Glisovic T, Bachorik JL, Yong J, Dreyfuss G. 2008. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* **582**: 1977-1986.
- Gozani O, Potashkin J, Reed R. 1998. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* **18**: 4752-4760.
- Graveley BR. 2005. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**: 65-73.
- Green MR. 1986. Pre-mRNA splicing. *Annu Rev Genet* **20**: 671-708.
- Haley B, Foys B, Levine M. 2010. Vectors and parameters that enhance the efficacy of RNAi-mediated gene disruption in transgenic *Drosophila*. *Proc Natl Acad Sci U S A* **107**: 11435-11440.
- Hausner TP, Giglio LM, Weiner AM. 1990. Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. *Genes Dev* **4**: 2146-2156.
- Hicks MJ, Mueller WF, Shepard PJ, Hertel KJ. 2010. Competing upstream 5' splice sites enhance the rate of proximal splicing. *Mol Cell Biol* **30**: 1878-1886.
- Hoffman BE, Grabowski PJ. 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes Dev* **6**: 2554-2568.
- Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, Wombacher R, Ramirez N, Cornish VW, Gelles J, Moore MJ. 2011. Ordered and dynamic assembly of single spliceosomes. *Science* **331**: 1289-1295.

Chapter 1: Introduction

- Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiue L, Hoon S, Brenner S et al. 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* **1**: 167-178.
- Inoue M, Muto Y, Sakamoto H, Yokoyama S. 2000. NMR studies on functional structures of the AU-rich element-binding domains of Hu antigen C. *Nucleic Acids Res* **28**: 1743-1750.
- Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD. 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**: 2141-2144.
- Kanaar R, Roche SE, Beall EL, Green MR, Rio DC. 1993. The conserved pre-mRNA splicing factor U2AF from *Drosophila*: requirement for viability. *Science* **262**: 569-573.
- Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H. 2008. *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**: 793-797.
- Keene JD, Komisarow JM, Friedersdorf MB. 2006. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* **1**: 302-307.
- Kent OA, Reayi A, Foong L, Chilibeck KA, MacMillan AM. 2003. Structuring of the 3' splice site by U2AF65. *J Biol Chem* **278**: 50572-50577.
- Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT, 2nd, Rosbash M. 2011. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev* **25**: 2502-2512.
- Kielkopf CL, Rodionova NA, Green MR, Burley SK. 2001. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* **106**: 595-605.
- Konarska MM, Sharp PA. 1986. Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell* **46**: 845-855.
- Kornblihtt AR. 2006. Chromatin, transcript elongation and alternative splicing. *Nat Struct Mol Biol* **13**: 5-7.
- Labourier E, Adams MD, Rio DC. 2001. Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. *Mol Cell* **8**: 363-373.

Chapter 1: Introduction

- Laski FA, Rio DC, Rubin GM. 1986. Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. *Cell* **44**: 7-19.
- Lee AL, Kanaar R, Rio DC, Wemmer DE. 1994. Resonance assignments and solution structure of the second RNA-binding domain of sex-lethal determined by multidimensional heteronuclear magnetic resonance. *Biochemistry* **33**: 13775-13786.
- Leung AK, Young AG, Bhutkar A, Zheng GX, Bosson AD, Nielsen CB, Sharp PA. 2011. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol* **18**: 237-244.
- Lewis HA, Musunuru K, Jensen KB, Edo C, Chen H, Darnell RB, Burley SK. 2000. Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100**: 323-332.
- Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. *Cell* **128**: 707-719.
- Licatalosi DD, Darnell RB. 2010. RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet* **11**: 75-87.
- Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**: 464-469.
- Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science* **327**: 996-1000.
- Mackereth CD, Madl T, Bonnal S, Simon B, Zanier K, Gasch A, Rybin V, Valcarcel J, Sattler M. 2011. Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* **475**: 408-411.
- MacMorris MA, Zorio DA, Blumenthal T. 1999. An exon that prevents transport of a mature mRNA. *Proc Natl Acad Sci U S A* **96**: 3813-3818.
- Madhani HD, Guthrie C. 1992. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71**: 803-817.
- Maroney PA, Romfo CM, Nilsen TW. 2000. Nuclease protection of RNAs containing site-specific labels: a rapid method for mapping RNA-protein interactions. *RNA* **6**: 1905-1909.

Chapter 1: Introduction

- Martinez-Contreras R, Fiset JF, Nasim FU, Madden R, Cordeau M, Chabot B. 2006. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol* **4**: e21.
- Maschhoff KL, Padgett RA. 1993. The stereochemical course of the first step of pre-mRNA splicing. *Nucleic Acids Res* **21**: 5456-5462.
- Mayeda A, Krainer AR. 1992. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**: 365-375.
- McManus CJ, Graveley BR. 2011. RNA structure and the mechanisms of alternative splicing. *Curr Opin Genet Dev* **21**: 373-379.
- Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J. 1999. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**: 838-841.
- Moore MJ, Sharp PA. 1993. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365**: 364-368.
- Moore MJ, Wang Q, Kennedy CJ, Silver PA. 2010. An alternative splicing network links cell-cycle control to apoptosis. *Cell* **142**: 625-636.
- Mount SM. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res* **10**: 459-472.
- Munoz MJ, Perez Santangelo MS, Paronetto MP, de la Mata M, Pelisch F, Boireau S, Glover-Cutter K, Ben-Dov C, Blaustein M, Lozano JJ et al. 2009. DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* **137**: 708-720.
- Ni JQ, Markstein M, Binari R, Pfeiffer B, Liu LP, Villalta C, Booker M, Perkins L, Perrimon N. 2008. Vector and parameters for targeted transgenic RNA interference in *Drosophila melanogaster*. *Nat Methods* **5**: 49-51.
- Ni JQ, Zhou R, Czech B, Liu LP, Holderbaum L, Yang-Zhou D, Shim HS, Tao R, Handler D, Karpowicz P et al. 2011. A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nat Methods* **8**: 405-407.
- Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**: 457-463.
- Okamura K, Balla S, Martin R, Liu N, Lai EC. 2008. Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat Struct Mol Biol* **15**: 998.

Chapter 1: Introduction

- Okamura K, Lai EC. 2008. Endogenous small interfering RNAs in animals. *Nat Rev Mol Cell Biol* **9**: 673-678.
- Okunola HL, Krainer AR. 2009. Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol Cell Biol* **29**: 5620-5631.
- Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**: 432-438.
- Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA. 1986. Splicing of messenger RNA precursors. *Annu Rev Biochem* **55**: 1119-1150.
- Pagani F, Baralle FE. 2004. Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* **5**: 389-396.
- Parks AL, Cook KR, Belvin M, Dompe NA, Fawcett R, Huppert K, Tan LR, Winter CG, Bogart KP, Deal JE et al. 2004. Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. *Nat Genet* **36**: 288-292.
- Patel AA, Steitz JA. 2003. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* **4**: 960-970.
- Perez-Canadillas JM, Varani G. 2001. Recent advances in RNA-protein recognition. *Curr Opin Struct Biol* **11**: 53-58.
- Peters L, Meister G. 2007. Argonaute proteins: mediators of RNA silencing. *Mol Cell* **26**: 611-623.
- Pinol-Roma S, Choi YD, Matunis MJ, Dreyfuss G. 1988. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev* **2**: 215-227.
- Ramos A, Grunert S, Adams J, Micklem DR, Proctor MR, Freund S, Bycroft M, St Johnston D, Varani G. 2000. RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J* **19**: 997-1009.
- Reed R. 2000. Mechanisms of fidelity in pre-mRNA splicing. *Curr Opin Cell Biol* **12**: 340-345.
- Robberson BL, Cote GJ, Berget SM. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* **10**: 84-94.
- Rudner DZ, Breger KS, Kanaar R, Adams MD, Rio DC. 1998a. RNA binding activity of heterodimeric splicing factor U2AF: at least one RS domain is required for high-affinity binding. *Mol Cell Biol* **18**: 4004-4011.

Chapter 1: Introduction

- Rudner DZ, Kanaar R, Breger KS, Rio DC. 1998b. Interaction between subunits of heterodimeric splicing factor U2AF is essential in vivo. *Mol Cell Biol* **18**: 1765-1773.
- Ruskin B, Krainer AR, Maniatis T, Green MR. 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* **38**: 317-331.
- Ruskin B, Zamore PD, Green MR. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**: 207-219.
- Ryter JM, Schultz SC. 1998. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J* **17**: 7505-7513.
- Sanford JR, Wang X, Mort M, Vanduy N, Cooper DN, Mooney SD, Edenberg HJ, Liu Y. 2009. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res* **19**: 381-394.
- Schneider M, Will CL, Anokhina M, Tazi J, Urlaub H, Luhrmann R. 2010. Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol Cell* **38**: 223-235.
- Schoemaker HJ, Schimmel PR. 1974. Photo-induced joining of a transfer RNA with its cognate aminoacyl-transfer RNA synthetase. *J Mol Biol* **84**: 503-513.
- Schwartz S, Meshorer E, Ast G. 2009. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* **16**: 990-995.
- Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Kramer A, Sattler M. 2003. Structural basis for the molecular recognition between human splicing factors U2AF(65) and SF1/mBBP. *Mol Cell* **11**: 965-976.
- Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL. 2008. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat Struct Mol Biol* **15**: 183-191.
- Shepard PJ, Hertel KJ. 2009. The SR protein family. *Genome Biol* **10**: 242.
- Sickmier EA, Frato KE, Shen H, Paranawithana SR, Green MR, Kielkopf CL. 2006. Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol Cell* **23**: 49-59.
- Siebel CW, Admon A, Rio DC. 1995. Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes Dev* **9**: 269-283.

Chapter 1: Introduction

- Siebel CW, Fresco LD, Rio DC. 1992. The mechanism of somatic inhibition of *Drosophila* P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. *Genes Dev* **6**: 1386-1401.
- Siebel CW, Kanaar R, Rio DC. 1994. Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. *Genes Dev* **8**: 1713-1725.
- Sims RJ, 3rd, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL, Reinberg D. 2007. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol Cell* **28**: 665-676.
- Singh R, Valcarcel J, Green MR. 1995. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**: 1173-1176.
- Smith CW, Valcarcel J. 2000. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* **25**: 381-388.
- Sun JS, Manley JL. 1995. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev* **9**: 843-854.
- Taliaferro JM, Alvarez N, Green RE, Blanchette M, Rio DC. 2011. Evolution of a tissue-specific splicing network. *Genes Dev* **25**: 608-620.
- Thibault ST, Singer MA, Miyazaki WY, Milash B, Dompe NA, Singh CM, Buchholz R, Demsky M, Fawcett R, Francis-Lang HL et al. 2004. A complementary transposon tool kit for *Drosophila melanogaster* using P and piggyBac. *Nat Genet* **36**: 283-287.
- Tomari Y, Du T, Zamore PD. 2007. Sorting of *Drosophila* small silencing RNAs. *Cell* **130**: 299-308.
- Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**: 1212-1215.
- Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**: 580-586.
- Ule J, Ule A, Spencer J, Williams A, Hu JS, Cline M, Wang H, Clark T, Fraser C, Ruggiu M et al. 2005. Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* **37**: 844-852.
- Valadkhan S, Manley JL. 2001. Splicing-related catalysis by protein-free snRNAs. *Nature* **413**: 701-707.

Chapter 1: Introduction

- Valcarcel J, Gaur RK, Singh R, Green MR. 1996. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science* **273**: 1706-1709.
- Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**: 672-676.
- Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S et al. 2012. Transcriptome-wide Regulation of Pre-mRNA Splicing and mRNA Localization by Muscleblind Proteins. *Cell* **150**: 710-724.
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470-476.
- Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**: 802-813.
- Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**: 57-63.
- Will CL, Luhrmann R. 2011. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**.
- Will CL, Schneider C, MacMillan AM, Katopodis NF, Neubauer G, Wilm M, Luhrmann R, Query CC. 2001. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J* **20**: 4536-4546.
- Witten JT, Ule J. 2011. Understanding splicing regulation through RNA splicing maps. *Trends Genet* **27**: 89-97.
- Wu JY, Maniatis T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**: 1061-1070.
- Wu S, Romfo CM, Nilsen TW, Green MR. 1999. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**: 832-835.
- Wuarin J, Schibler U. 1994. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol Cell Biol* **14**: 7219-7225.
- Xiao R, Tang P, Yang B, Huang J, Zhou Y, Shao C, Li H, Sun H, Zhang Y, Fu XD. 2012. Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Mol Cell* **45**: 656-668.

Chapter 1: Introduction

- Xu Q, Modrek B, Lee C. 2002. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res* **30**: 3754-3766.
- Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon YS, Zhang C, Yeo G, Black DL, Sun H et al. 2009. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell* **36**: 996-1006.
- Yean SL, Wuenschell G, Termini J, Lin RJ. 2000. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**: 881-884.
- Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* **16**: 130-137.
- Yu Y, Maroney PA, Denker JA, Zhang XH, Dybkov O, Luhrmann R, Jankowsky E, Chasin LA, Nilsen TW. 2008. Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. *Cell* **135**: 1224-1236.
- Zamore PD, Green MR. 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A* **86**: 9243-9247.
- Zhai B, Villen J, Beausoleil SA, Mintseris J, Gygi SP. 2008. Phosphoproteome analysis of *Drosophila melanogaster* embryos. *J Proteome Res* **7**: 1675-1682.
- Zhang C, Frias MA, Mele A, Ruggiu M, Eom T, Marney CB, Wang H, Licatalosi DD, Fak JJ, Darnell RB. 2010. Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* **329**: 439-443.
- Zhuang Y, Weiner AM. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* **46**: 827-835.
- Zorio DA, Blumenthal T. 1999. Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* **402**: 835-838.

SUMMARY

Alternative splicing of pre-mRNA is a highly regulated process that allows cells to change their genetic informational output. These changes are mediated by protein factors that directly bind specific pre-mRNA sequences. Although the transcripts regulated and bound by many splicing factors are known, there are relatively fewer instances where the mode of splicing regulation is explained in mechanistic detail. Here, we find that a member of this complex, P element somatic inhibitor (PSI), is phosphorylated *in vivo* by at least two different kinases, and that these phosphorylation events may be important for regulating protein-protein interactions. Finally, we show that PSI interacts with several proteins in *Drosophila* S2 tissue culture cells, the majority of which are splicing factors.

INTRODUCTION

Splicing of the third intron of the P element transposase

The *Drosophila* P transposable element encodes a transposase that has the ability to cut and paste the P element DNA from one place to another in the genome. The full-length P element contains four protein-coding exons and three introns. The third intron (IVS3) contains an in-frame stop codon and tissue-specific pre-mRNA splicing is used to restrict expression of the active transposase to the germline (Laski et al. 1986; Rio et al. 1986). In somatic cells, full-length transposase is not expressed, and instead, due to the premature stop codon, a truncated DNA binding and transposition-repressing form of the protein is expressed (Misra and Rio 1990).

It later became clear that there were sequences in the exon immediately upstream of the third intron that were crucial for this repression, now termed a splicing silencer (Siebel and Rio 1990; Siebel et al. 1992). Specifically, a sequence in the exon upstream of the accurate IVS3 5' splice site, which resembled a 5' splice site in sequence, but which was not used for splicing was termed a pseudo-5' splice site. In *Drosophila* somatic cell nuclear extracts, IVS3 splicing was inhibited in human cells, these upstream sequences, termed F1 and F2, were bound by U1 snRNP and a several other protein factors. Conversely, in human cell nuclear extracts, U1 snRNP bound to the accurate 5' splice site and human splicing extracts supported accurate IVS3 splicing (Siebel and Rio 1990; Siebel et al. 1992). Mutations in the F1/F2 sites or the unlabeled 5' exon RNA titrations activated IVS3 splicing in *Drosophila* nuclear extracts.

Biochemical fractionation of *Drosophila* nuclear extracts and RNA affinity chromatography led to the identification of two proteins that bound the P element splicing F1/F2 silencer element, a KH-domain protein called PSI which is related to mammalian KSRP and FBP proteins and a two RNP domain protein called hrp48 which is related to mammalian hnRNP A1. PSI is necessary for the repression of splicing, as addition of anti-PSI antibody to *in vitro* splicing reaction of the third intron restored splicing at the true 5' splice site (Siebel et al. 1994), but the inhibition of splicing could

Chapter 2: Phosphorylation of PSI

be restored by the addition of purified recombinant PSI to the reaction (Siebel et al. 1995).

An analysis of the expression pattern of PSI revealed that it was expressed solely in the nuclei of somatic cells, and not in the germline (Siebel et al. 1995), consistent with the patterns of splicing repression. Furthermore, overexpression of PSI in germline cells blocked P element IVS splicing in ovaries (Adams et al. 1997). Hrp48 was also found to bind the F1/F2 sites (Siebel et al. 1992; Siebel et al. 1994; Siebel et al. 1995) and mutations in *hrp48* activated P element IVS3 splicing in somatic cells (Hammond et al. 1997).

PSI, like its mammalian homologs, contains four N-terminal KH-type RNA binding domains. The repression of P element splicing, though, requires an interaction between a C-terminal region of PSI and the U1 snRNP 70K subunit (Labourier et al. 2001; Ignjatovic et al. 2005). Mechanistically, then, it is believed that recruitment of U1 snRNP to the F1/F2 pseudo-5' splice sites and formation of a stable RNA-protein complex at the silencer element would block binding of another U1 snRNP molecule to the accurate 5' splice site. A direct interaction between U1 snRNP and PSI might stabilize U1 snRNP binding to the F1/F2 sites and block the initial steps of spliceosome assembly (Labourier et al. 2002). Molecular genetic studies showed that PSI deletion mutants are embryonic lethal, and its protein-protein interaction with U1 snRNP 70K is necessary for male fertility (Labourier et al. 2002). Using splicing-sensitive microarrays, PSI was found to regulate 43 splicing events in S2 cells (Blanchette et al. 2005). Biochemical purification of P element silencer RNA-protein complexes and RNAi screens have identified *hrp36*, *hrp38* and PABPC as also interacting with the P element silencer and affecting IVS3 splicing *in vivo* (J. Yasuhara, unpublished data).

Phosphorylation of PSI

Although many studies have investigated the effects that splicing factors have on alternative splicing, few have looked at the regulation of the factors themselves. Many of these instances of regulation occur pre-translationally, often at the level of splicing. For example, many SR protein regulate their own splicing as well as that of heterologous SR proteins in a way that shunts those transcripts into the NMD pathway (Lareau et al. 2007; Anko et al. 2012).

Some splicing factors are known to be post-translationally modified. These events can affect the RNA binding capabilities of the protein (Xiao and Manley 1998) as well as the assembly of higher-order structures like the spliceosome (Wang et al. 1999). The spliceosomal protein SAP155 and NIPP1 are phosphorylated, and this modification is necessary for their interaction (Boudrez et al. 2002). SR proteins and other splicing factors are highly phosphorylated *in vivo* (Roth et al. 1990). Additionally, the RS domain of the splicing factor ASF/SF2 greatly affects the protein and RNA binding capabilities of the protein, and is necessary for splicing (Xiao and Manley 1997; Xiao and Manley 1998).

Chapter 2: Phosphorylation of PSI

We have identified two phosphorylation sites on PSI by mass spectrometry and identified two kinases that phosphorylate the N-terminus of PSI. These phosphorylation events may play a role in the ability of PSI to interact with other proteins. Additionally, we have identified several interaction partners of PSI, suggesting that PSI is present in cells as a member of large ribonucleoprotein complexes.

RESULTS

PSI is phosphorylated in vivo

Using a Polyoma (also known as Glu-Glu) tagged version of PSI, we purified PSI from *Drosophila* Kc cells. Interestingly, PSI purified from *Drosophila* cells migrated on SDS-PAGE gels as a doublet (Figure 1A, Figure 5B) while recombinant PSI purified from *E. coli* migrated as a single species (Figure 1A). We reasoned that PSI phosphorylation events occurring in *Drosophila* cells could be responsible for the doublet. Consistent with this idea, treatment of PSI purified from Kc cells with calf intestinal phosphatase (CIP) collapsed the doublet to a faster migrating band while having no effect on the migration of recombinantly produced PSI (Figure 1A).

To characterize this apparent phosphorylation, we digested purified endogenous *Drosophila* PSI with multiple proteases and analyzed the resulting peptides by multidimensional chromatography / mass spectrometry. The resulting data covered 84.5% of the sequence to an average depth of 10 observations per peptide. Manual evaluation of the spectra assigned to phosphopeptides confirmed two phosphorylation sites at Ser 42 and Ser 61 (Figure 1B). Spectra showing phosphorylation of at Ser 42 were measured 5 times in two different peptides. Spectra of the corresponding unmodified peptides were measured 86 and 5 times, respectively. The characteristics of the unmodified spectra supported the interpretation of the modified spectra (Figure 2).

The global phosphoproteomic analyses of *Drosophila* embryos by Zhai et al (Zhai et al. 2008) also identified Ser 42 and Ser 61 as phosphorylation sites. A third site, Ser 85, identified in that study was not detected on our analysis. Inspection of the supporting data from Zhai et al for Ser 85 showed that the CID spectrum contained no neutral loss peak; consequently, the identification is likely to be a false positive.

PSI is phosphorylated in *Drosophila* cells by casein kinase II

In order to identify the protein kinase or kinases responsible for the observed phosphorylations of PSI, we used chromatographic fractionation of *Drosophila* embryo nuclear extract and followed PSI-phosphorylating activity using recombinant PSI and radioactive γ -³²P-ATP as substrates (Figure 3A). To simplify the purification and exclude the phosphorylation of other residues, we used an N-terminal fragment of PSI

Chapter 2: Phosphorylation of PSI

that contained only the first 95 amino acids, including the two residues, Ser 42 and Ser 61, that we identified as being phosphorylated by mass spectrometry (Figure 1B). After five purification steps, we analyzed the protein content of the peak fraction of PSI kinase activity by SDS-PAGE and silver staining (Figure 3B). We detected two prominent protein species migrating at approximately 37 and 30 kDa that appeared to be at stoichiometric levels with each other. We excised these bands, as well as several other prominent bands, from the gel and performed mass spectrometry. The 37 and 30 kDa bands were identified as the alpha and beta subunits, respectively, of casein kinase II. No other known protein kinases were identified in this fraction.

Casein kinase II is a tetramer composed of two catalytic 40 kDa alpha subunits and two regulatory 25 kDa beta subunits (Glover et al. 1983) and has many known protein targets in *Drosophila* (Bourbon et al. 1995; Jaffe et al. 1997; Packman et al. 1997; Willert et al. 1997). Analytical gel filtration chromatography of the peak fraction of kinase activity showed an approximate size of 135 kDa for the kinase, consistent with a tetramer composed of two 40 kDa and two 25 kDa subunits.

Casein kinase II has a preferred recognition motif of *SXX(D/E) (Kemp and Pearson 1990). One of the identified phosphorylation sites, Ser 61, lies within this motif (*SGPE). We therefore hypothesized that casein kinase II was phosphorylating Ser 61. To confirm this, we made serine-to-alanine mutants at each phosphorylation site (Ser 42 and Ser 61) as well as a double mutant. We used both the peak activity fraction from the casein kinase II purification as well as purified recombinant human casein kinase II to phosphorylate these mutant PSI substrates *in vitro* (Figure 3C). The mutation of Ser 61 to alanine completely abolished the ability of casein kinase II to phosphorylate the substrate while the mutation of Ser 42 to alanine had little, if any, effect. Taken together, these data indicate that Ser 61 in PSI is a casein kinase II phosphorylation site.

PSI is phosphorylated in *Drosophila* cells by tousled-like kinase

Although the majority of the activity in the initial fractionation step resided in the 1M KCl fraction and was likely due to casein kinase II, we detected a smaller peak of activity in the 250 mM fraction. We further fractionated this peak of activity over several chromatographic columns (Figure 4A). After fractionation, we again visualized the peak activity fraction by silver stain (Figure 4B) and excised prominent bands from the gel and analyzed them by mass spectrometry. We identified tousled-like kinase (tlk) as a component of the prominent band migrating at 130 kDa. No other known protein kinases were identified in this fraction. A second, independent fractionation also identified tlk as the lone kinase in the final peak activity fraction.

We then tested the ability of the tlk-containing fraction to phosphorylate the serine-to-alanine mutant PSI substrates at Ser 42 and Ser 61 (Figure 4C). Although the tlk-containing fraction did efficiently phosphorylate the wildtype substrate, it also

Chapter 2: Phosphorylation of PSI

phosphorylated the two mutants. However, there are one threonine and fifteen serine residues in the PSI truncation substrate that may serve as alternate phosphorylation sites. No phosphorylation site could be found in the truncated PSI substrate by mass spectrometry, suggesting that the observed activity resulted in very low levels of modification and/or modification distributed over multiple sites.

PSI mutants show differential protein interaction profiles

PSI contains four KH domains (Amarasinghe et al. 2001), but the identified phosphorylation sites lie N-terminally to them. The phosphorylation sites are, however, near a very glycine-rich region. Glycine-rich regions are known to be mediators of protein-protein interactions, particularly in RNA binding proteins (Wang et al. 1997). We therefore hypothesized that the phosphorylation state of PSI may influence its ability to interact with other proteins.

To test this idea, we performed GST pulldowns in *Drosophila* Kc nuclear extract using GST-tagged wildtype and mutant N-terminal PSI truncations as bait (Figure 5A). As before, these truncations again consisted of the first 95 amino acids of PSI. We also performed the pulldowns using bait proteins that had been pre-phosphorylated by treatment with casein kinase II and ATP. We then analyzed the PSI-interacting proteins by mass spectrometry.

Interestingly, a 75 kDa protein interacted strongly with the wildtype and S61A PSI N-terminal truncations, but not with the S42A or S42A/S61A PSI truncations (Figure 5A). This protein also seemed to shift in mobility when casein kinase II had been added to the extract. Mass spectrometry analysis revealed several peptides of the 75 kDa Recombination Repair Protein 1 (Rrp1) in the wildtype and S61A pulldowns, but none in the S42A and S42A/S61A pulldowns. Interestingly, Rrp1 also contains two copies of the casein kinase II phosphorylation motif, indicating that its change in mobility when casein kinase II is added to the pulldown may be the result of its own phosphorylation by CKII.

To perform a more comprehensive analysis of the proteins that interact with PSI, we expressed epitope-tagged full-length PSI in S2 cells. We then immunoprecipitated the exogenous PSI using the polyoma (also known as Glu-Glu) epitope, eluted from this immunoprecipitation using free polyoma (EYMPME) peptide, and treated with RNase A. We then immunoprecipitated the eluate using a polyclonal anti-PSI antibody (Siebel et al. 1995) (Figure 5B, 5C). After eluting PSI from the antibody resin with acidified glycine, we analyzed the PSI-interacting proteins using mass spectrometry (Figure 5D). We identified several splicing factors among the interacting proteins, including snRNP70K, which had previously been shown to directly interact with the A/B domain of PSI (Labourier et al. 2001). These PSI-interacting proteins also included hrp48, a factor known to play a role in the P element splicing silencer (Siebel et al. 1994; Hammond et al. 1997), the hnRNP protein hrp59, the splicing factor PUF68, the RNA helicase Rm62

Chapter 2: Phosphorylation of PSI

and SR protein Srp54. Rrp1, identified as an interacting partner with truncated PSI, was present but with fewer peptides than proteins listed in Figure 5D. Interestingly, we also identified several cytoskeletal proteins, including actin and the alpha and beta subunits of tubulin. These proteins appear not to be non-specific contaminants, as they did not co-purify with non-specific rabbit IgG antibody done in parallel. Taken together these proteins appear to associate with PSI via protein-protein interactions in an RNase-insensitive manner and thus might be functioning along with PSI in the processing of specific nuclear pre-mRNAs.

DISCUSSION

We have identified two kinases that phosphorylate PSI. Although we could definitely show that the phosphorylation site for casein kinase II was Ser 61, we were unable to definitively show that Ser 42 was phosphorylated by tkk. As very little is known about its preferred motif (Pilyugin et al. 2009), it is difficult to assess whether tkk can phosphorylate Ser 42. Ser 42 is additionally in close proximity to several other serines, and thus its mutation may only shift phosphorylation to one or more nearby sites. However, given that we twice identified tkk as the only kinase present in a fraction that efficiently phosphorylated the PSI substrate, it is likely that tkk can phosphorylate PSI.

These phosphorylation events do not occur in or near the RNA-binding KH domains of PSI and are thus unlikely to modulate RNA-binding activity. They may, however, affect the protein interaction partners of PSI. Using GST pulldowns, we showed that Rrp1 requires the presence of Ser 42 to be able to interact efficiently with PSI. Rrp1 is an exonuclease that is involved in DNA damage repair (Szakmary et al. 1996). Connections between splicing and DNA repair have been previously recognized (Chaouki and Salz 2006). Interestingly, human tkk has also been implicated in DNA damage and repair (Carrera et al. 2003; Groth et al. 2003; Krause et al. 2003).

The interaction of PSI with several cytoskeletal proteins was unexpected. PSI is essentially exclusively nuclear (Labourier et al. 2002), and the immunoprecipitations and purifications performed here began with nuclear extracts. Both actin and tubulin, however, are known to exist in the nucleus (Menko and Tan 1980; Olave et al. 2002). Furthermore, certain transcriptional complexes are known to contain actin, and actin has been shown to interact with several hnRNP proteins (Zheng et al. 2009). The potential importance and role of the interaction of PSI with these cytoskeletal proteins will require further investigation as will a full understanding of the consequences of these two phosphorylation events in PSI.

MATERIALS AND METHODS

Purification of PSI

Chapter 2: Phosphorylation of PSI

PSI in pRSETa was grown in BL21 pLYS E cells to OD 0.4 at 37C, cooled to 16C, and induced with 0.5 mM IPTG overnight. The bacteria were spun down and resuspended in lysis buffer (20 mM HEPES pH 7.6, 10% glycerol, 1 M NaCl, 0.05% Tween, 0.5 mM DTT, 0.4 mM PMSF) and snap frozen in liquid nitrogen.

The lysate was then thawed, sonicated 4 times for 30 seconds each, and spun at 15000 RPM for 30 min to separate insoluble material. The soluble lysate was filtered and purified over a NiHiTrap 1 mL column using lysis buffer as the "A" buffer, and lysis buffer containing 500 mM imidazole as the "B" buffer. The His-tagged PSI was eluted from the column over 15 column volumes in 0.5 mL fractions.

Purification of hrp48

Hrp48 was purified in the same manner as PSI. Importantly, the bacterial cells were cooled to 16C before induction to promote the solubility of hrp48.

Purification of Polyoma-tagged full-length PSI from Kc cells

Nuclear extract from Kc cells was incubated with 50 μ L of anti-polyoma resin for 1 hr at 4°C. The resin was then washed three times with 1 mL IPB2 buffer (20 mM Tris pH 8.0, 2 mM EDTA, 400 mM NaCl, 0.2% NP-40, 1 mM dithiothreitol). The resin was then resuspended in 500 μ L of IPB2 and treated with 0.5 μ L of RNase A (Promega). The resin was then washed twice with 1 mL IPB2, and protein was eluted off the resin using elution buffer at 65°C (8M urea / 100 mM Tris pH 8.5). The samples were then prepared for mass spectrometry as described below.

Purification of N-terminal PSI truncation

A cDNA fragment containing amino acids 2-95 of PSI was cloned into pRSETA between the NdeI and KpnI restriction sites. Additionally, the cDNA contained a His6 tag on the N-terminal end and a polyoma (EYMPME) tag on the C-terminal end. The PSI truncation was expressed in BL21(DE3) pLYS E cells and purified using nickel affinity chromatography.

In vitro kinase assays

In vitro kinase assays contained the following: 5 μ L fraction to be assayed, 1 μ g PSI fragment, 0.5 μ L γ -32P-ATP (7000 Ci/mmol), 25 mM Hepes-KOH, pH 7.5, 0.5 mM EDTA, 5% glycerol, 5 mM MgCl₂, 100 mM KCl, 0.5 mM DTT, 0.4 mM PMSF, 10 mM NaF, 5 mM beta-glycerol-phosphate, 25 μ M ATP in a final volume of 50 μ L. The reactions were incubated at room temperature for 30 min. Ten μ L of protein sample buffer were added, and 15 μ L of the sample was then analyzed by SDS-PAGE. The gel was silver-stained, dried, and exposed to X-ray film for two hours to determine the chromatography fractions that contained PSI-phosphorylating activity.

Chapter 2: Phosphorylation of PSI

GST Pulldowns

cDNA fragments for each PSI mutant were cloned into pGEX-2TK and expressed as GST fusions in Rosetta(DE3) pLYS S E. coli cells. Lysate containing the overexpressed PSI fusion proteins was incubated with glutathione-sepharose beads for 3 hr at 4°C. The beads were then washed 3 times with Buffer A (50 mM Tris-HCl, pH 8.0, 1M NaCl, 0.5 mM DTT, 0.4 mM PMSF, 10% glycerol) and 2 times with 1X Casein Kinase II buffer (NEB P6010S). ATP was then added to all samples to 5 mM, and 1 uL purified human casein kinase II (NEB P6010S) was added to those samples that were to be phosphorylated. The samples were then incubated overnight at 30°C.

100 µL Kc nuclear extract was then added, and the beads were incubated at 4°C for 3 hrs. The beads were then washed three times with buffer B (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 0.02% NP-40, 0.5 mM DTT, 0.4 mM PMSF), and bound proteins were eluted by incubating with 50 µL elution buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10% glycerol, 20 mM glutathione, 0.5 mM DTT, 0.4 mM PMSF) at room temperature for 45 min. Bound proteins were analyzed by SDS-PAGE, silver-staining, and mass spectrometry.

PSI Immunoprecipitation

RNP extract (Pinol-Roma et al. 1990a; Pinol-Roma et al. 1990b) from *Drosophila* S2 cells stably expressing polyoma (also known as Glu-Glu) tagged PSI under the control of the metallothionein promoter was made after inducing expression of the transgene with 200 µM CuSO₄ for 36 hours. Briefly, cells were swelled in hypotonic buffer and nuclei were isolated using a dounce homogenizer. The nuclei were then sonicated, and the resulting lysate was passed over a 30% sucrose cushion to separate the light nucleoplasm from the dense chromatin-associated fraction.

The nucleoplasm was then collected and incubated at 4°C for 2 hrs with protein-A sepharose beads that had been pre-incubated with anti-polyoma antibody. The beads were then washed 3 times with wash buffer (20 mM Hepes-KOH pH 7.5, 200 mM KCl, 2 mM MgCl₂, 0.5 mM DTT, 0.05% NP-40, 0.4 mM PMSF). Proteins were then eluted from the resin by two incubations with 200 µL elution buffer (wash buffer supplemented with 100 µg/mL polyoma peptide (EYMPME)).

Elutions were then incubated for 1 hr at 4°C with protein-A Dynabeads (Invitrogen 100-01D) that had been pre-incubated with polyclonal anti-PSI antibody. As a control, the elutions were also incubated with protein-A Dynabeads that had been pre-incubated with rabbit IgG. The beads were then washed three times with wash buffer. Bound proteins were eluted off the beads in two elutions by incubating with 100 µL elution buffer (100 mM glycine, pH 2.5, 100 mM NaCl). The eluates from both the anti-PSI and rabbit IgG beads were analyzed by silver staining, western blotting, and mass

Chapter 2: Phosphorylation of PSI

spectrometry. Proteins identified as interacting with PSI by having at least 5 detected peptides are listed in Figure 5D.

Sample preparation and mass spectrometry for phosphopeptide analysis

Immunopurified PSI was adjusted to 40% methanol, 100 mM ammonium bicarbonate pH8.5, 5 mM TCEP, 1.5% ProteaseMAX (Promega) and subjected to carboxyamidomethylation of cysteines. Two samples were created and analyzed separately and then the results were combined. The first sample was digested with trypsin and chymotrypsin. The second sample was divided, and one portion was digested with a combination of trypsin and chymotrypsin. The other portion was digested with thermolysin. The two fractions were then recombined for analysis. All digestions were incubated overnight at 37°C and stopped by the addition of 5% formic acid. A 3 phase nano LC column was packed in a 100 µm inner diameter glass capillary with an emitter tip. The column consisted of 10 cm of Polaris C18 5 µm packing material (Varian), followed by 4 cm of Partisphere 5 SCX (Whatman), followed by another 2 cm of Polaris C18. The column was loaded by use of a pressure bomb and washed extensively with buffer A (see below). The column was then directly coupled to an electrospray ionization source mounted on a Thermo-Fisher LTQ XL linear ion trap mass spectrometer. Data collection was programmed so that neutral loss of phosphate would trigger the collection of an MS3 spectrum of the neutral loss peak. An Agilent 1200 HPLC equipped with a split line so as to deliver a flow rate of 30 nl/min was used for chromatography. Peptides were eluted using a 4-step MudPIT procedure (Washburn et al. 2001). Buffer A was 5% acetonitrile/ 0.02% heptafluorobutyric acid (HBFA); buffer B was 80% acetonitrile/ 0.02% HBFA. Buffer C was 250 mM ammonium acetate/ 5% acetonitrile/ 0.02% HBFA; buffer D was same as buffer C, but with 500 mM ammonium acetate. The programs SEQUEST and DTASELECT were used to identify peptides and proteins from the Drosophila database (A. J. M. Eng 1994; Tabb et al. 2002). Phosphopeptides were confirmed by manual inspection of the spectra.

Sample preparation and mass spectrometry for protein interaction analysis

The protein solution was adjusted to 8M urea, subjected to carboxyamidomethylation of cysteines, and digested with trypsin. The sample was then desalted using a C18 spec tip (Varian). A 2 phase nano LC column was packed and loaded as described above. The column consisted of 10 cm of Polaris C18 5 µm packing material (Varian), followed by 4 cm of Partisphere 5 SCX (Whatman). Chromatography, mass spectrometry and data analysis were as described above, except that no MS3 spectra were collected.

Mass spectrometry of gel bands

Excised gel bands were treated to cause carboxyamidomethylation of cysteines, digested with trypsin and the resulting peptides extracted. Samples were loaded on a

Chapter 2: Phosphorylation of PSI

100 μ M ID, 10 cm column of Polaris C18 and analyzed by LC-MS/MS with a linear gradient consisting of buffer A and buffer B as above.

REFERENCES

- A. J. M. Eng JY. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **5**: 976-989.
- Adams MD, Tarrng RS, Rio DC. 1997. The alternative splicing factor PSI regulates P-element third intron splicing in vivo. *Genes Dev* **11**: 129-138.
- Amarasinghe AK, MacDiarmid R, Adams MD, Rio DC. 2001. An in vitro-selected RNA-binding site for the KH domain protein PSI acts as a splicing inhibitor element. *RNA* **7**: 1239-1253.
- Anko ML, Muller-McNicoll M, Brandl H, Curk T, Gorup C, Henry I, Ule J, Neugebauer KM. 2012. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol* **13**: R17.
- Blanchette M, Green RE, Brenner SE, Rio DC. 2005. Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. *Genes Dev* **19**: 1306-1314.
- Boudrez A, Beullens M, Waelkens E, Stalmans W, Bollen M. 2002. Phosphorylation-dependent interaction between the splicing factors SAP155 and NIPP1. *J Biol Chem* **277**: 31834-31841.
- Bourbon HM, Martin-Blanco E, Rosen D, Kornberg TB. 1995. Phosphorylation of the Drosophila engrailed protein at a site outside its homeodomain enhances DNA binding. *J Biol Chem* **270**: 11130-11139.
- Carrera P, Moshkin YM, Gronke S, Sillje HH, Nigg EA, Jackle H, Karch F. 2003. Tousled-like kinase functions with the chromatin assembly pathway regulating nuclear divisions. *Genes Dev* **17**: 2578-2590.
- Chaouki AS, Salz HK. 2006. Drosophila SPF45: a bifunctional protein with roles in both splicing and DNA repair. *PLoS Genet* **2**: e178.
- Glover CV, Shelton ER, Brutlag DL. 1983. Purification and characterization of a type II casein kinase from Drosophila melanogaster. *J Biol Chem* **258**: 3258-3265.

Chapter 2: Phosphorylation of PSI

- Groth A, Lukas J, Nigg EA, Sillje HH, Wernstedt C, Bartek J, Hansen K. 2003. Human Tousled like kinases are targeted by an ATM- and Chk1-dependent DNA damage checkpoint. *EMBO J* **22**: 1676-1687.
- Hammond LE, Rudner DZ, Kanaar R, Rio DC. 1997. Mutations in the hrp48 gene, which encodes a Drosophila heterogeneous nuclear ribonucleoprotein particle protein, cause lethality and developmental defects and affect P-element third-intron splicing in vivo. *Mol Cell Biol* **17**: 7260-7267.
- Ignjatovic T, Yang JC, Butler J, Neuhaus D, Nagai K. 2005. Structural basis of the interaction between P-element somatic inhibitor and U1-70k essential for the alternative splicing of P-element transposase. *J Mol Biol* **351**: 52-65.
- Jaffe L, Ryoo HD, Mann RS. 1997. A role for phosphorylation by casein kinase II in modulating Antennapedia activity in Drosophila. *Genes Dev* **11**: 1327-1340.
- Kemp BE, Pearson RB. 1990. Protein kinase recognition sequence motifs. *Trends Biochem Sci* **15**: 342-346.
- Krause DR, Jonnalagadda JC, Gatei MH, Sillje HH, Zhou BB, Nigg EA, Khanna K. 2003. Suppression of Tousled-like kinase activity after DNA damage or replication block requires ATM, NBS1 and Chk1. *Oncogene* **22**: 5927-5937.
- Labourier E, Adams MD, Rio DC. 2001. Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. *Mol Cell* **8**: 363-373.
- Labourier E, Blanchette M, Feiger JW, Adams MD, Rio DC. 2002. The KH-type RNA-binding protein PSI is required for Drosophila viability, male fertility, and cellular mRNA processing. *Genes Dev* **16**: 72-84.
- Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. 2007. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**: 926-929.
- Laski FA, Rio DC, Rubin GM. 1986. Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. *Cell* **44**: 7-19.
- Menko AS, Tan KB. 1980. Nuclear tubulin of tissue culture cells. *Biochim Biophys Acta* **629**: 359-370.
- Misra S, Rio DC. 1990. Cytotype control of Drosophila P element transposition: the 66 kd protein is a repressor of transposase activity. *Cell* **62**: 269-284.
- Olave IA, Reck-Peterson SL, Crabtree GR. 2002. Nuclear actin and actin-related proteins in chromatin remodeling. *Annu Rev Biochem* **71**: 755-781.

Chapter 2: Phosphorylation of PSI

- Packman LC, Kubota K, Parker J, Gay NJ. 1997. Casein kinase II phosphorylates Ser468 in the PEST domain of the *Drosophila* I κ B homologue cactus. *FEBS Lett* **400**: 45-50.
- Pilyugin M, Demmers J, Verrijzer CP, Karch F, Moshkin YM. 2009. Phosphorylation-mediated control of histone chaperone ASF1 levels by Tousled-like kinases. *PLoS One* **4**: e8328.
- Pinol-Roma S, Choi YD, Dreyfuss G. 1990a. Immunological methods for purification and characterization of heterogeneous nuclear ribonucleoprotein particles. *Methods Enzymol* **181**: 317-325.
- Pinol-Roma S, Swanson MS, Matunis MJ, Dreyfuss G. 1990b. Purification and characterization of proteins of heterogeneous nuclear ribonucleoprotein complexes by affinity chromatography. *Methods Enzymol* **181**: 326-331.
- Rio DC, Laski FA, Rubin GM. 1986. Identification and immunochemical analysis of biologically active *Drosophila* P element transposase. *Cell* **44**: 21-32.
- Roth MB, Murphy C, Gall JG. 1990. A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *J Cell Biol* **111**: 2217-2223.
- Siebel CW, Admon A, Rio DC. 1995. Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes Dev* **9**: 269-283.
- Siebel CW, Fresco LD, Rio DC. 1992. The mechanism of somatic inhibition of *Drosophila* P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. *Genes Dev* **6**: 1386-1401.
- Siebel CW, Kanaar R, Rio DC. 1994. Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. *Genes Dev* **8**: 1713-1725.
- Siebel CW, Rio DC. 1990. Regulated splicing of the *Drosophila* P transposable element third intron in vitro: somatic repression. *Science* **248**: 1200-1208.
- Szakmary A, Huang SM, Chang DT, Beachy PA, Sander M. 1996. Overexpression of a Rrp1 transgene reduces the somatic mutation and recombination frequency induced by oxidative DNA damage in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **93**: 1607-1612.
- Tabb DL, McDonald WH, Yates JR, 3rd. 2002. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **1**: 21-26.

Chapter 2: Phosphorylation of PSI

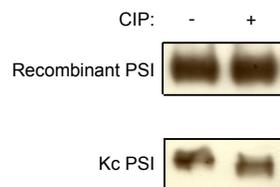
- Wang J, Dong Z, Bell LR. 1997. Sex-lethal interactions with protein and RNA. Roles of glycine-rich and RNA binding domains. *J Biol Chem* **272**: 22227-22235.
- Wang X, Bruderer S, Rafi Z, Xue J, Milburn PJ, Kramer A, Robinson PJ. 1999. Phosphorylation of splicing factor SF1 on Ser20 by cGMP-dependent protein kinase regulates spliceosome assembly. *EMBO J* **18**: 4549-4559.
- Washburn MP, Wolters D, Yates JR, 3rd. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**: 242-247.
- Willert K, Brink M, Wodarz A, Varmus H, Nusse R. 1997. Casein kinase 2 associates with and phosphorylates dishevelled. *EMBO J* **16**: 3089-3096.
- Xiao SH, Manley JL. 1997. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* **11**: 334-344.
- Xiao SH, Manley JL. 1998. Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *EMBO J* **17**: 6359-6367.
- Zhai B, Villen J, Beausoleil SA, Mintseris J, Gygi SP. 2008. Phosphoproteome analysis of *Drosophila melanogaster* embryos. *J Proteome Res* **7**: 1675-1682.
- Zheng B, Han M, Bernier M, Wen JK. 2009. Nuclear actin and actin-binding proteins in the regulation of transcription and gene expression. *FEBS J* **276**: 2669-2685.

Chapter 2: Phosphorylation of PSI

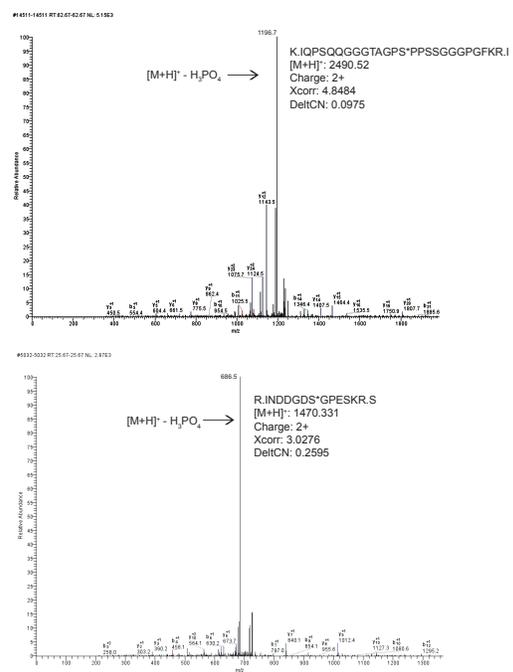
Figure 1. Biochemcial fractionation and analysis of the PSI kinase. A) Purified recombinant PSI and PSI purified from Kc cells was treated with calf intestinal phosphatase (CIP) and then visualized by immunoblotting. B) MS2 spectra identifying phosphopeptides found in PSI. B and Y series ions and neutral loss of phosphate are indicated. Inset: sequence of the phosphopeptide and SEQUEST statistics. MS3 spectra and corresponding spectra of unmodified peptides are given in Figure 4.

Chapter 2: Phosphorylation of PSI

A



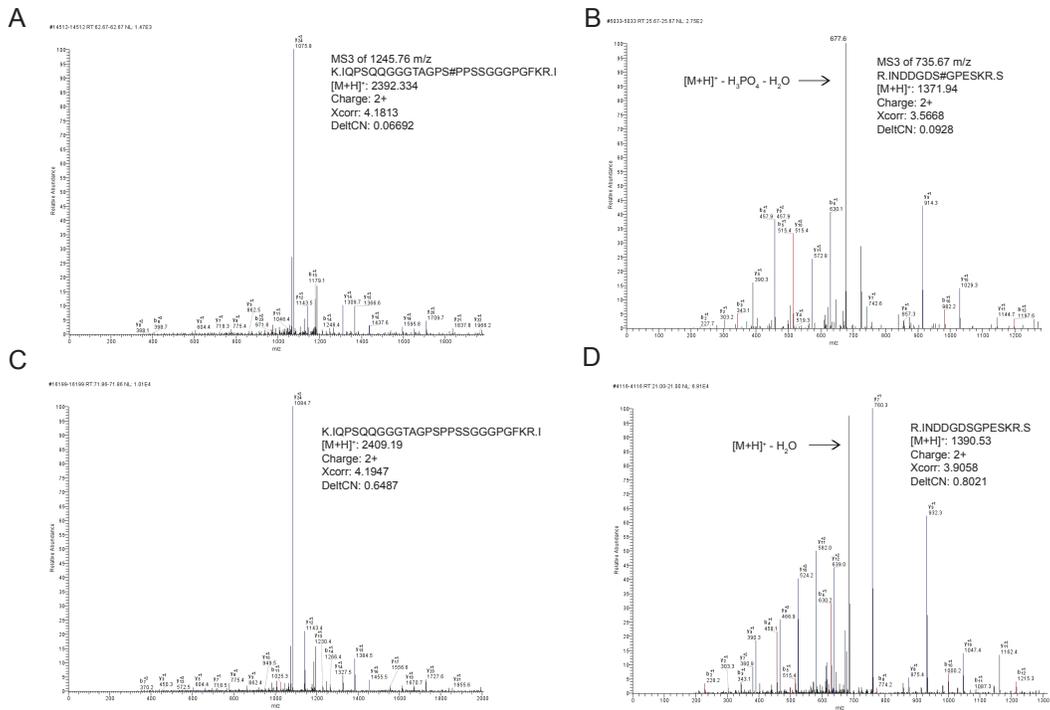
B



Chapter 2: Phosphorylation of PSI

Figure 2. Identification of phosphorylation sites by mass spectrometry. A and B) MS3 spectra of phosphopeptides. B and Y series ions and loss of water are indicated. Inset: sequence of the phosphopeptide and SEQUEST statistics. C and D) MS2 spectra of corresponding unmodified peptides.

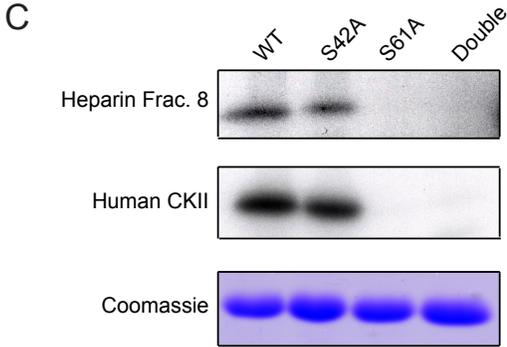
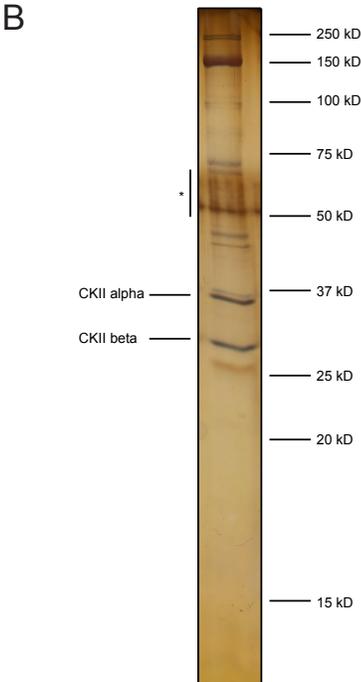
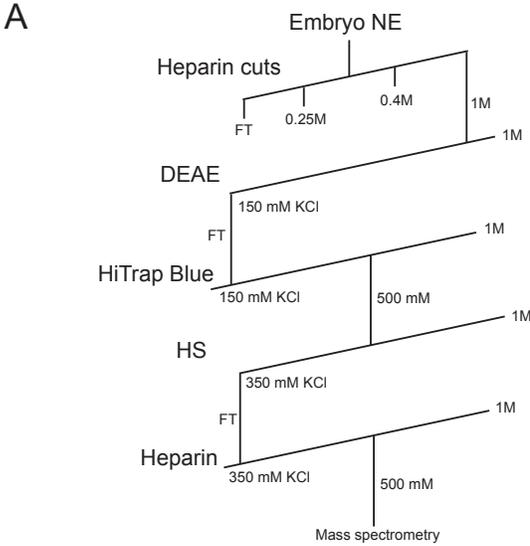
Chapter 2: Phosphorylation of PSI



Chapter 2: Phosphorylation of PSI

Figure 3. Biochemical purification of *Drosophila* casein kinase II. A) Purification strategy for endogenous casein kinase II. B) Protein composition of peak fraction of activity from the final heparin column visualized by SDS-PAGE and silver-staining. Species identified as casein kinase II alpha and beta are labeled. Bands labeled with an asterisk correspond to contaminating keratin. C) In vitro kinase assay of PSI mutant proteins. Serine to alanine PSI mutant proteins were phosphorylated in vitro using the peak fraction of activity from the final heparin column and using purified recombinant human casein kinase II (NEB P6010S). Assays were visualized using autoradiography, and, to ensure equal protein loading, coomassie staining.

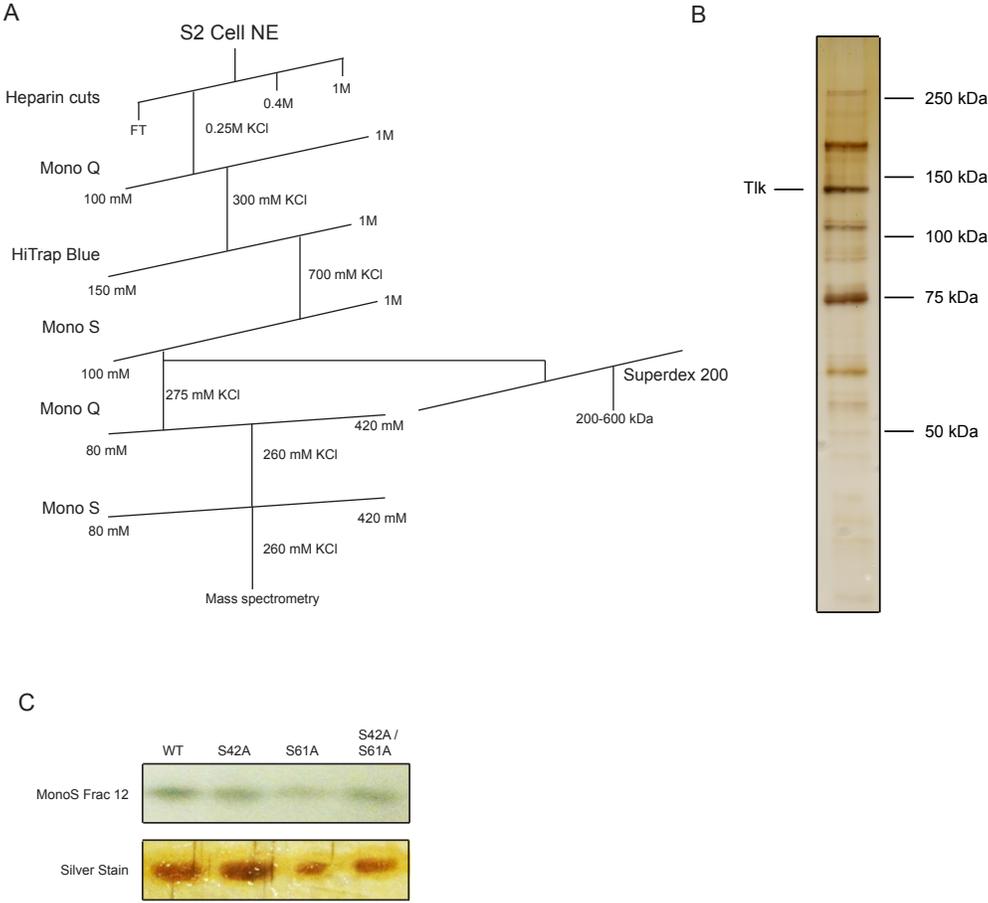
Chapter 2: Phosphorylation of PSI



Chapter 2: Phosphorylation of PSI

Figure 4. Purification of a second PSI kinase activity. A) Purification strategy for tousled-like (tlk) kinase. B) Protein composition of peak fraction activity from the final Mono S column visualized by silver staining. The species identified as tlk is labeled. C) In vitro kinase assay of PSI mutants. Serine to alanine PSI mutants were phosphorylated in vitro using the peak fraction of activity from the final Mono S column. Assays were visualized using autoradiography, and, to ensure equal protein loading, silver staining.

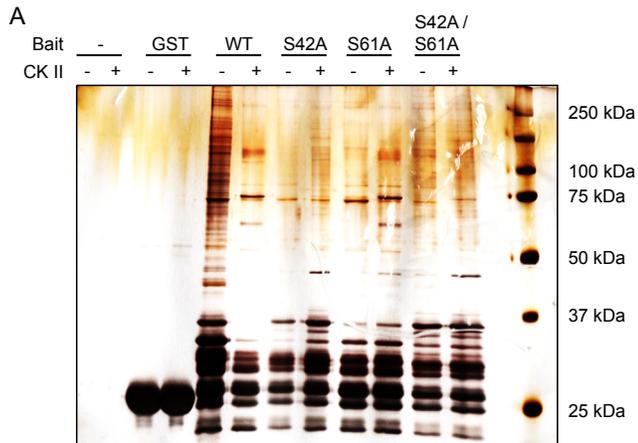
Chapter 2: Phosphorylation of PSI



Chapter 2: Phosphorylation of PSI

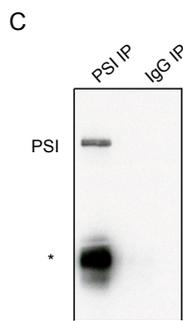
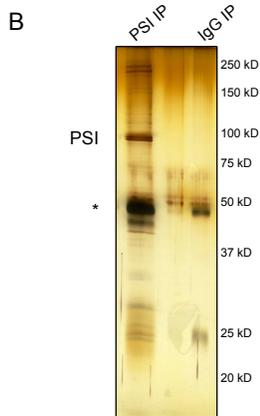
Figure 5. Protein-protein interactions of PSI. A) GST pulldown assay using PSI mutant proteins. GST-PSI fusion proteins carrying the serine to alanine PSI mutations were phosphorylated using purified human casein kinase II and incubated with Kc nuclear extract. The resulting glutathione resin eluates were analyzed by silver staining and mass spectrometry. B) Silver stain of PSI and interacting proteins following anti-polyoma and anti-PSI immunoprecipitations. The asterisk indicates antibody heavy chain. C) Immunoblot analysis of (B) using anti-PSI antibody. D) Mass spectrometry analysis of (B). Proteins identified as interacting with PSI and the number of peptides observed for each protein are listed.

Chapter 2: Phosphorylation of PSI



D

Protein	Peptides
PSI	94
betaTub56D	55
B52	48
Act5C	30
SF2	30
Act57B	26
betaTub85D	18
betaTub60D	16
betaTub97EF	15
hrp48	12
Lam	10
alphaTub85E	9
hrp59	9
zip	9
His4r	7
pUF68	7
Rm62	7
snRNP70K	7
CG7185	7
alphaTub84D	6
Srp54	6
YT521	6
His 3.3B	5
SmD3	5
Hsc70-3	5



SUMMARY

Transcription and pre-mRNA alternative splicing are highly regulated processes that play major roles in modulating eukaryotic gene expression. It is increasingly apparent that other pathways of RNA metabolism, including small RNA biogenesis, can regulate these processes. However, a direct link between alternative pre-mRNA splicing and small RNA pathways has remained elusive. Here we show that the small RNA pathway protein Argonaute-2 (Ago-2) regulates alternative pre-mRNA splicing patterns of specific transcripts in the *Drosophila* nucleus using genome-wide methods in conjunction with RNAi in cell culture and Ago-2 deletion or catalytic site mutations in *Drosophila* adults. Moreover, we show using ChIP-seq that nuclear Argonaute-2 binds to specific chromatin sites near gene promoters and negatively regulates the transcription of the Ago-2-associated target genes, which are distinct from the genes encoding the pre-mRNA splicing target transcripts. These transcriptional target genes are also bound by Polycomb group (PcG) transcriptional repressor proteins and change during development, implying that Ago-2 may regulate *Drosophila* development. Importantly, both of these activities were independent of the catalytic activity of Ago-2, suggesting a new role for the protein in the nucleus. Finally, we determined the nuclear transcriptome RNA-binding profile of Ago-2 using CLIP-seq, found that Ago-2 bound to several splicing target transcripts and identified a G-rich RNA binding site for Ago-2 that was enriched in these transcripts. Taken together, these results suggest two different nuclear roles for Ago-2: one in pre-mRNA splicing and one in PcG-mediated transcriptional repression, thus identifying new functions for Argonaute-2 in nuclear RNA polymerase II transcription regulation and pre-mRNA processing that extend beyond its well-established roles in the canonical cytoplasmic RISC-related small interfering RNA pathways.

INTRODUCTION

Small RNAs have been known for some time to play vital roles in RNA-mediated regulation, beginning with the discovery of their role in developmental regulation in *C. elegans* (Lee et al. 1993; Wightman et al. 1993). In *Drosophila*, there is a functional separation between the pathway that produces microRNAs, which tend to act without cleaving the target mRNA, and the pathway that produces siRNAs, which tend to involve cleavage of the target endogenous or viral mRNA (Forstemann et al. 2007). This separation is based on the level of complementarity between the strands of the double-stranded RNA (Tomari et al. 2007).

Although Ago-1 bound microRNAs arise from specific genomic loci, Ago-2 bound small interfering RNAs (siRNAs) can originate from several different sources. Viral RNA replication intermediates, convergent transcription events, and complementary regions

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

in different transcripts can all lead to the formation of functional double-stranded RNA that can be cleaved by Dicer-2 and loaded into Ago-2 (Czech et al. 2008; Kawamura et al. 2008). This opens up a large part of the genome to potential small RNA-mediated regulation.

A few examples of splicing regulation through small RNAs have already been reported. The splicing of the serotonin receptor 2C gene in human tissue culture cells has been shown to be affected by the snoRNA HBII-52 in a base-pairing dependent manner between the snoRNA and the serotonin receptor 2C pre-mRNA (Kishore and Stamm 2006). This principle has also been used in a therapeutic setting to modulate the splicing of the mouse SMN2 gene, whose ortholog in humans is inefficiently correctly spliced (Hua et al. 2011). Using antisense oligos targeting the SMN2 pre-mRNA, the productive splicing efficiency could be increased, leading to the production of more functional SMN2 protein (Hua et al. 2010; Hua et al. 2011).

Previous work in other organisms has also demonstrated links between chromatin state, histone marks, and alternative splicing (Spies et al. 2009). These studies showed that these effects are mediated through recruitment of factors to specific histone marks (Luco et al. 2010) or by modulating the rate of transcription through certain chromatin regions (Batsche et al. 2006). Some of these effects were also shown to be dependent on Argonaute-1 and siRNA-mediated transcriptional gene silencing in human cells (Allo et al. 2009). Links between RNAi components and chromatin structure have also been made in other systems, including *Drosophila* and *S. pombe* (Buhler and Moazed 2007; Pushpavalli et al. 2012).

It has been shown that Argonaute family members in humans and *C. elegans* can bind throughout the length of pre-mRNA transcripts (Chi et al. 2009; Zisoulis et al. 2010; Leung et al. 2011). Interestingly, there are also hints of a small RNA-independent mode of mouse Argonaute-2 mRNA binding (Leung et al. 2011).

An interplay between the splicing and small RNA machineries has been shown to exist in *Drosophila* cells. In a genome-wide screen, several splicing factors, including PSI and snRNP70K were shown to have an influence on the production or efficacy of either microRNAs or siRNAs (Zhou et al. 2008). We now ask the reciprocal question, that is: do the components of the small RNA pathway have an effect on the output of the splicing machinery? We find that Ago-2, but not Dicer-2, regulates many splice junctions in *Drosophila*. Furthermore, we find that Ago-2 binds to chromatin at the promoter regions of genes, acts as a transcriptional repressor in conjunction with PcG proteins, and that this repression is independent of its catalytic slicer activity. Finally, we use CLIP-seq on nuclear extract to determine the nuclear RNA binding profile of Ago-2 in *Drosophila* tissue culture cells and detect interaction of Ago-2 with some of the splicing-affected target pre-mRNAs.

RESULTS

***Drosophila* Argonaute-2 regulates nuclear alternative pre-mRNA splicing patterns in S2 cells**

We investigated the effects that four *Drosophila* small RNA-related proteins, Dicer-1 (Dcr-1), Dicer-2 (Dcr-2), Argonaute-1 (Ago-1) and Argonaute-2 (Ago-2) had on pre-mRNA splicing patterns in *Drosophila* tissue culture cells by first using RNAi to knockdown expression of each protein and then analyzing the resulting RNA by splice junction microarray (Blanchette et al. 2005; Blanchette et al. 2009; Taliaferro et al. 2011). Importantly, RNAi knockdown of components of the RNAi pathway themselves was possible and efficient, as evidenced by knockdown efficiencies of greater than 90% (Figure 1A). The knockdown of Dcr-1, Dcr-2, and Ago-1 had very modest effects on alternative splicing and only resulted in changes in splicing efficiency of 12, 4, and 2 splice junctions, respectively. By contrast, the knockdown of Ago-2 resulted in changes at 116 junctions (Figure 2A), a number similar to the splicing changes seen upon knockdown of known splicing factors (Blanchette et al. 2005; Blanchette et al. 2009). Several of these splicing changes were validated using semi-quantitative RT-PCR (Figure 1B).

We then set out to address if the knockdown of Ago-2 was leading to changes in splicing through a direct or indirect mechanism. If Ago-2 directly affects pre-mRNA splicing, we reasoned that Ago-2 must at least be partially nuclear-localized. We created a *Drosophila* S2 cell line that expressed FLAG-tagged Ago-2 under the control of an inducible promoter. We induced expression of this tagged Ago-2 such that the expression of the transgene was very low compared to the expression of endogenous Ago-2 (data not shown). We then fractionated extract from these cells into three components: cytoplasm, nucleoplasm, and chromatin-associated (Figure 2B). We detected, by immunoblotting with an anti-FLAG antibody, localization of Ago-2 in all three subcellular fractions.

To determine if Ago-2 binds the splicing target transcripts shown by the splice junction microarray *in vivo*, we crosslinked RNA-protein complexes using 0.05% formaldehyde and then immunoprecipitated complexes containing Ago-2 using a specific monoclonal antibody (Miyoshi et al. 2005). We detected an enrichment of transcripts whose splicing was affected by Ago-2 in the Ago-2 immunoprecipitations compared to the input and control immunoprecipitations (Figure 1C, Figure 2C).

Data from the splice junction microarrays was also used to determine the change in RNA expression level for 84% of *Drosophila* genes upon Ago-2 knockdown (Figure 1D). Of the 51 genes whose mRNA expression levels changed, only one, *hrp36*, was a known splicing factor. The mRNA expression of *hrp36* decreased approximately 2-fold upon Ago-2 knockdown. However, the overall change in expression of all genes upon Ago-2 knockdown was unlike that observed upon knockdown of any other tested

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

splicing factor, including *hrp36*, demonstrating that Ago-2 most likely mediates changes in splicing independent of its role in modulating *hrp36* mRNA levels (Figure 2D).

Argonaute-2 null mutants show defects in adult pre-mRNA splicing patterns

We took advantage of two Ago-2 mutant fly strains to investigate the role Ago-2 has in splicing in adult flies. One strain, Ago-2^{51B}, was an Ago-2 null mutant caused by an imprecise P element excision that deleted the start codons for Ago-2 (Xu et al. 2004). The other, Ago-2^{V966M}, was a point mutant that is deficient in catalytic slicer activity (Kim et al. 2007). Homozygous Ago-2^{51B} flies showed no detectable expression of Ago-2 by immunoblotting while homozygous Ago-2^{V966M} flies showed expression of Ago-2 to a level similar to that seen in w¹¹¹⁸ control flies (Figure 3A).

To determine differences in pre-mRNA splicing patterns in the Ago-2 mutant strains, we used high-throughput mRNA-seq to analyze the mRNA populations present in 0-16 hour post-eclosion males from homozygotes and heterozygotes from both strains. After filtering and mapping, we obtained between 58 and 95 million mapped reads for each strain (Figure 4A). Consistent with the immunoblot data, the level of Ago-2 mRNA transcripts was dramatically reduced in the homozygous Ago-2^{51B} sample (Figure 4B). We then used JuncBASE to determine the differential splicing patterns present in each sample (Brooks et al. 2011). The percent spliced in (Ψ) value for every splice junction in each sample was computed (Venables et al. 2008). As a reference, we used Ψ values from the Ago2^{51B} heterozygote, and any Ψ value for a particular sample and junction which was at least 5 above or below the reference value for that junction was identified as being significantly changed. Using these criteria, we observed 334, 173, and 155 significantly changing splice junctions in the Ago-2^{51B} homozygote, Ago-2^{V966M} heterozygote, and Ago-2^{V966M} homozygote samples, respectively (Figure 3B). We observed almost twice as many splicing changes in the homozygous Ago-2 heterozygous deletion mutant than in any other sample, including the homozygous catalytic mutant. This is in spite of the fact that the homozygous deletion mutants were siblings of the reference samples, while the catalytic mutants were not related to the reference samples. This is also consistent with the splice junction microarray data in which many junctions were sensitive to Ago-2 knockdown, but only a few splice junctions were sensitive to knockdown of the siRNA-producing enzyme Dcr-2 (Figure 2A). Importantly, no known splicing factor showed a significant change in RNA expression between any of the four samples (Figure 4C). Selected splicing changes were verified using semi-quantitative RT-PCR on RNA collected from adult males of the appropriate genotype (Figure 3C).

The chromatin association of Argonaute-2 is distinct from its regulation of pre-mRNA splicing

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

Recent reports have implicated chromatin structure and the presence of chromatin-associated factors in the regulation of alternative splicing (Batsche et al. 2006; Spies et al. 2009; Luco et al. 2010), and Argonaute proteins have been known to affect chromatin structure (Buhler and Moazed 2007; Pushpavalli et al. 2012). To address whether a similar mechanism was responsible for the observed Ago-2-mediated splicing changes, we performed ChIP-seq on endogenous Ago-2 in *Drosophila* S2 cells. The ChIP-seq immunoprecipitations were performed on biological replicates using a previously characterized antibody (Miyoshi et al. 2005) and were specific for the precipitation of Ago-2 (Figure 5A). After filtering for quality, approximately 20 million reads mapped uniquely for each experimental replicate (Figure 5B), and there was a high degree of overlap between the replicates (Figure 5C).

Most peaks of Ago-2 chromatin binding occurred at promoters (Figure 6A). This ChIP-seq data from embryonically-derived S2 cells was also compared to ChIP-CHIP data for Ago-2 from larvae that was produced by the modENCODE consortium (Celniker et al. 2009)(Figure 6A). This comparison showed a high degree of overlap of Ago-2 bound genomic regions, but also a significant number of differences, possibly indicating the changing occupancy of these regions by Ago-2 during development. Furthermore, this promoter enrichment was even more pronounced when considering only those genes whose mRNA expression levels changed upon Ago-2 knockdown in the splice junction microarray experiment (Figure 6B, red). However, Ago-2 chromatin occupancy did not correlate with whether or not splicing of target transcripts was regulated by Ago-2 (Figure 6B, purple).

Since Ago-2 has no known DNA binding domains, we reasoned that Ago-2 localization to chromatin may be due to its interaction with one or more recruiting factors. Previous studies had documented a physical interaction between Ago-2 and the chromatin binding factor CP190, as well as a co-localization on chromatin of Ago-2 with some of the trithorax and Polycomb group (PcG) proteins (Moshkovich et al. 2011). We again used ChIP-CHIP data from the modENCODE consortium (Celniker et al. 2009) to compare our observed Ago-2 ChIP data in S2 cells with that of 20 other proteins (Figures 5D and 6C). Ago-2 binding to chromatin co-localized strongly with the binding of several other factors, including many PcG group genes. We detected the enrichment of several sequence motifs in Ago-2 bound regions, including motifs that closely match the known binding sites for *Drosophila* CTCF (Cuddapah et al. 2009) (Figure 5E).

Argonaute-2 functions in transcriptional repression

In addition to changes in pre-mRNA splicing patterns, many changes in gene expression were observed in the Ago-2 mutant adult male mRNA-seq samples. Generally, there was considerably greater misregulation of RNA expression in the homozygous Ago-2 deletion mutant than in the Ago-2 catalytic mutant (Figure 7A). This is consistent with the greater misregulation of splicing in the deletion mutant and with

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

the idea of Ago-2 having nuclear roles beyond its catalytic RNA slicer activity (Moshkovich et al. 2011). Consistent with a previous report (Cernilogar et al. 2011), we observed a general up-regulation of heat shock proteins in the homozygous deletion mutant. However, this up-regulation did not occur in the homozygous catalytic mutant (Figure 8A).

In order to determine the genes whose expression was sensitive to the presence of Ago-2 and not to its catalytic activity, we took the following approach: we began with the genes whose FPKM values were significantly different between the homozygous and heterozygous deletion mutant samples (Trapnell et al. 2010) and subtracted the genes that also showed significant changes between the homozygous and heterozygous catalytic mutant samples (Figure 8B). To control for the slight differences in genetic background between the two mutants, we then subtracted the genes that showed significant changes between the two heterozygous samples (Figures 8C and E). Finally, if Ago-2 has RNA regulatory roles beyond its catalytic activity, it would be expected to find a high degree of similarity in the patterns produced in the heterozygous deletion mutant and homozygous catalytic mutant, because both samples contain “structurally” intact Ago-2. We observed such an overlap containing 483 genes (Figures 7B and 8D).

Of these 483 genes, 53 of them were among the 968 bound at the chromatin level by Ago-2 in S2 cells. This represents a modest but significant overlap between Ago-2 bound genes and those whose expression is sensitive to Ago-2 (p value = 0.001, Chi-square with Yates correction). Of the 53 bound and regulated genes, RNA levels for 49 of them are greater in the homozygous than the heterozygous Ago-2 deletion mutant (Figures 7C, 8F and 8G). Considering that Ago-2 is chromatin-bound at these genes, it is likely that Ago-2 is acting as a transcriptional repressor in these cases.

Using the splice junction microarray, we were also able to visualize switching events between alternative promoters upon Ago-2 knockdown. There were cases where we detected significant Ago-2 chromatin binding centered over one of these alternative promoters (Figure 8H). Additionally, usage of the bound promoter increased upon Ago-2 knockdown. Taken together, these data extend the previously observed Ago-2 chromatin binding to identify a subset of genes whose promoters are bound and repressed by Ago-2, thus implicating Ago-2 directly in transcriptional repression.

Argonaute-2 binds RNA throughout the transcriptome

In order to directly determine the RNA molecules bound by Ago-2 in the *Drosophila* nucleus, we performed CLIP-seq on UV crosslinked nuclear extract from S2 cells (Licatalosi et al. 2008) using a homemade polyclonal antibody (Figures 9A and B). Importantly, we sequenced pre-mRNAs and mRNAs, not siRNAs, bound to Ago-2, since the vast majority of CLIP tags were between 30 and 40 nt long. After using randomized barcodes to remove potential PCR duplicates (Konig et al. 2010), we aligned the reads

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

to the repeat masked *Drosophila* genome and determined significantly enriched regions of tag density (Webb, Kudla, Tollervey and Granneman, in preparation). We identified 837 clusters, of which 811 overlapped with annotated genomic features with 93% of those mapping to the mRNA sense strand. The majority of these clusters overlapped with protein coding genes (Figure 10A), though we also observed significant overlap with tRNAs and snoRNAs (Figures 9C and D) (Taft et al. 2009; Haussecker et al. 2010).

In some cases, we observed overlap between Ago-2 splicing targets identified by the splice junction microarray and Ago-2 CLIP-seq clusters. As an example, we observed a cluster of CLIP tag reads immediately next to the affected cassette exon in the *dco* gene (Figure 10B).

We also detected significantly-enriched sequence motifs in the clusters, most of which were G-rich (Figure 10C). This is consistent with a previous report suggesting that mouse Ago-2 may have a specific preference for G-rich motifs, independent of whether it is loaded with a small RNA (Leung et al. 2011). Tags containing the most enriched motif, GGCGG, were likely to have mutations, possible indicators of the site of crosslinking (Hafner et al. 2010; Konig et al. 2010), immediately next to the motif (Figure 10D). The 25th and 100th most enriched motifs, CCGAT and CAGCC, respectively, showed no such enrichment of mutations. We then created a motif composed of the top four enriched sequences in the CLIP clusters, weighted by their prevalence in the clusters (Figure 9F). This motif was again very G-rich. When we searched for the presence of this motif near splice sites, we detected an enrichment of the motif near Ago-2-sensitive splice sites relative to those splice sites that were insensitive to Ago-2 (Figure 10E). This enrichment seemed to be contained in the flanking exons and not within the affected intron (Figure 9G).

DISCUSSION

Although mammalian argonaute proteins are primarily known for their cytoplasmic functions, it has become clear that they can spend at least part of their time in the nucleus (Figure 2B) (Weinmann et al. 2009), and that once in the nucleus they can perform important functions relevant to chromatin formation and transcriptional silencing (Verdel et al. 2004; Janowski et al. 2006). We have now expanded that nuclear repertoire to *Drosophila* and shown that it includes direct transcriptional repression and regulation of alternative pre-mRNA splicing.

Previous reports have linked Ago-2 to higher-order chromatin structures, insulator function through interaction with CP190 and CTCF and to transcriptional silencing of heat-shock loci through small RNAs in *Drosophila* (Cernilogar et al. 2011; Moshkovich et al. 2011). Indeed, the studies on the effects of Ago-2 on the heat shock loci used Ago-2-associated small RNAs and suggest a role for Ago-2 in RNA polymerase II pausing or elongation. However, the transcriptional repression we report here is independent of the catalytic activity of Ago-2, but is related to its chromatin-binding. It is

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

unlikely that Ago-2 is directly contacting the DNA to mediate this repression since Ago-2 has no known DNA binding domains. Formaldehyde is an efficient protein-protein crosslinker, and perhaps interaction with other proteins is required to mediate association with DNA. Although the enrichment we observed in the CHIP-seq reads was modest, the binding sites are quite consistent with those previously reported (Moshkovich et al. 2011) and can be readily connected to repression of mRNA steady-state levels in Ago-2 null mutants, but not in Ago-2 catalytic mutants (Figure 7D).

If Ago-2 is not directly binding DNA, it is likely colocalizing with other chromatin-associated DNA binding factors, such as the Polycomb group (PcG) proteins, as we have observed. In fact, previous reports have shown that members of the RNAi machinery are required for proper PcG function in *Drosophila* cells (Grimaud et al. 2006). The fact that the overwhelming majority of direct transcriptional targets of Ago-2 were repressive events seems to fit nicely with the PcG complex being a repressive chromatin complex. In fact, 80% of these direct transcriptional targets are bound by at least one Polycomb protein in addition to Ago-2 ($p = 2.3 \times 10^{-7}$, Fisher's exact). This repression may also be developmentally regulated since the locations of Ago-2 chromatin binding change slightly between embryonically-derived S2 cells and larvae.

Small RNAs affecting splicing patterns have been reported (Kishore and Stamm 2006; Kishore et al. 2010). In particular, snoRNAs have been implicated in changing alternative splicing outcomes. We observed Ago-2 strongly binding to several snoRNAs in our CLIP assays, indicating one potential mechanism for the action of Ago-2 on alternative splicing.

We also observed Ago-2 binding to long stretches of RNA across many gene transcripts in CLIP-seq assays. This was qualitatively different from what has been reported for the binding patterns for other splicing factors (Huelga et al. 2012). Although bound genes were enriched for being more highly expressed (Figure 9E), many highly expressed genes were not bound at all, and control CLIP experiments did not yield the same extended patterns of binding. These findings suggest that the observed extended binding of Ago-2 across many transcripts may be legitimate. Further experimental work will be needed to determine the significance of this observation, as well as a unified mechanism for the action of Ago-2 on alternative splicing and its interactions with other splicing factors and spliceosome components.

MATERIALS AND METHODS

Polyclonal Ago-2 antibody

MBP-Ago-2delQ was received from the lab of Rachel Green. This construct contained MBP fused to amino acids 403-1214 of *Drosophila* Ago-2. The plasmid was transformed into BL21 pLYSE cells, and 1 L was grown at 37C to OD = 0.6, then cooled to 16C, induced with 1 mM IPTG, and left overnight. The bacteria were then spun down

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

and resuspended in 20 mL lysis buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 0.4 mM PMSF), frozen, thawed, and sonicated 3 times for 30 seconds. NP-40 was then added to 0.1%, and the lysate was spun at 15K RPM for 30 min. The soluble supernatant was then applied to 2 mL amylose resin. The resin was washed with 30 mL column buffer (20 mM Tris pH 7.5, 200 mM KCl, 1 mM EDTA, 0.5 mM DTT, 10% glycerol) and eluted in ten 1 mL fractions of elution buffer (column buffer supplemented with 20 mM maltose).

The resulting protein was reasonably clean, with the ~135 kDa fusion protein being by far the most abundant species. This protein was then mixed with purified *Pseudomonas* exotoxin, and injected into rabbits by Josman, LLC. Two further booster injections were performed, and a total of seven bleeds were made. The resulting serum was then purified using MBP-Ago-2delQ crosslinked to cyanogen bromide-activated sepharose.

RNAi and splice-junction microarrays

S2 cells were treated with 400 bp dsRNA targeting each protein for 4 days. dsRNA molecules were designed using SnapDragon (http://www.flyrnai.org/cgi-bin/RNAi_find_primers.pl) to minimize potential off-target effects. After 4 days, total RNA was recovered. Knockdown efficiencies were monitored by western blotting (Figure 1A).

For each microarray hybridization, 1 μ g of total RNA from the experimental knockdown and 1 μ g of total RNA from a nonspecific knockdown were amplified and converted to aRNA using the MessageAmp II aRNA Amplification kit following the manufacturer's recommendations (Ambion) and labeled with Cy5 and Cy3 monoreactive dye, respectively (GE Healthcare). The custom splicing-sensitive microarray used was based on FlyBase version 5.15 and interrogates 48,550 annotated splicing events from 11,368 different genes with three overlapping 36-nt oligonucleotide probes: one centered at the splice junction, and two probes offset by 3 nt on each side of the splice junction. In addition, one fully exonic probe per 100 nt of each mRNA, on average, were added. The 389,068 different probes were distributed randomly onto custom Agilent 500K arrays and used for hybridization of each cDNA sample. The microarrays were then processed and scanned following the manufacturer's recommendations (Agilent Technologies). The feature extraction reports were loaded into R (<http://www.r-project.org>) and Lowess-normalized using the marray package (Gentleman et al. 2004; Smyth 2004). The genes with affected alternative splicing were first identified using ANOVA, comparing the group of exonic probes common to all transcripts with the different groups of splice junction probes corresponding to every splicing event of a given gene. The genes with Q-values < 0.001 (adjusted using Benjamini-Hochberg correction) were then subjected to t-tests to identify the group of junction probes significantly affected with a P-value of < 0.001.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

To determine the degree of overlap in regulation between RNA binding proteins, Pearson correlations of the net expression of all genes were compared after RNAi knockdown of each protein.

RNA immunoprecipitation (RIP) of predicted Ago-2 target pre-mRNAs

Endogenous Ago-2 protein was immunoprecipitated from S2 nuclear RNP-enriched extracts that were crosslinked with 0.05% formaldehyde (Pinol-Roma et al. 1990). The extracts were stored in HNEB2 (10 mM Hepes-KOH pH 7.6, 100 mM KCl, 2.5 mM MgCl₂, 0.2% NP-40, 0.2 mM PMSF). 900 µL of extract was incubated with protein-A dynabeads (Invitrogen) that had been pre-incubated with monoclonal anti-Ago-2 antibody (Miyoshi et al. 2005). As a control, an immunoprecipitation using protein-A dynabeads pre-incubated with IgG was also done. The reaction was incubated for 4 h at 4°C. The beads were then washed four times with 1 mL of wash buffer (20 mM HEPES pH 7.6, 500 mM KCl, 5 mM MgCl₂, 0.5 mM PMSF, 50 U / mL RNasin (Promega)). The beads were then resuspended in 100 µL of 1X RQ1 DNase buffer (Promega). Five units of RQ1 DNase I (Promega) was then added and the reaction was incubated for 1 h at 37°C. Crosslinks were reversed by adding EDTA to 20 mM and incubating at 65°C for 1 h. RNA was then eluted by phenol/chloroform extracting the beads and ethanol precipitating. The pellet was washed twice with 1 mL of 70% ethanol and resuspended in 15 µL water. RNA concentration was measured using a Nanodrop spectrophotometer. Equal amounts of Ago-2 and IgG-immunoprecipitated RNA and RNA isolated from the starting RNP-enriched extracts were then used for RT-PCR using random hexamers. Individual bound transcripts were then assayed using HotStart PCR (Qiagen) and gene-specific primers.

mRNA-seq from adult files

Total RNA was collected by Trizol extraction from five 0-16 hr post-eclosion male adult files. RNA was collected from homozygotes and heterozygotes from the following strains: *w¹¹¹⁸* ; + ; Ago-2^{V966M}/TM3, *twi-GFP*, *Sb*, *Ser* and *yw* ; + ; Ago-2^{51B}/TM3, *Ser*, *y⁺*. Poly A⁺ RNA was then selected twice by incubating 10 µg RNA with oligo-dT dynabeads (Invitrogen). Starting with 200 ng poly A⁺ RNA, then, mRNA-seq libraries were made by first chemically fragmenting the RNA (Ambion AM8740) and constructing a library using the NEBNext mRNA Library Prep Reagent Set for Illumina (NEB E6100S). Duplicate libraries were made from separate RNA preps. Samples were barcoded and sequenced using 50 bp single-end sequencing on an Illumina HiSeq2000. Reads were de-multiplexed and aligned to the *Drosophila* genome using Tophat 1.4.0 (Trapnell et al. 2009) and the following options: `-a 6 -i 40 -coverage-search -butterfly-search -segment-mismatches 2`. We obtained between 58 and 97

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

million reads for each sample, and ~97% of those reads mapped uniquely. Of those reads, ~5% crossed known splice junctions.

Transcript abundances were then calculated using Cufflinks (Trapnell et al. 2010) and the following options: -u -m 280 -N -min-intron-length 40. Output from Cufflinks and Cuffdiff was visualized using cummeRbund (<http://compbio.mit.edu/cummeRbund/index.html>).

Alternative splicing analysis from RNA-seq data was performed using JuncBASE (Brooks et al. 2011). Briefly, Percent Spliced In (Ψ) values were calculated for each splice junction in each sample using both reads that crossed splice junctions and exonic reads that were specific to one particular outcome of a splicing event. A virtual reference of PSI values was created from the median PSI value of all four samples. PSI values in a particular sample that deviated from this reference PSI value by at least 5 were then marked as significantly changing splicing events. In all, 1039 total splicing events were recorded as significantly changing in at least one sample.

Ago-2 chromatin immunoprecipitation sequencing (ChIP-seq)

Briefly, S2 cells were crosslinked with 2% formaldehyde for 1 min. The remaining formaldehyde was then quenched with 250 mM glycine, pH8.0. Cells were then lysed in SDS and sonicated using a Covaris sonicator. Ago-2 was then immunoprecipitated using a monoclonal antibody (Miyoshi et al. 2005) bound to protein A Dynabeads (Invitrogen), washed very stringently, and eluted using SDS. Crosslinks were then reversed by incubating at 65°C overnight, and libraries were made from immunoprecipitated DNA using NEB ChIP-Seq Sample Prep Reagent Kit (NEB E6200S). Samples were barcoded and sequenced using 50 bp single-end sequencing on an Illumina HiSeq2000. As controls, libraries made from input and IgG samples were also made and sequenced. Reads were demultiplexed and aligned to the *Drosophila* genome using Bowtie 0.12.7 (Langmead et al. 2009). We obtained between 14 and 28 million reads for each sample, of which between 72 and 91% mapped uniquely. Enriched peaks of chromatin binding were determined using MACS (Zhang et al. 2008).

To compare Ago-2 binding to that of other factors, ChIP-CHIP and ChIP-Seq data was retrieved from the modEncode depository (Celniker et al. 2009) (<http://modencode.org/>). To determine the degree of overlap between binding regions of each protein, bound regions, as determined by the modEncode consortium, for a particular protein were compared both to bound regions for every other protein and to samples in which the bound regions for every protein had been randomized. The degree of enrichment of colocalization for two proteins was then calculated as the number of bp overlap between Protein A and Protein B divided by the number of bp overlap between randomized bound regions of Protein A and randomized bound regions of Protein B. To determine

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

enriched motifs under Ago-2 binding sites, we used the HOMER software package (Heinz et al. 2010). To determine the location of Ago-2 binding peaks in relation to annotated genes, the CEAS package was used (Shin et al. 2009).

Ago-2 CLIP-seq

CLIP-seq on S2 nuclear extract was performed as an adaption of the iCLIP procedure (Konig et al. 2010; Konig et al. 2011). S2 cells were irradiated three times with 400 mJ/cm² UV at 254 nm. Nuclei were then collected by swelling the cells in hypotonic buffer, douncing, and centrifuging. Briefly, cells were pelleted and resuspended in PBS. The cells were then pelleted and resuspended in hypotonic buffer (10 mM HEPES-KOH pH 7.6, 10 mM KCl, 1.5 mM MgCl₂, 0.5 mM DTT, 0.4 mM PMSF) and incubated on ice for 10 min. The cells were then pelleted, resuspended in hypotonic buffer, and dounced 15 times to lyse cells. The nuclei were then pelleted and resuspended in lysis buffer (50 mM Hepes-KOH, pH 7.6, 100 mM NaCl, 0.5 mM EDTA) and sonicated 6 times for 30 sec each using a tip sonicator, Triton X-100 was then added to 1 % and NP-40 was added to 0.1%.

Extracts were treated with DNase I and varying amounts of RNase I (Ambion AM2295) and then Ago-2 was immunoprecipitated using a homemade polyclonal antibody bound to protein A Dynabeads (Invitrogen). RNA from immunoprecipitated complexes was dephosphorylated on the beads using PNK (NEB), and an RNA linker was ligated to the 3' end of the immunoprecipitated RNA using T4 RNA ligase (NEB). The 5' end of the RNA was then phosphorylated with ³²P-ATP using PNK (NEB) and the complexes were eluted from the beads and run on an SDS-PAGE gel. The gel was exposed to X-ray film and slices of the gel corresponding to the known mobility of Ago-2 + 15 kDa were excised. Complexes were electroeluted from the gel at 200V for 3 hr (Hafner et al. 2010). The complexes were then proteinase K-treated, and RNA was recovered by ethanol precipitation. Reverse transcription was performed using oligonucleotides that contained both a multiplexing barcode and a 5 nt random barcode that would allow determination of PCR amplification events. After reverse transcription, the resulting cDNA was gel purified on a 6% TBE-urea gel, circularized using Circligase II (Epicentre). A complementary oligonucleotide annealed to the ligation junction and then the partially duplex region restriction was digested using BamHI to leave single stranded cDNA with known sequences on both ends of the insert. The cDNA was then amplified using Illumina primers PE 1.01 and PE 2.01 for 25 cycles using Phusion DNA polymerase (NEB). CLIP libraries were then sequenced using 50 bp single-end sequencing on an Illumina HiSeq2000.

After demultiplexing and collapsing all potential PCR duplicates, we obtained between 1.7 and 4.2 million reads for each sample. Of these, between 1.0 and 2.8 million aligned to the *Drosophila* genome using Bowtie (Langmead et al. 2009) allowing three mismatches (options -5 9 -v 3 -S), but only between 225K and 622K aligned uniquely.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

Many mapped to chrUextra, which contains scaffolds of repeats that are unable to be assembled into the genome.

Clusters were determined using pyCRAC software (Webb, Kudla, Tollervey and Granneman, in preparation). Clusters were required to contain at least 8 overlapping tags, and at least one nucleotide in each cluster was required to be mutated in at least 20 percent of tags in the cluster.

Comparisons to S2 cell mRNA-seq data were made by retrieving SAM files from the modEncode consortium (Celniker et al. 2009) and analyzing them using Cufflinks (Trapnell et al. 2010) and the following options: -u -m 200 -N -min-intron-length 40.

REFERENCES

- Allo M, Buggiano V, Fededa JP, Petrillo E, Schor I, de la Mata M, Agirre E, Plass M, Eyras E, Elela SA et al. 2009. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol* **16**: 717-724.
- Batsche E, Yaniv M, Muchardt C. 2006. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol* **13**: 22-29.
- Blanchette M, Green RE, Brenner SE, Rio DC. 2005. Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. *Genes Dev* **19**: 1306-1314.
- Blanchette M, Green RE, MacArthur S, Brooks AN, Brenner SE, Eisen MB, Rio DC. 2009. Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. *Mol Cell* **33**: 438-449.
- Brooks AN, Yang L, Duff MO, Hansen KD, Park JW, Dudoit S, Brenner SE, Graveley BR. 2011. Conservation of an RNA regulatory map between Drosophila and mammals. *Genome Res* **21**: 193-202.
- Buhler M, Moazed D. 2007. Transcription and RNAi in heterochromatic gene silencing. *Nat Struct Mol Biol* **14**: 1041-1048.
- Celniker SE, Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD, MacAlpine DM et al. 2009. Unlocking the secrets of the genome. *Nature* **459**: 927-930.
- Cernilogar FM, Onorati MC, Kothe GO, Burroughs AM, Parsi KM, Breiling A, Lo Sardo F, Saxena A, Miyoshi K, Siomi H et al. 2011. Chromatin-associated RNA interference components contribute to transcriptional regulation in Drosophila. *Nature* **480**: 391-395.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

- Chi SW, Zang JB, Mele A, Darnell RB. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**: 479-486.
- Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. 2009. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* **19**: 24-32.
- Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R et al. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453**: 798-802.
- Forstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD. 2007. *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* **130**: 287-297.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**: R80.
- Grimaud C, Bantignies F, Pal-Bhadra M, Ghana P, Bhadra U, Cavalli G. 2006. RNAi components are required for nuclear clustering of Polycomb group response elements. *Cell* **124**: 957-971.
- Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp AC, Munschauer M et al. 2010. PAR-CLIP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J Vis Exp*.
- Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA. 2010. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* **16**: 673-695.
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.
- Hua Y, Sahashi K, Hung G, Rigo F, Passini MA, Bennett CF, Krainer AR. 2010. Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes Dev* **24**: 1634-1644.
- Hua Y, Sahashi K, Rigo F, Hung G, Horev G, Bennett CF, Krainer AR. 2011. Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature* **478**: 123-126.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

- Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiue L, Hoon S, Brenner S et al. 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* **1**: 167-178.
- Janowski BA, Huffman KE, Schwartz JC, Ram R, Nordsell R, Shames DS, Minna JD, Corey DR. 2006. Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat Struct Mol Biol* **13**: 787-792.
- Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H. 2008. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**: 793-797.
- Kim K, Lee YS, Carthew RW. 2007. Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes. *RNA* **13**: 22-29.
- Kishore S, Khanna A, Zhang Z, Hui J, Balwierz PJ, Stefan M, Beach C, Nicholls RD, Zavolan M, Stamm S. 2010. The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Hum Mol Genet* **19**: 1153-1164.
- Kishore S, Stamm S. 2006. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* **311**: 230-232.
- Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**: 909-915.
- Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J.. 2011. iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J Vis Exp*.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843-854.
- Leung AK, Young AG, Bhutkar A, Zheng GX, Bosson AD, Nielsen CB, Sharp PA. 2011. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol* **18**: 237-244.
- Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**: 464-469.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

- Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science* **327**: 996-1000.
- Miyoshi K, Tsukumo H, Nagami T, Siomi H, Siomi MC. 2005. Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev* **19**: 2837-2848.
- Moshkovich N, Nisha P, Boyle PJ, Thompson BA, Dale RK, Lei EP. 2011. RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. *Genes Dev* **25**: 1686-1701.
- Pinol-Roma S, Swanson MS, Matunis MJ, Dreyfuss G. 1990. Purification and characterization of proteins of heterogeneous nuclear ribonucleoprotein complexes by affinity chromatography. *Methods Enzymol* **181**: 326-331.
- Pushpavalli SN, Bag I, Pal-Bhadra M, Bhadra U. 2012. *Drosophila* Argonaute-1 is critical for transcriptional cosuppression and heterochromatin formation. *Chromosome Res* **20**: 333-351.
- Shin H, Liu T, Manrai AK, Liu XS. 2009. CEAS: cis-regulatory element annotation system. *Bioinformatics* **25**: 2605-2606.
- Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: Article3.
- Spies N, Nielsen CB, Padgett RA, Burge CB. 2009. Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* **36**: 245-254.
- Taft RJ, Glazov EA, Lassmann T, Hayashizaki Y, Carninci P, Mattick JS. 2009. Small RNAs derived from snoRNAs. *RNA* **15**: 1233-1240.
- Taliaferro JM, Alvarez N, Green RE, Blanchette M, Rio DC. 2011. Evolution of a tissue-specific splicing network. *Genes Dev* **25**: 608-620.
- Tomari Y, Du T, Zamore PD. 2007. Sorting of *Drosophila* small silencing RNAs. *Cell* **130**: 299-308.
- Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105-1111.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511-515.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

Venables JP, Klinck R, Bramard A, Inkel L, Dufresne-Martin G, Koh C, Gervais-Bird J, Lapointe E, Froehlich U, Durand M et al. 2008. Identification of alternative splicing markers for breast cancer. *Cancer Res* **68**: 9525-9531.

Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**: 672-676.

Weinmann L, Hock J, Ivancevic T, Ohrt T, Mutze J, Schwille P, Kremmer E, Benes V, Urlaub H, Meister G. 2009. Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell* **136**: 496-507.

Wightman B, Ha I, Ruvkun G. 1993. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**: 855-862.

Xu K, Bogert BA, Li W, Su K, Lee A, Gao FB. 2004. The fragile X-related gene affects the crawling behavior of *Drosophila* larvae by regulating the mRNA level of the DEG/ENaC protein pickpocket1. *Curr Biol* **14**: 1025-1034.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

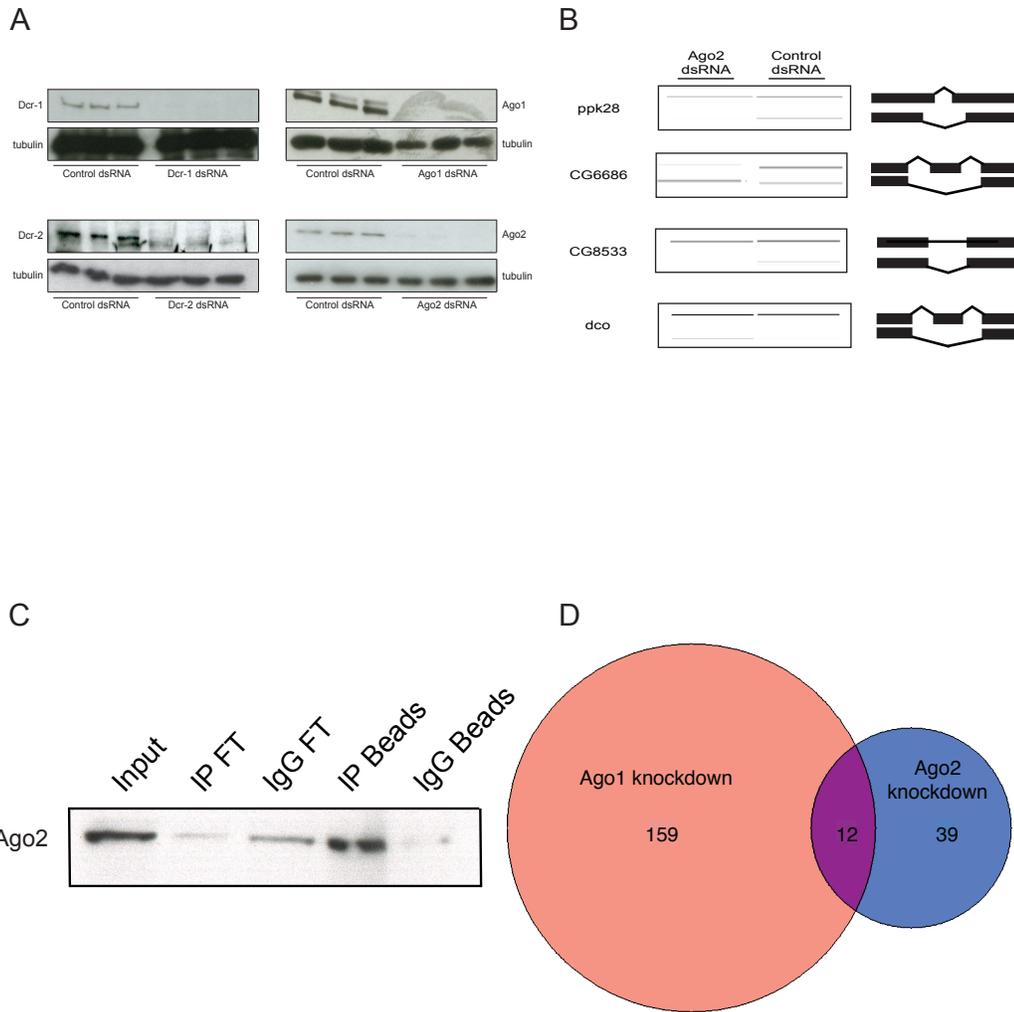
Zhou R, Hotta I, Denli AM, Hong P, Perrimon N, Hannon GJ. 2008. Comparative analysis of argonaute-dependent small RNA pathways in *Drosophila*. *Mol Cell* **32**: 592-599.

Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE, Yeo GW. 2010. Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nat Struct Mol Biol* **17**: 173-179.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

Figure 1. RNAi knockdowns in *Drosophila* S2 cells. A) Immunoblots of Dcr-1, Dcr-2, Ago-1, and Ago-2 in S2 cells after treatment with control or experimental double-stranded RNA for four days. B) RT-PCR validation of four Ago-2 splicing targets predicted by the splice junction microarray in S2 cells. Splice isoform levels are shown after treatment with control double-stranded RNA and double-stranded RNA directed against Ago-2. C) Immunoblot of Ago-2 after Ago-2 immunoprecipitation from S2 cell nuclear extract. D) Number of genes whose transcript levels changed after knockdown of Ago-1 and Ago-2 in S2 cells.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression



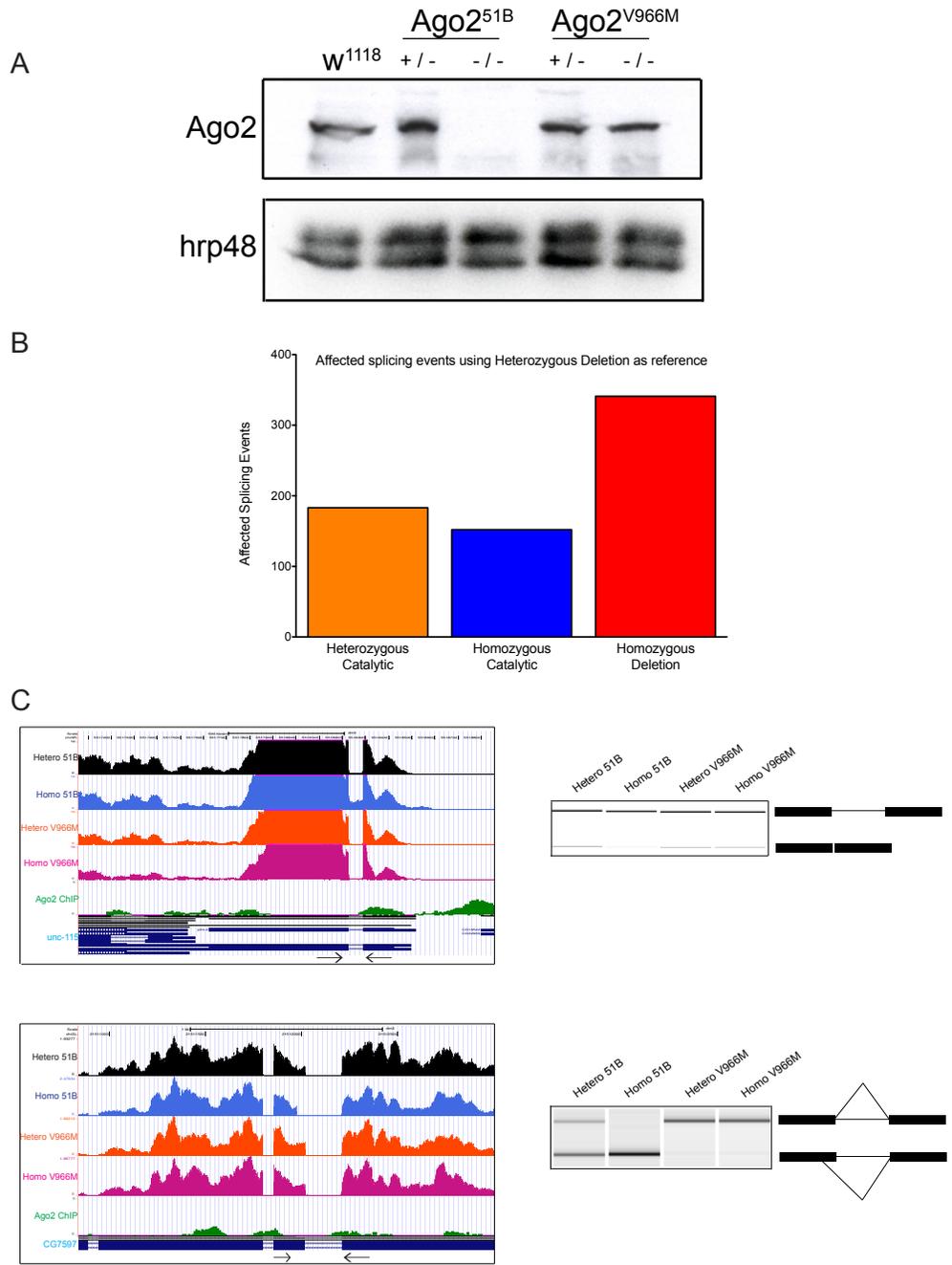
Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

Figure 2. Analysis of Ago-2 in *Drosophila* cell culture. A) Splice junction microarray summary. The number of upregulated and downregulated junction probes in response to knockdown of each protein is indicated. B) Subcellular localization of FLAG-tagged Ago-2. PSI is used as a nucleoplasmic marker, and tubulin is used as a cytoplasmic marker. Expression of the tagged Ago-2 is under the control of a copper-sensitive promoter. C) RT-PCR of RNA recovered after immunoprecipitation of Ago-2 from S2 nuclear extract. D) Clustered heatmap of Pearson correlation coefficients of global RNA expression levels after knockdown of many RNA-binding proteins.

Figure 3. Analysis of splicing events affected in Ago-2 mutant *Drosophila* adults.

A) Ago-2 immunoblot of whole fly extracts from Ago-2 mutant strains. Hrp48 is used as a loading control. B) Number of unique splicing changes observed in each Ago-2 mutant strain. The heterozygous deletion strain is used as a reference. Splice junctions with PSI values differing by at least 5 from the heterozygous deletion PSI value are reported here. C) Two novel alternative splicing events detected in the homozygous Ago-2^{51B} strain and validated by RT-PCR using RNA isolated from 0-16 hour post-eclosion adult males. The positions of the primers used to perform the RT-PCR are indicated as black arrows.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression



Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

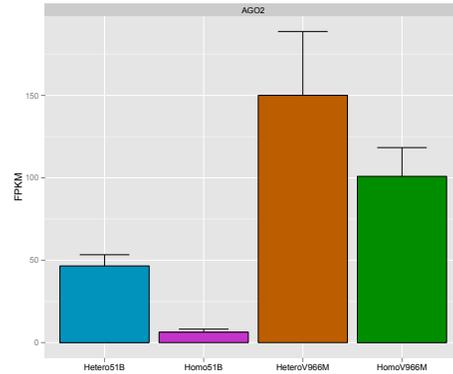
Figure 4. mRNA-seq analysis of Ago-2 mutant strains. A) Number of total and mapped mRNA-seq reads for each Ago-2 mutant strain. B) Ago-2 FPKM levels in each Ago-2 mutant strain. C) FPKM values for 34 different RNA-binding proteins in each Ago-2 mutant strain.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

A

Sample	Reads	Mapped
Hetero 51B	59.9 M	59.5 M (99.3%)
Homo 51B	96.8 M	94.5 M (97.6%)
Hetero V966M	58.4 M	58.1 M (99.5%)
Homo V966M	62.2 M	60.5 M (95.7%)

B



C

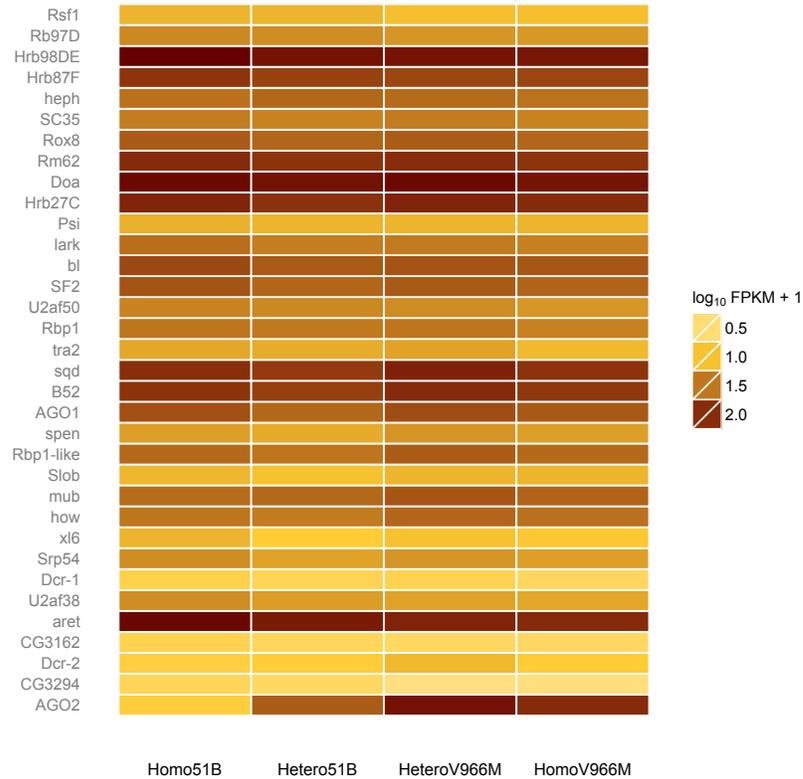
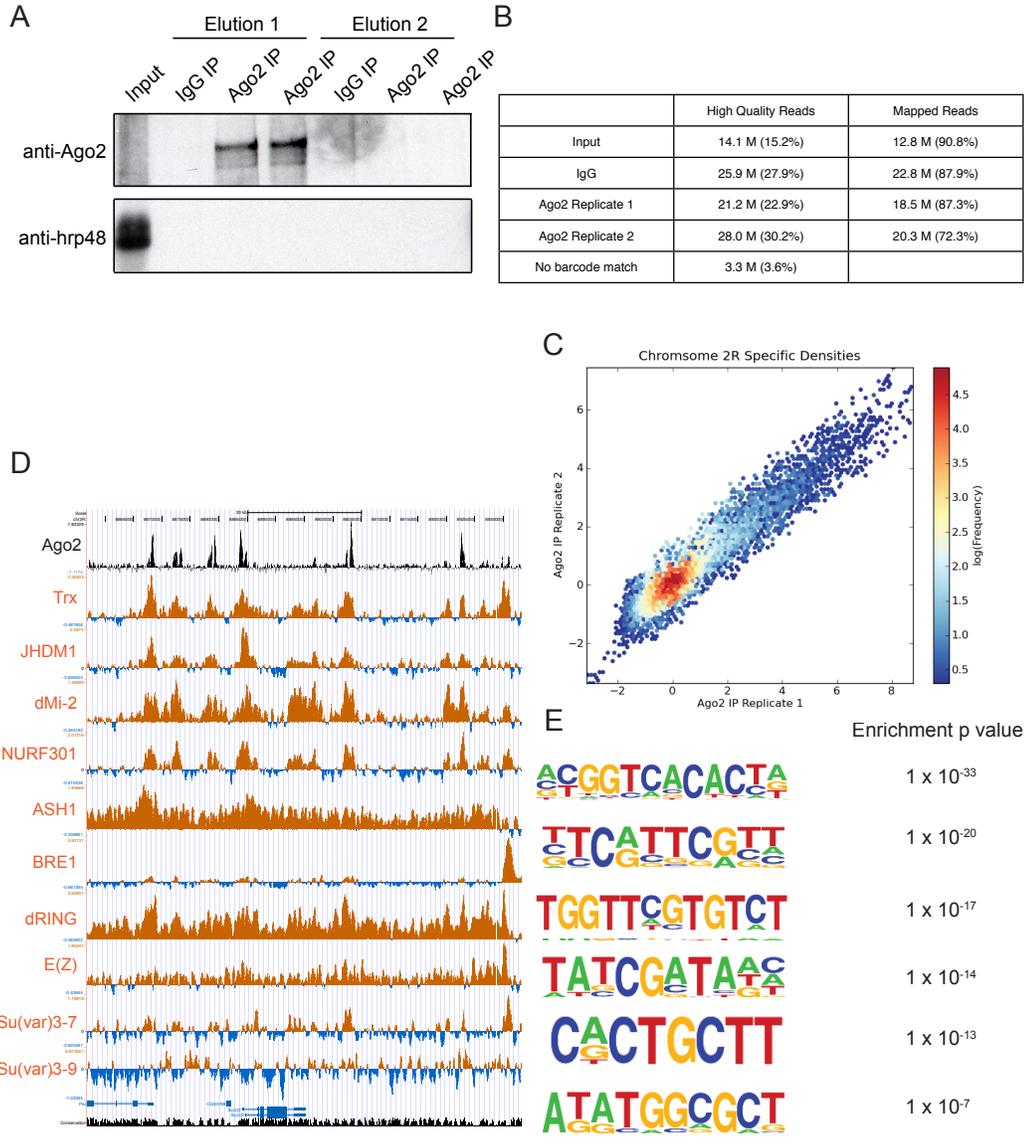


Figure 5. Chromatin immunoprecipitation-sequencing (ChIP-seq) analysis of *Drosophila* Ago-2. A) Immunoblot of Ago-2 after immunoprecipitation from formaldehyde-crosslinked S2 cell chromatin extract. B) Number of total and mapped ChIP-seq reads for each Ago-2 IP sample. C) Number of reads per million mapped into each 10 bp bin across chromosome 2R for both replicates of the Ago-2 immunoprecipitation, with the corresponding density of each bin from the input sample subtracted away. D) Representative browser graphic showing chromatin binding profiles of Ago-2 (ChIP-seq, black) and several other chromatin-binding proteins (ChIP-CHIP, orange) (Celniker et al. 2009). E) Enriched motifs found under Ago-2-bound chromatin sequences as detected by HOMER (Heinz et al. 2010) The motif similar to that of CTCF is listed at the bottom (Cuddapah et al. 2009).

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression



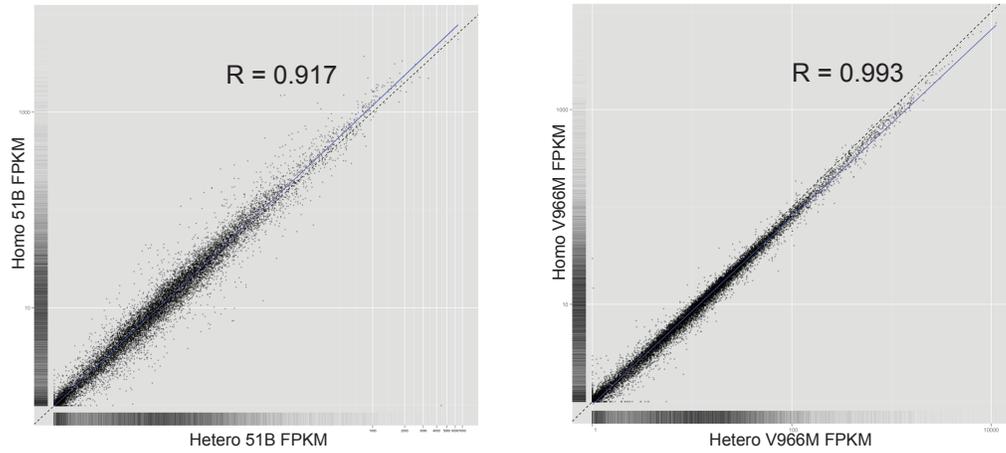
Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

Figure 6. Chromatin association of Ago-2 in Drosophila cell culture. A) Representative browser graphic of Ago-2 ChIP performed on larvae (orange) and embryonically-derived S2 cells (green). B) Profile of Ago-2 ChIP tag density of all genes (black), genes whose mRNA expression level was sensitive to Ago-2 in S2 cells (red), and Ago-2 splicing targets in S2 cells (purple), represented along a metagene. C) Clustered heatmap showing enrichment of colocalization on chromatin by ChIP-seq and ChIP-CHIP between 21 different chromatin binding factors (Celniker et al. 2009).

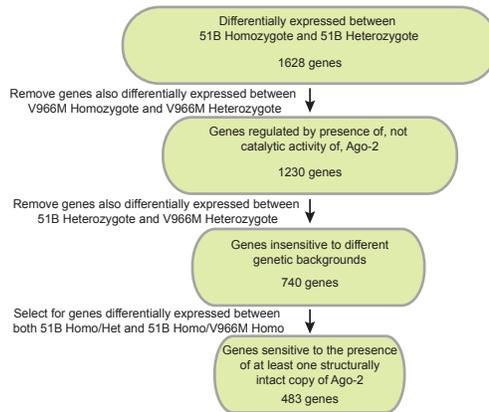
Figure 7. mRNA-seq analysis of Ago-2 mutant strains. A) Scatter plots of FPKM values of all genes, comparing heterozygotes and homozygotes of Ago-2 deletion mutants (51B, left) and catalytic mutants (V966M, right). B) Strategy to identify genes that are sensitive to the presence of Ago-2 but not to its catalytic activity. C) mRNA levels of genes that were both sensitive to the presence, but not catalytic activity, of Ago-2 in whole flies and bound by Ago-2 at the chromatin level in S2 cells. Expression levels were normalized against those seen in the heterozygous Ago-2^{51B} sample.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

A



B



C

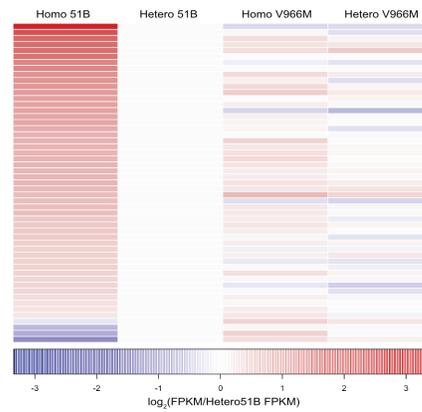


Figure 8. Effects of Ago-2 mutations on gene expression. A) Expression levels of heat shock genes in the four mutant strains. Expression levels were normalized against those seen in the heterozygous Ago-2^{51B} sample. B) Venn diagram showing genes whose FPKM values significantly changed when comparing the Ago-2^{51B} homozygous and heterozygous mutant samples or the Ago-2^{V966M} homozygous and heterozygous mutant samples. C) Same as B, with the addition of genes whose FPKM values changed significantly when comparing the Ago-2^{51B} heterozygous and Ago-2^{V966M} heterozygous mutant samples. D) Same as C, with the addition of genes whose FPKM values changed significantly when comparing the Ago-2^{51B} homozygous and Ago-2^{V966M} homozygous mutant samples. In this analysis, the large overlap between the Ago-2^{51B} homo/het and Ago-2^{51B} homo/Ago-2^{V966M} homo comparison indicates genes that are sensitive to the presence of but not the catalytic activity of Ago-2. These two comparisons should yield similar results as both the Ago-2^{51B} het and Ago-2^{V966M} homo have at least one structurally intact copy of the Ago-2 gene. E) Dendrogram based on Jensen-Shannon distances between the four samples when comparing FPKM values for all genes. Overall, homozygous and heterozygous samples from one strain were more similar to each other than to either sample from the other strain. This indicates that genetic background is having a noticeable effect. F) Clustered heatmap of FPKM values for the 483 genes that were specifically sensitive to the presence of Ago-2 and not its catalytic activity (see Figure 7B). G) Dendrogram based on Jensen-Shannon distances between the four samples when comparing FPKM values for those genes that were both sensitive to the presence of Ago-2 and bound by Ago-2 at the chromatin level. H) Promoter switching in 14-3-3zeta. Splice junction microarray data is visualized in orange (increase in splicing) and blue (decrease in splicing) boxes. Here, usage of the proximal and distal promoters is measured by probes spanning splice junctions exclusive to each promoter. Ago-2 ChIP-seq data is shown in green. Upon knockdown of Ago-2, usage of the proximal promoters increases (tall orange boxes connecting proximal promoter with second exon).

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

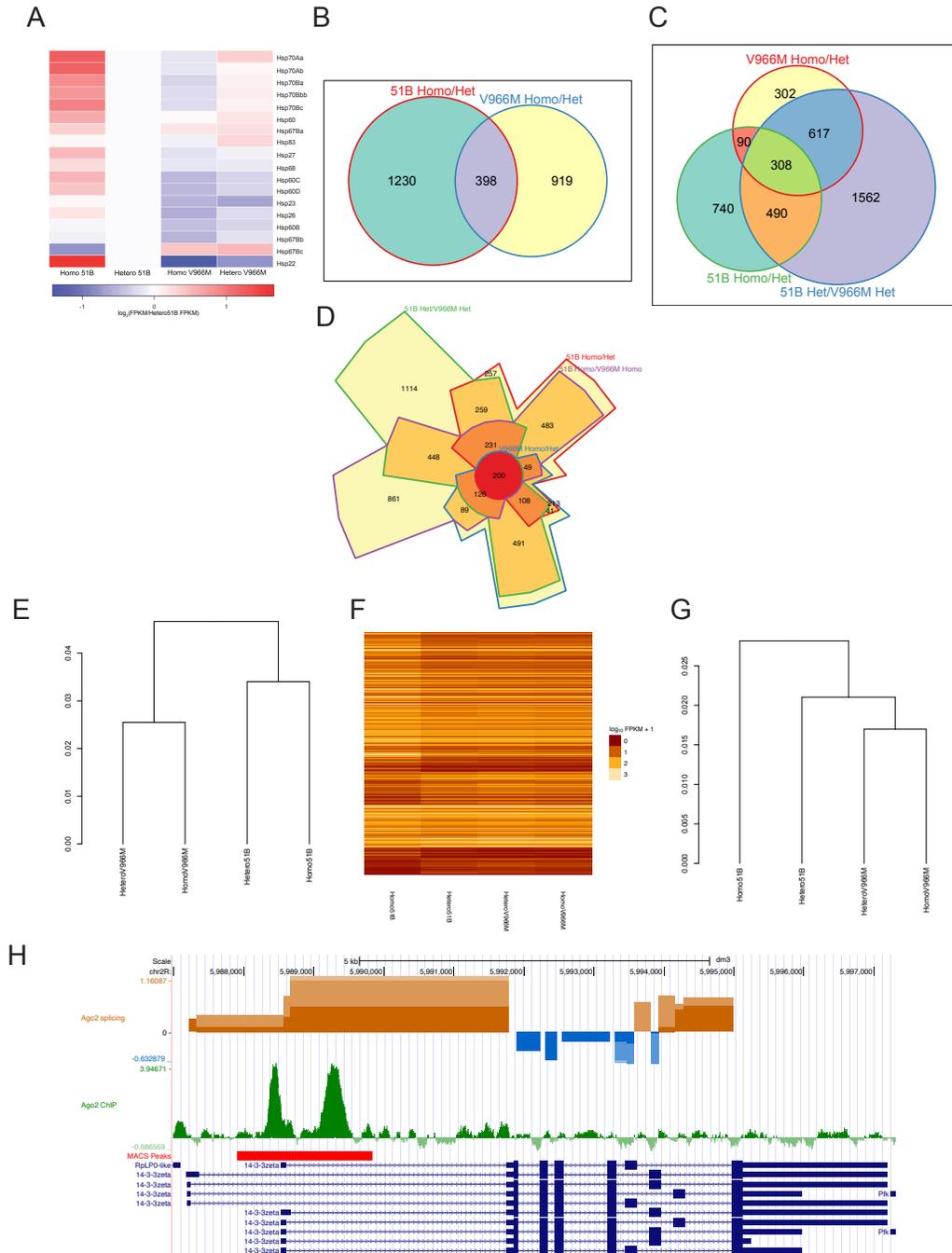


Figure 9. CLIP-seq of Ago-2 in *Drosophila* cell culture. A) immunoblot analysis of Ago-2 CLIP SDS-PAGE gel. RNase dilutions on top indicate the amount of dilution from an RNase I stock. The IgG sample received the 1:50 RNase dilution. Ago-2 migrates at 130 kDa. Immunoprecipitation and blot probing were performed with homemade polyclonal Ago-2 antibody. B) Autoradiogram of Ago-2 CLIP SDS-PAGE gel. RNase dilutions are as in A. The –UV and IgG samples received the 1:50 RNase dilution. Sections cut from the gel were from the low amount of RNase (1:500 dilution) and from the size of Ago-2 upward 15 kDa (130 to 145 kDa). C) Browser graphic showing Ago-2 CLIP tags (blue) and modEncode-determined S2 cell mRNA-seq levels (black) (Celniker et al. 2009). Tags shown are clustering strongly on tRNA:D:96A. D) Browser graphic as in C, except that now Ago-2 CLIP tags are in red as they map to the minus strand. Tags shown are clustering strongly on snoRNA-MeU2-C28. E) Expression levels (from modEncode mRNA-seq data) of all genes expressed in S2 cells (FPKM > 0) and those containing Ago-2 CLIP tag clusters. Boxes represent the 25th-75th percentiles, while whiskers represent 5th to 95th percentiles. F) The motif obtained upon combining the top four enriched motifs in Ago-2 CLIP clusters and weighting them based on their prevalence within the clusters. G) Enrichment of the motif in Supplementary Figure 5F around Ago-2-sensitive splice sites. Log-odds scores of finding the motif at each position were calculated for the Ago-2-sensitive and Ago-2-insensitive splice junctions. The values shown are the ratios (sensitive/insensitive) of the median scores at each position.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression

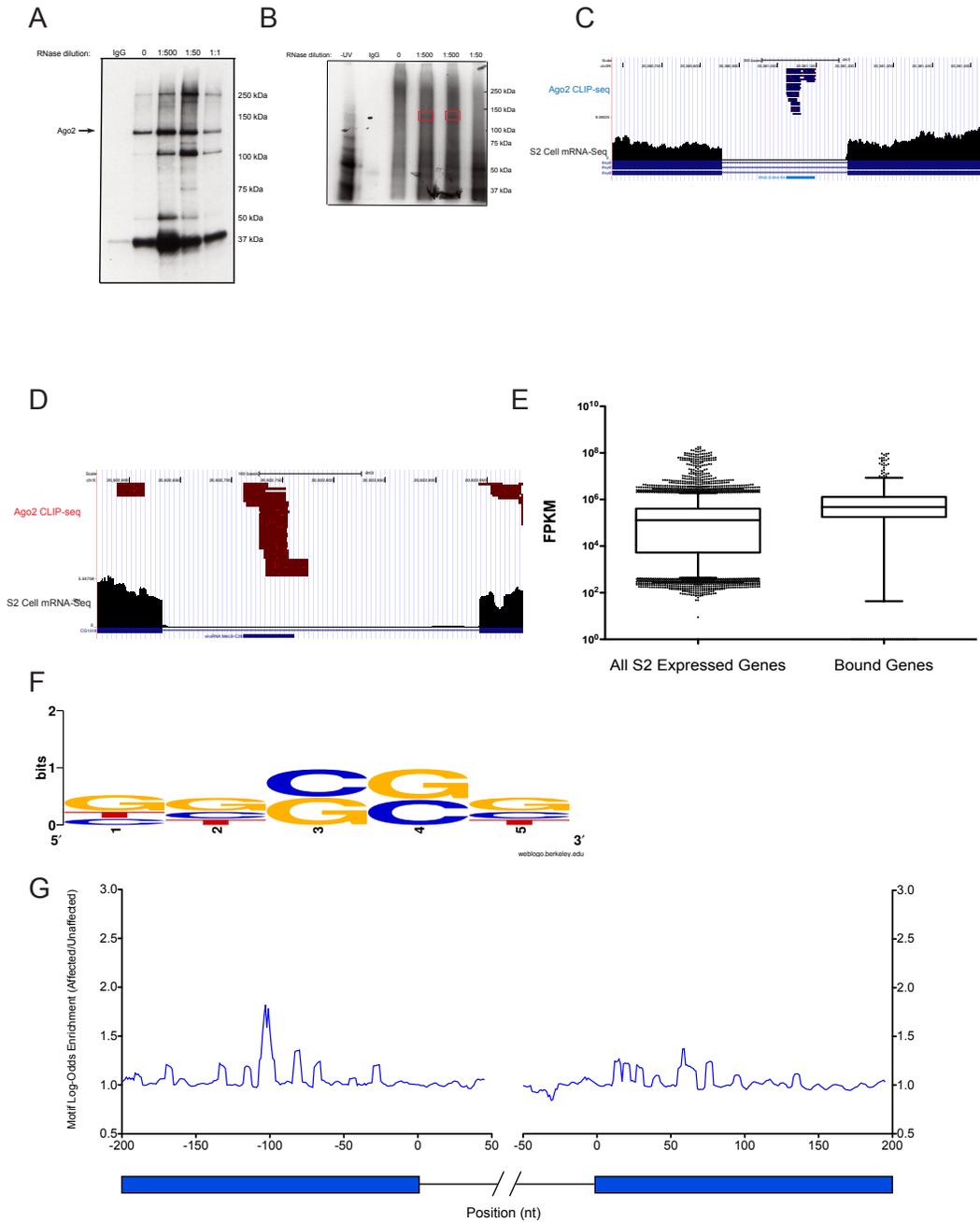
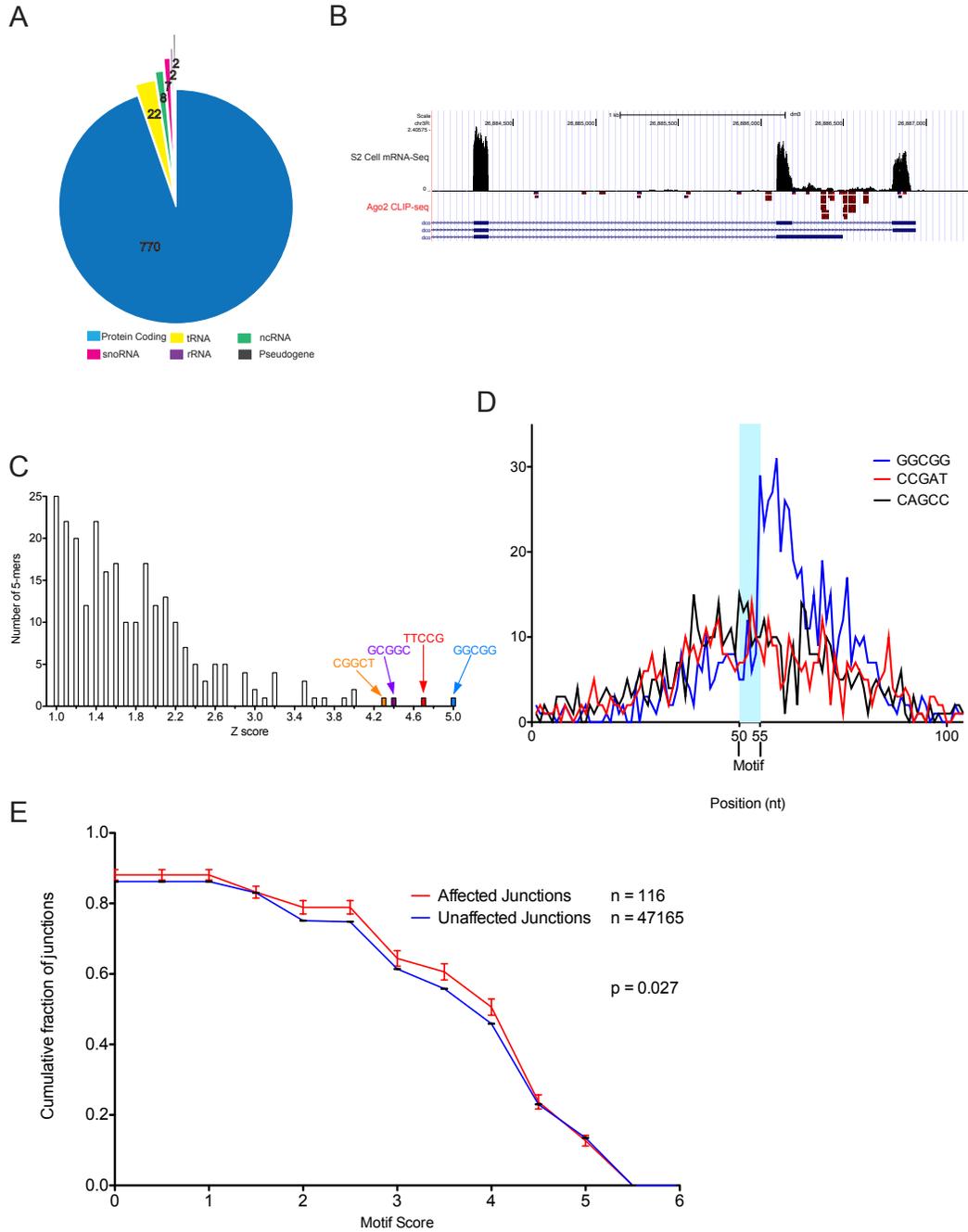


Figure 10. CLIP-seq analysis of *Drosophila* Ago-2. A) Pie chart indicating where significantly enriched clusters of Ago-2 binding were found using CLIP-seq. B) Browser graphic showing S2 cell mRNA expression in black (as determined by the ModEncode consortium) of the 5' end of the *dco* gene. The cassette exon shown was shifted towards exclusion after knockdown of Ago-2. Ago-2 CLIP tags are shown in red. C) Enriched 5-mers contained in clusters of Ago-2 CLIP tags. The highest enriched motifs are shown. D) Mutation density around enriched motifs in Ago-2 CLIP tag clusters. The frequency of mutations in 50 nucleotides upstream and downstream of the motif were calculated for the most enriched (blue) 25th most enriched (red) and 100th most enriched (black) motifs. E) Enrichment of the G-rich motif obtained from the CLIP-seq experiment around Ago-2-sensitive splice junctions. Sequences were restricted to 200 nt into the exon and 50 nt into the intron surrounding the 5' and 3' splice sites of Ago-2-sensitive (red) or insensitive (blue) junctions.

Chapter 3: New roles for Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression



Chapter 4: Evolution of a tissue-specific splicing network

Portions of this chapter were taken from: Taliaferro JM, Alvarez N, Green RE, Blanchette M, Rio DC. Evolution of a tissue-specific splicing network. *Genes and Development* **25**(6): 608-620.

SUMMARY

Alternative splicing of pre-mRNA is a strategy employed by most eukaryotes to increase transcript and proteomic diversity. Many metazoan splicing factors are members of multi-gene families, with each member having different functions. How these highly related proteins evolve unique properties has been unclear. Here we characterize the evolution and function of a new *Drosophila* splicing factor, termed LS2, that arose from a gene duplication event of dU2AF⁵⁰, the large subunit of the highly conserved heterodimeric general splicing factor U2AF. The quickly evolving LS2 gene has diverged from the splicing-promoting, ubiquitously expressed dU2AF⁵⁰ such that it binds a markedly different RNA sequence, acts as a splicing repressor, and is preferentially expressed in testes. Target transcripts of LS2 are also enriched for performing testes-related functions. We therefore propose a path for the evolution of a new splicing factor in *Drosophila* that regulates specific pre-mRNAs and contributes to transcript diversity in a tissue-specific manner.

INTRODUCTION

Alternative splicing is the complex process by which many different eukaryotic mRNAs are generated from nuclear precursor mRNAs (pre-mRNAs). The splicing of one transcript in several different ways allows the generation of vast proteomic diversity from a comparatively smaller number of genes (Nilsen and Graveley 2010). These alternatively spliced transcripts are often restricted to particular tissues and encode proteins that are critical to proper tissue function (Wang and Burge 2008). Regulation of pre-mRNA splicing is achieved through the interaction of RNA sequence elements and a variety of related RNA binding protein factors (Black 2003; Ben-Dov et al. 2008; Wang and Burge 2008).

Many different alternative splicing patterns exist (Black 2003). All of these involve the employment of one splice site over another. The efficiency with which splice sites are recognized and their ability to recruit functionally competent spliceosome components regulate splice site utilization (Nelson and Green 1988; Yu et al. 2008). These efficiencies can be modulated by the binding of factors that enhance or repress splice site use (Blanchette et al. 2005). The recognition and determination of 3' splice sites is primarily carried out by U2 associated factor (U2AF) (Ruskin et al. 1988; Zamore and Green 1989).

Chapter 4: Evolution of a tissue-specific splicing network

The essential, highly conserved U2AF general splicing factor is a heterodimer composed of large (U2AF^{LS}) and small (U2AF^{SS}) subunits that promotes spliceosome assembly (Ruskin et al. 1988; Singh et al. 1995). U2AF is conserved among all eukaryotic species from *S. pombe* to humans. U2AF^{LS} (dU2AF⁵⁰ in *Drosophila*) recognizes the polypyrimidine tract at the 3' end of the intron (Zamore and Green 1989; Kanaar et al. 1993) while its cooperating partner U2AF^{SS} (dU2AF³⁸ in *Drosophila*) interacts with the intron-terminal AG dinucleotide (Merendino et al. 1999; Wu et al. 1999; Zorio and Blumenthal 1999). U2AF^{LS} additionally cooperates with the branchpoint adenosine-binding SF1 through interactions in its C-terminal pseudo-RNA recognition motif (Kent et al. 2003; Selenko et al. 2003). Following these contacts, the 3' end of the intron is then competent for interaction with U2 snRNP. U2AF therefore functions to promote spliceosome assembly.

Much work has been done concerning the evolutionary conservation of the *cis*-acting RNA sequence elements. Many sequence elements are widely conserved even across vast evolutionary distances and often lead to similar splicing patterns in the orthologous transcripts (Brooks et al. 2010). However, little is understood about how related family members of the RNA binding proteins that mediate these splicing effects arise and diverge to acquire distinct and diverse functions (Baek and Green 2005; Akerman et al. 2009). These distinct functions allow evolutionarily related proteins to form regulatory networks with each member controlling the splicing of specific transcripts through the recognition of specific sequence motifs. Here, we have identified and characterized the appearance and evolutionary divergence of a *Drosophila* splicing factor that we have termed LS2 (Large Subunit 2, also known as CG3162). LS2 arose from a retro-duplicated copy of the highly conserved, positively acting dU2AF⁵⁰, and has diverged sufficiently from dU2AF⁵⁰ such that it is highly specialized in its specificity, function and expression.

RESULTS

LS2 evolved from a retro-duplicated copy of dU2AF⁵⁰

LS2 and dU2AF⁵⁰ are 55% identical and 70% similar at the primary sequence level (Figure 1). Using the amino acid sequences of several U2AF large subunit and LS2 genes, we determined that the LS2 gene arose via a duplication event before the most recent common ancestor of all *Drosophila* (Figure 2A). The LS2 orthologs are in syntenic positions in each *Drosophila* genome. We could not detect an LS2 ortholog in mosquito or honeybee. Given the estimated ages of the most recent common ancestor of *Drosophila* and mosquito and the most recent common ancestor of the 12 *Drosophila* species analyzed (Tamura et al. 2004), we conclude that the duplication event that gave

Chapter 4: Evolution of a tissue-specific splicing network

rise to LS2 occurred between 60 and 250 million years ago.

Sequence analysis of the dU2AF⁵⁰ orthologs revealed little divergence between the orthologs, consistent with the conserved function of the U2AF large subunit and its requirement for viability (Kanaar et al. 1993). However, the LS2 orthologs were comparatively highly diverged. Thus, while the dU2AF⁵⁰ orthologs are under much constraint and negative selection to retain their current function, the LS2 orthologs may be free to acquire new functions and may be under positive selection. While the dU2AF⁵⁰ in *D. melanogaster* contains five introns, LS2 does not contain any introns. This implies the use of an RNA intermediate during the gene duplication process, consistent with the idea of a retro-duplication event.

LS2 controls splicing of a transcript pool that is distinct from that of dU2AF⁵⁰

To determine whether LS2 was simply a redundant form of dU2AF⁵⁰, we used *Drosophila* splice junction microarrays to determine the splicing events sensitive to dU2AF⁵⁰, dU2AF³⁸, and LS2 after RNA interference (RNAi) knockdown in *Drosophila* S2 cells (Blanchette et al. 2005). We verified that LS2 expression was efficiently knocked down (Figure 3A). Analysis of the microarray results revealed that dU2AF⁵⁰, dU2AF³⁸, and LS2 affected the splicing of 378, 497 and 311 splice junctions in 206, 276 and 168 genes, respectively (Figure 3B). dU2AF⁵⁰ is a core splicing factor and as such may not be expected to specifically regulate distinct transcripts. Nevertheless, our data is consistent with previous studies in which core spliceosomal factors did have such specificity (Park et al. 2004; Sridharan et al. 2010). Although the collection of splice junctions sensitive to dU2AF⁵⁰ and LS2 depletion overlapped to a small extent, the majority of them were unique to either protein (Figure 2B). To more precisely characterize the relationship between the splice junction targets of each protein, we incorporated the magnitude and direction of splice junction changes upon knockdown. These plots showed little correlation between the responses to dU2AF⁵⁰ and LS2 knockdown (Figure 2C). Thus, these two proteins have distinct splice junction specificities and functions. By contrast, there was a strong correlation of the responses to dU2AF⁵⁰ and dU2AF³⁸ knockdown (Figure 3C), consistent with their known physical and functional interactions (Rudner et al. 1998b). We also observed an intermediate correlation of the responses to LS2 and dU2AF³⁸ knockdown, implying a possible functional interaction. Finally, we validated several of the predicted splicing changes predicted by the microarray using semi-quantitative RT-PCR (Figure 3D).

LS2 has diverged from dU2AF⁵⁰ in RNA sequence recognition specificity

To directly determine whether dU2AF⁵⁰ and LS2 recognize similar or different RNA

Chapter 4: Evolution of a tissue-specific splicing network

binding sites, we used in vitro binding site selection (SELEX) to determine an optimized RNA binding sequence motif for LS2. Similar analyses with the large subunit of U2AF showed that U2AF^{LS} preferentially recognizes pyrimidine-rich sequences, consistent with its role in spliceosome assembly through recognition of the polypyrimidine tract (Singh et al. 2000; Sickmier et al. 2006). By contrast, although the LS2 and dU2AF⁵⁰ proteins are highly related in primary sequence throughout their RRM domains (Figure 1), the purified LS2 protein preferentially binds to a G-rich RNA motif with much less degeneracy at specific positions (Figure 4A). This RNA binding specificity was confirmed using quantitative electrophoretic mobility shift RNA binding assays using an RNA probe containing the SELEX-derived motif (Figure 4B) and a mutant probe that much more closely resembled a polypyrimidine-rich RNA (Figure 4C). Similar to the measured equilibrium dissociation constant (K_d) of 2.2 μ M for purified dU2AF⁵⁰ binding to a polypyrimidine RNA, the apparent K_d of the LS2 protein for its RNA SELEX motif was 1.9 μ M (Figure 4D). The LS2 protein showed a much lower affinity for the mutant probe. Additionally, as was also the case for dU2AF⁵⁰, the highly positively charged, N-terminal RS domain was required for high-affinity RNA binding but did not play a role in sequence specificity (Figure 4D) (Rudner et al. 1998a). Finally, purified recombinant LS2 / dU2AF³⁸ heterodimer bound RNA much more tightly than LS2 monomer alone (Figure 5). The apparent equilibrium dissociation constant (K_d) of the heterodimer for a G-rich RNA was 150nM, similar to the affinity of U2AF heterodimer for a polypyrimidine RNA (Rudner et al. 1998a). The heterodimer also showed greater non-specificity in RNA binding, that may be due to the presence of an additional RS domain provided by dU2AF³⁸. However, the LS2-dU2AF³⁸ heterodimer still bound preferentially to a G-rich RNA. The increased non-specificity for G-rich versus pyrimidine-rich DNA of the LS2-dU2AF³⁸ heterodimer compared to the LS2 monomer is also consistent with the previously documented RNA binding properties of human and *Drosophila* U2AF (Rudner et al. 1998a).

If the derived SELEX motif for LS2 binding is correct and the target transcript pool from the LS2 RNAi-knockdown splice junction microarray data are direct targets of the LS2 protein, we reasoned that the G-rich LS2-binding motif should be enriched in the LS2-affected genes over all other unaffected transcripts. Similar patterns of RNA binding motif enrichment have been observed previously with known splicing factors with well-defined RNA binding motifs and from in vivo transcript-binding data (Blanchette et al. 2009). We detected such an enrichment (p value $< 1 \times 10^{-5}$) of preferred LS2 RNA binding motifs in the 168 LS2-affected genes (Figure 4E).

LS2 interacts with dU2AF³⁸ in an RNA-independent manner

U2AF^{LS} functions in spliceosome assembly in conjunction with the small U2AF subunit, U2AF^{SS}. U2AF^{SS} functions as a core splicing factor whose role is to recognize the

Chapter 4: Evolution of a tissue-specific splicing network

intron-terminal AG dinucleotide (Merendino et al. 1999; Wu et al. 1999; Zorio and Blumenthal 1999). U2AF^{LS} and U2AF^{SS} in humans and *Drosophila* interact through a hydrophobic interface (Zamore and Green 1989; Rudner et al. 1998b; Kielkopf et al. 2001) that in both LS2 and dU2AF⁵⁰ is located in between the RS domain and the first RRM. Both dU2AF⁵⁰ and LS2 contain the critical hydrophobic residues necessary for this interaction (Figure 1). To test whether LS2 can physically interact with dU2AF³⁸, we performed GST-pulldown interaction assays with recombinant LS2 and dU2AF³⁸ proteins. GST-tagged dU2AF⁵⁰ and LS2 bound recombinant dU2AF³⁸ (Figure 6A lanes 3 and 4) and this interaction was dependent on the presence of the putative U2AF^{SS} interaction domain (Figure 6A lanes 5 and 6). Additionally, recombinant LS2 and dU2AF³⁸ co-eluted from an ion exchange column at 900 mM KCl, consistent with a hydrophobic interaction between the two proteins (data not shown).

To test whether LS2 and dU2AF³⁸ interact in *Drosophila* cells, we used a stably transfected S2 cell line that expressed an epitope-tagged LS2. Endogenous dU2AF³⁸ could be co-immunoprecipitated with polyoma (also known as Py or Glu-Glu)-tagged LS2 from these S2 cell nuclear extracts (Figure 6B lanes 1 and 2). This interaction was resistant to RNase treatment (Figure 6B lanes 3 and 4), indicating that these two proteins were physically interacting, not simply bound to the same RNA. However, dU2AF³⁸ could not be immunoprecipitated using a polyoma antibody from S2 cell extract containing a FLAG-tagged version of LS2, indicating the specificity of the interaction (Figure 6B, lanes 5 and 6). Additionally, there is likely to be a functional interaction between LS2 and dU2AF³⁸ *in vivo* based on the moderate correlation and overlap of splice junction population changes in response to LS2 and dU2AF³⁸ knockdown (Figure 2B, Figure 3C). We therefore propose that LS2 has co-opted a fraction of the cellular dU2AF³⁸ population for use on its distinct transcript pools in extra-spliceosomal functions.

Expression of LS2 is highly enriched in testes

Many alternative splicing events are specific to a particular cell or tissue type. A common mechanism for achieving this specificity is to restrict expression of the necessary splicing factors to the appropriate tissues, as is the case for the mammalian nervous system-specific factors nPTB (Kikuchi et al. 2000; Markovtsov et al. 2000) and Nova (Buckanovich et al. 1993). Although dU2AF⁵⁰ expression is ubiquitous, consistent with its function as a general splicing factor, FlyAtlas expression microarray data indicated that LS2 mRNA was preferentially expressed in the testes (Chintapalli et al. 2007). To confirm that this is also true for LS2 protein, we performed immunoblot analysis for LS2 using whole males, whole females, heads, and testes. While expression of LS2 in whole flies and in heads compared to the loading control was negligible, we detected significant expression in testes, consistent with the mRNA

Chapter 4: Evolution of a tissue-specific splicing network

expression array results (Figure 7A). Moreover, gene ontology analysis on the LS2-affected transcripts revealed several gene ontology terms consistent with a role in testes function, gamete production, and cellular regulation through phosphorylation (Al-Shahrour et al. 2006) (Figure 7B, Figure 8A). Fewer GO term enrichments were seen for genes affected by dU2AF⁵⁰ and dU2AF³⁸ (Figure 8B,C), consistent with their ubiquitous expression and function as general, spliceosome-associated splicing factors.

If expression of LS2 was highly enriched in testes, we hypothesized that expression of the LS2 target transcripts should also be testes-enriched. Using FlyAtlas tissue expression data, we found that 87.4% of all genes expressed in S2 cells are also expressed in testes (Chintapalli et al. 2007). However, 97.5% of LS2 targets identified from S2 cells are expressed in testes, representing a significant enrichment ($p < 0.0001$, chi-squared test). Furthermore, when the magnitude of expression is taken into account, the LS2 mRNA targets tend to be much more highly expressed in testes than either all *Drosophila* genes or those present in S2 cells (Figure 7C).

LS2 acts as a splicing repressor *in vitro* and *in vivo*

We then asked where positional enrichments of LS2 recognition motifs were located in the endogenous target transcripts of LS2 that were identified by the RNAi-splice junction microarrays. Analysis of the location of the LS2 recognition motifs near affected cassette exon junctions showed an enrichment of motifs associated with exon skipping just upstream of the cassette exon (Figure 9A). This peak was approximately 60 nt upstream of the 3' splice site.

In order to investigate the molecular mechanism by which LS2 affects splicing of specific transcripts, we modified the efficiently spliced *Drosophila ftz* intron by adding a G-rich LS2 binding site motif 65 nucleotides upstream of the 3' splice site (Figure 9B). This placed the LS2 binding site upstream of both the polypyrimidine tract and the branchpoint adenosine. We then monitored splicing of this modified pre-mRNA in HeLa cell nuclear splicing extract in the presence or absence of purified recombinant LS2/dU2AF³⁸ heterodimer protein or LS2 protein alone. In these *in vitro* splicing assays, the splicing efficiency of the LS2 binding motif-containing pre-mRNA was significantly decreased in the presence of purified recombinant LS2/dU2AF³⁸ heterodimer (Figure 9C lanes 9 and 10, Figure 9D), as well as in the presence of the uncomplexed LS2 protein (Figure 9E, Figure 10), indicating that LS2 has repressive activity, even in human splicing extracts. Additionally, both LS2 alone and the LS2/dU2AF³⁸ heterodimer repressed splicing of the G-rich motif-containing RNA in a concentration-dependent manner (Figure 9E, Figure 11). However, splicing of the substrate lacking the G-rich LS2 binding motif was unaffected by the addition of LS2/dU2AF³⁸ heterodimer or uncomplexed LS2 protein (Figure 9C lanes 4 and 5, Figure 9D,E), indicating that the

Chapter 4: Evolution of a tissue-specific splicing network

effect of LS2 is specific and dependent upon its ability to bind RNA through its specific recognition motif. The ability of LS2 to repress splicing without the need for the dU2AF³⁸ small subunit is consistent with the ability of dU2AF⁵⁰ and human U2AF⁶⁵ to activate splicing without dU2AF³⁸ or U2AF³⁵, respectively (Zamore et al. 1992; Kanaar et al. 1993).

LS2 was not able to substitute for the 3' splice site definition activity of dU2AF⁵⁰. That is, LS2/dU2AF³⁸ could not activate the splicing of substrates in which the polypyrimidine tract had been replaced by the G-rich LS2 recognition motif (Figure 11B).

Next, we asked whether LS2 also displayed similar activities *in vivo*. A minigene construct made from the *Drosophila* PEP gene containing a cassette exon was used to test the effect of LS2 in S2 cells (Figure 12A). Here, we inserted a G-rich LS2 recognition motif in the first intron 60 nt upstream of the 3' splice site. This motif was again upstream of both the polypyrimidine tract and the branchpoint adenosine. In this assay, splicing repression would be manifested near the cassette exon, leading to increased skipping of the internal exon. We monitored the exon inclusion levels of both the wild-type and motif-inserted minigenes by RT-PCR in response to the overexpression of LS2.

The basal level of inclusion of the cassette exon in this minigene was approximately 90% (Figure 12B,C). Overexpression of LS2 had very little effect on the splicing of the wild-type construct. Similarly, the insertion of a neutral, unrelated sequence motif 60 nt upstream of the 3' splice site had a very modest effect. However, insertion of an LS2 recognition motif significantly shifted the splicing toward exclusion of the cassette exon, likely due to the action of endogenous LS2. Moreover, unlike the wild-type construct, the splicing of the motif-containing construct was sensitive to the level of LS2 because over-expression of LS2 further shifted the splicing toward exon exclusion (Figure 12B,C). These results are consistent with the repressive activities of LS2 detected *in vitro*, its role as a potent splicing repressor, and the bioinformatically predicted positional enrichments of LS2 RNA binding motifs.

We also detected another LS2 binding motif enrichment, associated with exon inclusion, located approximately 120 nucleotides 3' of the downstream splice site (Figure 9A). Repression at the downstream splice site may kinetically allow splicing to occur at the cassette exon, causing its inclusion. Both enrichments are therefore consistent with the proposed function of LS2 as a splicing repressor.

Although previous studies had identified G-runs as important splicing regulatory motifs in mammals (Xiao et al. 2009), these runs were mainly associated with 5' splice sites and are bound by hnRNP H. The LS2 recognition sequence is not a G-run but rather a motif with guanosines enriched at specific positions. Additionally, its action as a splicing

Chapter 4: Evolution of a tissue-specific splicing network

repressor is greatly increased by over-expression of LS2, demonstrating its functional dependence on LS2 as a trans-acting factor.

LS2 interacts with its predicted targets in *Drosophila* S2 cells and has functionally diverged from dU2AF⁵⁰

To determine if LS2 interacts with its targets as predicted by the splice junction microarray, we performed immunoprecipitations of LS2 protein from stably transfected cells expressing polyoma-epitope-tagged LS2. Using PSI protein as a negative control, we determined that the immunoprecipitation was specific for LS2 (Figure 12D). We then used RT-PCR of anti-LS2-or non-immune IgG-immunoprecipitated RNA with gene-specific primers to assay for specific LS2 protein-RNA interactions. We detected a significant enrichment of several predicted target transcripts in the LS2 immunoprecipitates over both the input and negative control non-immune IgG immunoprecipitates (Figure 12E).

The action of LS2 as a splicing repressor rather than an activator demonstrates its functional divergence from dU2AF⁵⁰. Consistent with this divergence, activation of 3' splice sites could not be detected in constructs where the normal polypyrimidine tract was replaced by the LS2 G-rich sequence motifs (Figure 11B). If LS2 could serve as a surrogate, albeit opposite, form of dU2AF⁵⁰, we hypothesized that the polypyrimidine tracts of LS2-affected splice junctions would be weaker than expected. That is, several of the pyrimidines in the polypyrimidine tract would be replaced by guanines to allow LS2 binding. Toward this end, we analyzed the polypyrimidine tracts and 3' splice sites of targets of several alternative splicing factors, including LS2, as well as those of all 48,550 introns interrogated by the splice junction microarray, using MaxEntScore (Yeo and Burge 2004). This analysis only takes into account the last 22 nucleotides of the intron and therefore would not be expected to detect the motif enrichment that was detected 60 nt upstream of the 3' splice site (Figure 9A). Using this data, we conclude that the polypyrimidine tracts of LS2 target splice junctions are not any stronger or weaker than expected and do not contain anything resembling the G-rich LS2 SELEX motif (Figure 13A), indicating that LS2 has diverged from the 3' splice site-centric role of dU2AF⁵⁰. Interestingly, we also noted that the introns affected by LS2 are significantly longer than those affected by other characterized alternative splicing factors. (Figure 13B). While the median lengths of all *Drosophila* introns and those affected by dU2AF⁵⁰ knockdown were 85 and 121 nt, respectively, the median length of introns affected by LS2 knockdown was 422 nt (p value < 0.0001, student's t-test).

Although it is curious that knockdown of a core splicing factor like dU2AF⁵⁰ resulted in splicing changes at specific junctions and not a global downregulation in splicing, this was consistent with previous studies that had shown similar effects with *S. pombe*

Chapter 4: Evolution of a tissue-specific splicing network

U2AF temperature-sensitive mutants and RNAi depletion of *Drosophila* core spliceosome proteins (Sridharan et al. 2010; Park et al. 2004).

LS2 is specifically expressed in differentiated cells in the *Drosophila* testes

Recent mRNA-seq studies have shown that cells in the *Drosophila* testes undergo extensive changes in alternative splicing patterns upon differentiation, and testes, like the brain and nervous system, are known hotspots of alternative splicing in mammals (Venables and Eperon 1999; Elliott and Grellscheid 2006; Gan et al. 2010). In general, the overall complexity of alternative splicing events decreases upon differentiation in the *Drosophila* testis, when the testes stem cell population adopts more restricted cell fates as the spermatocytes develop and mature. Consistent with this decrease in the overall complexity of alternative splicing patterns, there is a concomitant decrease in expression of a majority of splicing factors during testes differentiation (Gan et al. 2010).

By contrast, this recent mRNA-seq study reported that LS2 is one of the few splicing factors whose expression increases dramatically upon testes differentiation (Gan et al. 2010). Consistent with this mRNA profiling data, immunofluorescence localization studies using flies expressing GFP-tagged Histone-2Av and affinity-purified anti-LS2 antibody indicated that while LS2 protein was expressed in differentiated spermatocytes, it was not expressed in the undifferentiated stem cells at the testis tip (Figure 14). These testis tips were phenotypically normal, however, as evidenced by the ample GFP fluorescence from the tagged histone in the tip.

LS2 interacts with SF1

Next, we wanted to investigate the mechanism by which LS2 represses the use of splice sites. If LS2 interacts with dU2AF³⁸ to recognize 3' splice site-like sequences, we reasoned that it may also interact with SF1 to recognize the branch point. Toward this end, we performed GST pulldowns similar to those in Figure 6A. Bacteria expressing full length SF1 cloned into pRSETa would not grow, presumably because of either the insolubility or toxicity of the protein to the bacterial cells. However, previous reports (Selenko et al 2003) had determined the region of human SF1 that was responsible for interacting with human U2AF⁶⁵. Although the domains of *Drosophila* SF1 seemed to be shuffled a bit relative to human SF1, it was possible to align the two sequences and determine the complementary region of *Drosophila* SF1 that would interact with dU2AF⁵⁰, and, possibly, LS2. This was contained within residues 200-375 of *Drosophila* SF1. This fragment then, was cloned into pRSETa with a FLAG tag and expressed in *E. coli*. GST pulldowns indicated that this fragment could interact with both dU2AF⁵⁰

Chapter 4: Evolution of a tissue-specific splicing network

and LS2 (Figure 16). Furthermore, Selenko et al. identified residues in RRM3 of U2AF⁶⁵ that were critical for interaction with SF1. These residues comprised part of one face of an alpha helix and were negatively charged. The corresponding residues in LS2 were also negatively charged, and point mutation of those residues to lysines abolished the ability of LS2 to interact with SF1 (Figure 16).

LS2 RNAi fly lines show no phenotype

To see if flies deficient in LS2 were sterile or showed some other phenotype, we constructed fly lines expressing hairpins against LS2 or, as a control, mCherry. A region of LS2 was targeted such that it would have no predicted off-target effects in the *Drosophila* genome. The hairpin-producing insert was cloned into pValium22 (<http://www.flyrnai.org/TRiP-REA.html>) and injected into embryos. After driving expression of the hairpin using an Act5C-Gal4 driver, expression of LS2 was knocked down by approximately 80-90% (Figure 17). However, neither males nor females from these lines were sterile, nor were there any other readily visible phenotypes, in the testes or otherwise.

DISCUSSION

Although the evolutionary patterns of splice sites and splice signals have been well documented (Brooks et al. 2010), little is known about how the proteins that recognize these sites and signals acquired their distinct functions and specificities. Many of these factors belong to large multigene families, with the SR proteins and hnRNP proteins being two notable examples (Dreyfuss et al. 1993; Shepard and Hertel 2009). It has been difficult, though, to determine how and when these family members diverged.

Our findings indicate that the *Drosophila* genome acquired a new gene encoding a novel splicing factor, LS2, more than 60 million years ago through a retro-transposition gene duplication event. The quickly-evolving LS2 gene subsequently diverged from its progenitor in its RNA binding sequence specificity, expression pattern and function to become an independent factor with a vastly different regulatory capacity and influence. We believe this to be the clearest example yet described of how gene duplication and divergence can result in the many related, yet distinct, splicing factors found in mammalian genomes. Furthermore, these results give an example of how these processes can transform a duplicated copy of a ubiquitously expressed and generally acting splicing factor into a tissue-specifically expressed and highly specialized component of a dedicated biological system.

Chapter 4: Evolution of a tissue-specific splicing network

Generally, new genes in *Drosophila* that are formed by retrotransposition events show a propensity to leave the X chromosome for the autosomes (Betran et al. 2002). More specifically, the phenomenon of acquisition of male-specific expression and function following gene duplication of an X-linked parental copy has been described for a multitude of genes in the *Drosophila* genome (Parisi et al. 2003). This may be due to the possibly disadvantageous overexpression of X-linked genes in males due to dosage compensation (Baker et al. 1994) or the increased risk of uncomplemented deleterious mutations due to X chromosome hemizyosity in males (Oliver 2002). The autosomal LS2 gene appears to be a very old instance of this phenomenon. dU2AF⁵⁰ is X-linked, and the LS2 ortholog is found on chromosome 2R in the same syntenic context in all 12 sequenced *Drosophila* genomes. The burst of protein sequence evolution common to all of the *Drosophila* LS2 orthologs combined with the maintenance of an intact, but fast-evolving open-reading frame, may allow for identification of specific LS2 amino acid residues that have undergone positive selection, a common fate for such genes (Proschel et al. 2006) during establishment and evolution.

In addition to showing a general male bias in expression and function, it has also been observed that many duplications of X-linked genes in *Drosophila* end up with a large testes-specific bias in expression (Bai et al. 2008), as is the case for the LS2 gene. Many of these genes, including LS2, have specifically identified motifs in their promoter regions that may contribute to a testes-biased expression pattern (Bai et al. 2009). Consistent with this, ChIP-seq data from the modEncode consortium show a large enrichment of acetylation at histone H3 lysine 9 (H3K9), usually associated with transcriptionally active regions, in the promoter region of LS2 in males but not in females (Liang et al. 2004; Celniker et al. 2009). LS2 gene expression increases significantly upon testis-cell differentiation and through its action as a splicing repressor may serve to suppress the many possible alternative splicing events typical of an undifferentiated stem cell in order to funnel the population of mature spliced mRNA isoforms toward a simpler, cell type-specific pattern. Although we cannot distinguish whether LS2 expression and function is a cause or consequence of testes differentiation, we suggest that LS2 acts to promote testes differentiation through its action on testes-important target pre-mRNA transcripts.

Because in mammals many RNA binding proteins are members of multigene families (Martinez-Contreras et al. 2007), similar evolutionary associations among related RNA binding protein family members are likely to exist in other organisms. Many of these factors and their functions have yet to be characterized. The relationship between LS2 and dU2AF⁵⁰ investigated here may provide a conceptual framework for future studies of the appearance and evolution of other splicing factors, including those of the multigene families commonly found in all mammalian genomes.

MATERIALS AND METHODS

dU2AF⁵⁰/LS2 sequence alignments

Sequence alignments of dU2AF⁵⁰ and LS2 were performed using ClustalW (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). Visualizations were done using Jalview.

dU2AF⁵⁰/LS2 phylogenetic analysis

Annotated orthologs of dU2AF⁵⁰ (CG9998) and LS2 (CG3162) were extracted from the eleven other *Drosophila* genus sequenced genomes. For each, the genomic context (neighboring genes) was manually inspected to confirm orthology. The closest homologs in mosquito (*Anophele gambiae*), honey bee (*Apsis mellifera*), and human were identified by blastp search. For each of these outgroups, the most similar protein sequence to **both** CG9948 and CG3162 was a single gene: XP_311994.3 for mosquito, XP_623055.1 for honey bee, and NP_001012496.1 for human. The genomic context in mosquito, honey bee, and human genomes was inspected and did not support closer relationship to dU2AF⁵⁰ or LS2 – none of the neighboring genes was shared with any of these species. Of note, however, is that in no case was the mosquito, honey bee, or human homolog X-linked.

This set of protein sequences was aligned using muscle v3.7 using default parameters. The protein sequence alignment was then used as a fixed guide to align the corresponding codons of each genes' coding sequence. Phylogeny of these sequences was inferred using MrBayes (v3.1.2) under a 3-partition model in which the 1st and 2nd codon positions evolve in a 2-state model with gamma rate-variation. The 3rd codon position was modeled to evolve under a 6-state model with a separate gamma rate. MCMC sampling was allowed to proceed for 900,000 generations of which 100,000 were discarded as burn-in. The resulting consensus tree and its clade credibility values are shown in Figure 2A.

RNAi

Double-stranded RNA (dsRNA) molecules of about 400 nucleotides in length were designed against dU2AF⁵⁰, dU2AF³⁸, and LS2 using SnapDragon to minimize off-target effects (http://www.flyrnai.org/cgi-bin/RNAi_find_primers.pl). dsRNA was prepared in vitro from a PCR fragment template containing T7 RNA polymerase start sites at both 5' ends. The resulting RNA was purified using RNeasy Midi Kit (Qiagen) and heated and cooled to anneal the strands. Control dsRNA was made from the backbone of pBluescript and contained no known matches to the *Drosophila* genome.

Chapter 4: Evolution of a tissue-specific splicing network

2 mL of S2 cells were seeded in M3 media supplemented with 5% FBS at approximately 1×10^6 cells /mL one day before dsRNA treatment. The media was then aspirated and 1 mL of serum-free M3 media was added with 10 μ g of dsRNA. The cells were incubated at 25°C for 1 hr, followed by the addition of 1 mL of M3 media supplemented with 10% FBS. The cells were then incubated at 25°C for 4 days. Total RNA was then harvested using RNeasy Mini Kit (Qiagen).

Splice-junction microarray analysis

For each microarray hybridization, 1 μ g of total RNA from the LS2 knock-down and 1 μ g of total RNA from a non-specific knockdown were amplified and converted to aRNA using the MessageAmp™ II aRNA Amplification Kit following the manufacturer recommendations (Ambion) and labeled with Cy5 and Cy3 mono-reactive dye, respectively (GE Healthcare). The custom splicing sensitive microarray used was based on FlyBase v5.15 and interrogates 49,364 annotated splicing events from 13,344 different genes with three overlapping 36 nt oligonucleotide probes, one centered at the splice junction and two probes offset by 3nt on each side of the splice junction. In addition, one fully exonic probe per 100 nt of each mRNA on average were added. The 348,650 different probes were randomly distributed onto two custom Agilent 220K arrays and used for hybridization of each cDNA sample. The microarrays were then processed and scanned following the manufacturer's recommendation (Agilent Technologies). The Feature Extraction reports were loaded into R and Lowess normalized using the marray package (Gentleman et al. 2004; Smyth 2004; Team 2010). The genes with affected AS were first identified using ANOVA comparing the group of exonic probes common to all transcripts to the different groups of splice junction probes corresponding to every splicing events of a given genes. The genes with Q-values smaller than 0.001 (adjusted using Benjamini-Hochberg correction) were then subjected to t-tests to identified the group of junction probes significantly affected with p-value ≤ 0.001 .

Array validations

To validate individual splice-junction microarray results, total RNA was isolated from S2 cells that had been treated with LS2 dsRNA or control dsRNA, as above. Reverse transcription was performed with random hexamers and HotStart PCR (Qiagen) was performed with gene specific primers that surrounded the junction of interest. The ratios of individual isoforms were quantitated with an Agilent 2100 Bioanalyzer.

LS2 and LS2/dU2AF³⁸ heterodimer purification

We purified bacterially expressed GST-tagged LS2 protein from Rosetta pLYS S cells after inducing expression overnight at 16°C with 1 mM IPTG. The cells were lysed using sonication in buffer A (50mM Tris pH 8.0, 100 mM KCl, 1M NaCl, 0.05% NP-40, 1 mM EDTA pH 8.0, 0.5 mM DTT, 1mM PMSF). The lysate was then incubated with glutathione-agarose beads (Sigma) and washed extensively with buffer A. GST-LS2 protein was then eluted using buffer A lacking NaCl and supplemented with 20 mM glutathione. Following dialysis in buffer B (20 mM HEPES-KOH pH 7.6, 20 mM KCl, 5% glycerol, 0.5 mM DTT, 0.4 mM PMSF), GST-LS2 was then further purified using a 1 mL Poros HS column on an AKTA protein purification system (GE Healthcare) using buffer B with 0M or 1M KCl. GST-LS2 Δ RS typically eluted at approximately 250 mM KCl, while GST-LS2 typically eluted at approximately 750 mM KCl. RNA SELEX was performed with GST-LS2 Δ RS to avoid possible nonspecific protein-RNA interactions due to the presence of the RS domain. See Figure 15 for an example of purified GST-LS2.

To purify the GST-LS2/dU2AF³⁸ heterodimer, a similar strategy was used. GST-LS2 protein was expressed as above. dU2AF³⁸ protein was expressed in pRSETA (Invitrogen) using Rosetta pLYS S cells, also inducing overnight at 16°C with 1 mM IPTG. Following lysis by sonication, lysates from 1L of GST-LS2-expressing cells were mixed with lysates from 4L of dU2AF³⁸-expressing cells at 4°C for 2 hr. Purification using glutathione-agarose beads and Poros HS was then performed as above. GST-LS2/dU2AF³⁸ heterodimer typically eluted from the heparin column at approximately 900 mM KCl.

LS2 SELEX

The random RNA library was made using an oligo with a T7 promotor at the 5' end and a 15 nt random portion made double-stranded by filling in with Klenow fragment DNA polymerase. In vitro transcription was done in the presence of $\alpha^{32}\text{P}$ -UTP. The starting random pool of RNA had the following sequence:

5'GACACGACGCTCTTCCGATCT(N₁₅)AGATCGGAAGAGCTCGTATGAAGCTTGTAAGCTTCATACG3'

Because of its low RNA affinity (inferred from previous experiments, later confirmed by gel shifts) 100 μL of 33 μM GST-LS2 Δ RS protein was incubated with 28 pmol (15,550 library coverages) RNA. LS2 Δ RS was used to minimize nonspecific interactions mediated by the highly positively charged RS domain. Binding was done in binding buffer (20 mM HEPES-KOH pH 7.6, 100 mM KCl, 2 mM MgCl₂, 0.5 mM DTT) at 25°C

Chapter 4: Evolution of a tissue-specific splicing network

for 30 min. The reaction was then applied to a nitrocellulose membrane (Millipore) pre-wet with binding buffer. The reaction was pulled through under gentle vacuum then washed with 5 mL binding buffer. A control reaction was done containing no protein. The percentage of RNA that was bound by protein was calculated by subtracting the counts per minute from the control filter from the counts per minute of the reaction filter and dividing by the counts per minute of the amount of RNA put into the binding reaction. The RNA was then recovered by phenol/chloroform extracting the membrane in 2.5M urea and ethanol precipitating. Reverse transcription and PCR were then done to prepare the template for the next cycle of enrichment. Five cycles of enrichment were performed, after which the selected molecules were sequence by Illumina sequencing using the following primers:

5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA
TCT 3'

5' CAAGCAGAAGACGGCATAACGAGCTCTTCCGATCT 3'

Motif analysis was done using MEME (Bailey and Elkan, 1994).

Electrophoretic mobility shift assays

RNAs containing the indicated sequences (Figure 4B,C) were synthesized using DNA oligonucleotides containing T7 promoters and α - 32 P-UTP. The RNA was then gel purified using a 15% polyacrylamide denaturing gel. The RNA was then dephosphorylated using calf intestine phosphatase (CIP; Roche) and then end-labeled using γ - 32 P-ATP and Optikinase (USB) to increase its specific activity.

Binding was done using the GST-LS2 or GST-LS2 Δ RS proteins and 0.1 nM RNA in a total volume of 20 μ L. The binding reaction contained 20 mM HEPES-KOH pH 7.6, 125 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 2 mM MgCl₂, and protein concentrations ranging from 10 μ M to 305 pM in 2-fold serial dilutions. Reactions were allowed to come to equilibrium for 45 min at room temperature. 4 μ L of 30% glycerol were then added and the reactions were run on a 4% polyacrylamide (60:1 acrylamide:bis) gel in 0.5X TBE buffer at 180V for 2 hr at 4°C. Gels were then dried, imaged using a Typhoon phosphorimager, and quantification of the reaction products was done using ImageQuant software (GE Healthcare).

Binding constants were determined using the fractions of bound and unbound RNA from the ImageQuant quantification and Prism 5 (Graphpad Software). The fractions of bound and unbound RNA were plotted and fitted to a standard binding model equation using Prism 5 (Graphpad Software). Data was plotted using the equation $Y = B_{max}(X) /$

Chapter 4: Evolution of a tissue-specific splicing network

($K_d + X$) where Y is the %RNA bound, B_{max} is the maximum %RNA bound, and X is the LS2 concentration.

LS2 / dU2AF³⁸ and LS2/SF1 interaction assays

GST-pulldowns were performed using GST-tagged dU2AF⁵⁰ protein or GST-tagged LS2 protein, purified as described above. Recombinant dU2AF³⁸ protein was also expressed as described above. To 1 mL of dU2AF³⁸-expressing *E. coli* lysate, 50 µg of purified dU2AF⁵⁰ large subunit was added. The final concentrations of the GST-LS2 and dU2AF³⁸ proteins were approximately 650 nM and 400 nM, respectively. The reaction mixture was then rotated at 4°C for 1 hr. 50 µL of glutathione agarose beads, washed in buffer A (see GST-LS2 protein purification, above) were then added and the reaction was rotated at 4°C for another hour. The beads were then pelleted and washed four times with 1 mL buffer A. The beads were then boiled in 50 µL of protein sample buffer. Samples were then run on an SDS-PAGE gel and analyzed by coomassie staining (data not shown) or immunoblotting.

SF1 200-375 was FLAG-tagged and cloned into pRSETa and expressed in the same manner as dU2AF³⁸. This fragment was reasonably well expressed and soluble. Pulldowns were performed exactly as described above with dU2AF³⁸, with the exception that 400 µL of SF1-overexpressed extract was added to 50 µg of LS2. Point mutants in LS2 were made using overlapping PCR and were purified in the same manner as the wildtype protein. There was no decrease in solubility or expression of the mutants.

For the coimmunoprecipitation of LS2 and dU2AF³⁸ proteins, RNP-enriched nuclear extracts from S2 cells stably expressing polyoma-tagged LS2 protein were used (Pinol-Roma et al., 1990). RNP-enriched extract was stored in HNEB2 (10 mM HEPES pH 7.6, 100 mM KCl, 2.5 mM MgCl₂, 0.2% NP-40, 0.2 mM PMSF). 25 µL of protein G beads (GE Healthcare) containing crosslinked anti-polyoma (GLU-GLU) antibodies were washed 3 times with 1 mL HNEB2. 50 µL of RNP-enriched extract were then added and the reaction was incubated at 4°C for one hr with rotation. For RNase-treated extracts, the extracts were pre-treated with 20 µg/mL RNase A for 20 min at room temperature and then allowed to continue digestion during incubation with the anti-polyoma beads (4°C for 1 hr). The beads were then washed 4 times with 1 mL wash buffer (20 mM HEPES-KOH pH 7.6, 400 mM LiCl, 2.5 mM MgCl₂, 0.2% NP-40, 0.2 mM PMSF, 0.5 mM DTT). The beads were then boiled in 25 µL SDS protein sample buffer and analyzed by SDS-PAGE and immunoblotting.

Motif enrichment in affected transcripts

The SELEX data were analyzed using MEME (Bailey and Elkan 1994) and the position-specific scoring matrix (PSSM) of the preferred LS2 binding sites was used to search the LS2-affected transcripts identified from the splice junction microarray, compared to the rest of the expressed transcriptome not affected by LS2 RNAi-knockdown. The relative fraction of transcripts containing at least 1 LS2 binding site with different motif scores was calculated and plotted. The error bars correspond to the bootstrapped standard deviation of the population of transcripts at the different motif score.

Motif placement in affected transcripts

In order to analyze the enrichment of the LS2 binding motif(s) in genes with LS2 RNAi knockdown-affected AS, only simple AS patterns corresponding to alternative cassette exons, competing donor sites, competing acceptor sites and intron retention events were considered for modeling purposes. The affected AS events from every simple splice pattern type were further divided into two groups, either positively or negatively affecting exon inclusion of the longer isoform. A 400 nt region surrounding each affected splice site of the corresponding splicing events was used to identify the best motif score within a window of 50 nt. For each window, a t-test was performed to compare the population of motif scores from the affected events to the population of the best motif scores of the corresponding window in all of the other known *Drosophila* AS events of a given type but NOT affected in the LS2 RNAi-knockdown samples. The p-values from these tests were plotted underneath every splice type analyzed. The red bars correspond to regions of at least five consecutive best score windows with p-values ≤ 0.05 separated by at most three windows above the p-value cutoff.

Testis enrichment of LS2 protein and its pre-mRNA targets

Whole cell lysates containing approximately $\frac{1}{8}$ of a whole fly, 1 head, and 1 pair of testes were separated by SDS-PAGE and immunoblotted for the PSI and LS2 proteins.

Gene ontology analysis of the LS2-affected transcripts was done using BABELOMICS v3.2 (Al-Shahrour et al. 2006) (<http://babelomics3.bioinfo.cipf.es/>).

The testis mRNA expression levels of LS2 RNAi knockdown-affected transcripts were calculated using expression microarray data from FlyAtlas (Chintapalli et al. 2007) (www.flyatlas.org). The mean fluorescence level from four independent Affymetrix Dros2.0 expression arrays was used as the testes expression level of that particular gene. For the S2 cell data sets, only genes identified as present in S2 cells in at least 1

Chapter 4: Evolution of a tissue-specific splicing network

out of 4 microarray experiments were used. For the LS2 RNAi knockdown-affected data set, only the 168 genes whose splicing was changed upon knockdown of LS2 were used.

In vivo splicing assays

Cassette exon constructs containing exons 1-3 of the PEP (CG6143) gene either without or with 2 LS2-recognition motifs (GGCGGCGGTGGGGGGTGGTGGCGGG) or a neutral motif (TGCACCCTCTGATGCACCCTCTGA) inserted 60 nt upstream of the cassette exon were created using overlap PCR. These constructs were cloned into pMT-V5-His (Invitrogen) and their expression was under the control of the metallothionein promoter. To overexpress LS2, we used an LS2-cDNA cloned into pUC-hyg-MT such that expression of LS2 was also under control of the metallothionein promoter.

24 hr before transfection, 2mL of 1×10^6 cells /mL were seeded in a 6-well plate. The cells were then transfected with 0.5 μ g each of PEP-motif, PEP+motif, *or* PEP+neutral motif-containing plasmid DNA, and pUC-hyg-MT-LS2 *or* blank pUC-hyg-MT plasmids. Transfections were done using Effectene (Qiagen). One day later, Cu₂SO₄ was added to 50 μ M. Two days after copper addition, the cells were harvested and total RNA was isolated. Reverse transcription was done using random hexamers, and PCR was done using specific primers that amplify the exogenous PEP and not the endogenous PEP. The quantities and sizes of the RT-PCR products were then analyzed using an Agilent 2100 Bioanalyzer.

We also performed *in vivo* splicing assays as described in the Experimental Procedures with a construct containing an LS2 motif insertion in the downstream exon (see Figure 12D). However, we could not recover RNA from this construct, likely due to destabilization of the RNA with this exonic insertion. The stability of RNAs containing intronic insertions (see Figure 12A-C) did not seem to be affected.

In vitro splicing assays

In vitro splicing assays were performed as previously described (Padgett et al., 1983).

Similar to the *in vivo* splicing assays, an LS2 recognition motif was inserted into the ftz intron 65 nucleotides upstream of the 3' splice site. The ftz intron was transcribed *in vitro* using T7 RNA polymerase and α^{32} P-UTP. It was then gel purified using a 5% polyacrylamide denaturing gel. In vitro splicing reactions were then set up in 20 μ L final volumes with 8 μ L HeLa nuclear extract, 8 μ L 2.5X SP mix, 3 μ L LS2/dU2AF³⁸

Chapter 4: Evolution of a tissue-specific splicing network

heterodimer (~500 ng) or blank buffer, and 1 μ L RNA (20 fmol, ~20,000 cpm). 2.5X SP mix contained the following: 5 mM ATP, 50 mM creatine phosphate, 25% glycerol, 50 mM HEPES pH 7.6, 7.5% PEG 8000, 62.5 mM potassium glutamate, 10 mM $MgCl_2$. The KCl concentrations of the HeLa nuclear extract and heterodimer fractions were both 100 mM. The final concentrations of glutamate, chloride, and potassium were therefore 25 mM, 55 mM, and 80 mM, respectively.

The reactions were incubated at 30°C for 3 hr, then phenol/chloroform extracted, ethanol precipitated, and washed once with 70% ethanol. They were then resuspended in 10 μ L urea/bromophenol blue/xylene cyanol and subjected to denaturing gel electrophoresis on a pre-run 0.4 mm thick 12% polyacrylamide-urea gel for 7 hr at 25W. The gel was then fixed for 20 min in 10% methanol, 10% acetic acid. It was then dried and exposed using a phosphorimager. Quantitation was done using a Typhoon Phosphorimager with ImageQuant software (GE Healthcare).

Quantitation was done by first normalizing the intensity of each band according to its length in nucleotides. The spliced ratio was then calculated by adding up the intensities of all splicing intermediates and products and dividing by the unspliced pre-mRNA. Those results were then normalized by setting the splicing efficiency of WT ftz pre-mRNA in the absence of heterodimer to 1.0.

RNA immunoprecipitation (RIP) of predicted LS2 target pre-mRNAs

Polyoma (Glu-Glu) tagged LS2 protein was immunoprecipitated from S2 nuclear RNP-enriched extracts (Pinol-Roma et al. 1990). The extracts were stored in HNEB2 (see above). 900 μ L extract were incubated with 100 μ L beads containing anti-polyoma antibodies that had been washed four times with 1 mL HNEB2. As a control, an immunoprecipitation using IgG was also done. The reaction was incubated at 4°C for 4 hr. The resin was then washed four times with 1 mL wash buffer (20 mM HEPES pH 7.6, 100 mM KCl, 5 mM $MgCl_2$, 0.5 mM DTT, 0.2 mM PMSF, 50 U/mL RNasin (Promega)). The beads were then resuspended in 100 μ L 1X RQ1 DNase buffer (Promega). 5 units of RQ1 DNase (Promega) were then added and the reaction was incubated at 37°C for 1 hr. RNA was then eluted by phenol/chloroform extracting the beads and ethanol precipitating. The pellet was washed twice with 1 mL 70% ethanol. The pellet was then resuspended in 15 μ L H_2O . RNA concentration was measured using a Nanodrop spectrophotometer. Equal amounts of polyoma or IgG immunoprecipitated RNA and RNA isolated from the starting RNP-enriched extracts were then used for RT-PCR using random hexamers. Individual bound transcripts were then assayed using HotStart PCR (Qiagen) and gene-specific primers.

Chapter 4: Evolution of a tissue-specific splicing network

Testes immunofluorescence

Testes from 1-3 day old males expressing GFP-tagged His2Av were dissected into cold Ringers solution (0.35 g NaCl in 50 mL water). Approximately ten pairs of testes were then fixed using 1X PBX (PBS supplemented with 0.1% Triton X-100 and 0.5% BSA) with 4% formaldehyde at room temperature for 15 min. The testes were then washed 3 times with 1X PBX for 2 min. each.

Blocking was done for 1 hr at room temperature in 1mL 2% normal goat serum. Fixed testes were then incubated with primary antibody diluted 1:500 in 1X PBX overnight at 4°C. They were then washed 3 times with 1mL 1X PBX for 15 min each. Secondary (donkey anti-rabbit AlexaFluor 568) was then added at 1:400 dilution for 2 hr at room temperature. Testes were then washed 3 times with 1mL 1X PBX for 15 min. each. Testes were then mounted on a slide and imaged using an Axioimager 373 microscope.

Splice site strength and intron length analyses

The strength of individual splice sites was scored using MaxEntScore (Yeo and Burge, 2004) (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html). Data concerning all introns in the *Drosophila* genome was retrieved from FlyBase (http://flybase.org/static_pages/downloads/ID.html). All plots were made using Prism v5.0 (GraphPad Software).

LS2 RNAi flies

The following oligos were ligated in to pValium22 (contains UAS sequences and v⁺) that had be treated with NheI and EcoRI:

Top

5'

CTAGCAGTCGATTGCAGCGACGTTTCAGTTAGTTATATTCAAGCATAACTGAAACGT
CGCTGCAATCGGCG 3'

Bottom

5'

AATTCGCCGATTGCAGCGACGTTTCAGTTATGCTTGAATATAACTAACTGAAACGTC
GCTGCAATCGACTG 3'

Chapter 4: Evolution of a tissue-specific splicing network

These oligos, upon annealing, made overhanging ends that were complimentary to those produced by the restriction digestion of the plasmid. This construct was then injected into yv embryos by Genetic Services Inc. into attP2 sites on chr3L. These were then crossed to yv ; Dr / TM3 to create a yv ; LS2HP/ TM3 stock. Homozygotes of this stock were viable and fertile and were crossed to various Gal4 driver stocks. The most efficient knockdown was seen with the Act5C-Gal4 driver.

REFERENCES

Akerman M, David-Eden H, Pinter RY, Mandel-Gutfreund Y. 2009. A computational approach for genome-wide mapping of splicing factor binding sites. *Genome Biol* **10**: R30.

Al-Shahrour F, Minguez P, Tarraga J, Montaner D, Alloza E, Vaquerizas JM, Conde L, Blaschke C, Vera J, Dopazo J. 2006. BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Res* **34**: W472-476.

Amarasinghe AK, MacDiarmid R, Adams MD, Rio DC. 2001. An in vitro-selected RNA-binding site for the KH domain protein PSI acts as a splicing inhibitor element. *RNA* **7**: 1239-1253.

Baek D, Green P. 2005. Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. *Proc Natl Acad Sci U S A* **102**: 12813-12818.

Bai Y, Casola C, Betran E. 2008. Evolutionary origin of regulatory regions of retrogenes in *Drosophila*. *BMC Genomics* **9**: 241.

Bai Y, Casola C, Betran E.. 2009. Quality of regulatory elements in *Drosophila* retrogenes. *Genomics* **93**: 83-89.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**: 28-36.

Baker BS, Gorman M, Marin I. 1994. Dosage compensation in *Drosophila*. *Annu Rev Genet* **28**: 491-521.

Chapter 4: Evolution of a tissue-specific splicing network

Ben-Dov C, Hartmann B, Lundgren J, Valcarcel J. 2008. Genome-wide analysis of alternative pre-mRNA splicing. *J Biol Chem* **283**: 1229-1233.

Betran E, Thornton K, Long M. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Res* **12**: 1854-1859.

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

Blanchette M, Green RE, Brenner SE, Rio DC. 2005. Global analysis of positive and negative pre-mRNA splicing regulators in *Drosophila*. *Genes Dev* **19**: 1306-1314.

Blanchette M, Green RE, MacArthur S, Brooks AN, Brenner SE, Eisen MB, Rio DC. 2009. Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the *Drosophila* hnRNP A/B family members. *Mol Cell* **33**: 438-449.

Brooks AN, Yang L, Duff MO, Hansen KD, Park JW, Dudoit S, Brenner SE, Graveley BR. 2010. Conservation of an RNA regulatory map between *Drosophila* and mammals. *Genome Res*.

Buckanovich RJ, Posner JB, Darnell RB. 1993. Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron* **11**: 657-672.

Celniker SE, Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD, MacAlpine DM et al. 2009. Unlocking the secrets of the genome. *Nature* **459**: 927-930.

Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet* **39**: 715-720.

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

Elliott DJ, Grellscheid SN. 2006. Alternative RNA splicing regulation in the testis. *Reproduction* **132**: 811-819.

Chapter 4: Evolution of a tissue-specific splicing network

Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K, Chen X. 2010. Dynamic regulation of alternative splicing and chromatin structure in *Drosophila* gonads revealed by RNA-seq. *Cell Res* **20**: 763-783.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**: R80.

Kanaar R, Roche SE, Beall EL, Green MR, Rio DC. 1993. The conserved pre-mRNA splicing factor U2AF from *Drosophila*: requirement for viability. *Science* **262**: 569-573.

Kent OA, Reayi A, Foong L, Chilibeck KA, MacMillan AM. 2003. Structuring of the 3' splice site by U2AF65. *J Biol Chem* **278**: 50572-50577.

Kielkopf CL, Rodionova NA, Green MR, Burley SK. 2001. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* **106**: 595-605.

Kikuchi T, Ichikawa M, Arai J, Tateiwa H, Fu L, Higuchi K, Yoshimura N. 2000. Molecular cloning and characterization of a new neuron-specific homologue of rat polypyrimidine tract binding protein. *J Biochem* **128**: 811-821.

Liang G, Lin JC, Wei V, Yoo C, Cheng JC, Nguyen CT, Weisenberger DJ, Egger G, Takai D, Gonzales FA et al. 2004. Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc Natl Acad Sci U S A* **101**: 7357-7362.

Markovtsov V, Nikolic JM, Goldman JA, Turck CW, Chou MY, Black DL. 2000. Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol Cell Biol* **20**: 7463-7479.

Martinez-Contreras R, Cloutier P, Shkreta L, Fiset JF, Revil T, Chabot B. 2007. hnRNP proteins and splicing control. *Adv Exp Med Biol* **623**: 123-147.

Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J. 1999. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG.

Chapter 4: Evolution of a tissue-specific splicing network

Nature **402**: 838-841.

Nelson KK, Green MR. 1988. Splice site selection and ribonucleoprotein complex assembly during in vitro pre-mRNA splicing. *Genes Dev* **2**: 319-329.

Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**: 457-463.

Oliver B. 2002. Genetic control of germline sexual dimorphism in *Drosophila*. *Int Rev Cytol* **219**: 1-60.

Padgett RA, Hardy SF, Sharp PA 1983. Splicing of adenovirus RNA in a cell free transcription system. *Proc. Nat. Acad. Sci. USA* **80**: 5230-5234.

Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Oliver B. 2003. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* **299**: 697-700.

Park JW, Parisky K, Celotto AM, Reenan RA, Graveley BR. 2004. Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc Natl Acad Sci U S A* **101**: 15974-15979.

Pinol-Roma S, Swanson MS, Matunis MJ, Dreyfuss G. 1990. Purification and characterization of proteins of heterogeneous nuclear ribonucleoprotein complexes by affinity chromatography. *Methods Enzymol* **181**: 326-331.

Proschel M, Zhang Z, Parsch J. 2006. Widespread adaptive evolution of *Drosophila* genes with sex-biased expression. *Genetics* **174**: 893-900.

R Development Core Team. 2010. R: A Language and Environment for Statistical Computing.

Rudner DZ, Breger KS, Kanaar R, Adams MD, Rio DC. 1998a. RNA binding activity of heterodimeric splicing factor U2AF: at least one RS domain is required for high-affinity binding. *Mol Cell Biol* **18**: 4004-4011.

Chapter 4: Evolution of a tissue-specific splicing network

Rudner DZ, Kanaar R, Breger KS, Rio DC. 1998b. Interaction between subunits of heterodimeric splicing factor U2AF is essential in vivo. *Mol Cell Biol* **18**: 1765-1773.

Ruskin B, Zamore PD, Green MR. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**: 207-219.

Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Kramer A, Sattler M. 2003. Structural basis for the molecular recognition between human splicing factors U2AF(65) and SF1/mBBP. *Molecular Cell* **11**: 965-976.

Shepard PJ, Hertel KJ. 2009. The SR protein family. *Genome Biol* **10**: 242.

Sickmier EA, Frato KE, Shen H, Paranawithana SR, Green MR, Kielkopf CL. 2006. Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol Cell* **23**: 49-59.

Singh R, Banerjee H, Green MR. 2000. Differential recognition of the polypyrimidine-tract by the general splicing factor U2AF65 and the splicing repressor sex-lethal. *RNA* **6**: 901-911.

Singh R, Valcarcel J, Green MR. 1995. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**: 1173-1176.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: Article3.

Sridharan V, Heimiller J, and Singh R. 2010. Genomic mRNA profiling reveals compensatory mechanisms for the requirement of the essential splicing factor U2AF. *Mol Cell Biol*. Epub ahead of print.

Tamura K, Subramanian S, Kumar S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* **21**: 36-44.

Venables JP, Eperon I. 1999. The roles of RNA-binding proteins in spermatogenesis and male infertility. *Curr Opin Genet Dev* **9**: 346-354.

Chapter 4: Evolution of a tissue-specific splicing network

Wang Z, Burge CB. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**: 802-813.

Wu S, Romfo CM, Nilsen TW, Green MR. 1999. Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**: 832-835.

Xiao X, Wang Z, Jang M, Nutiu R, Wang ET, Burge CB. 2009. Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat Struct Mol Biol* **16**: 1094-1100.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**: 377-394.

Yu Y, Maroney PA, Denker JA, Zhang XH, Dybkov O, Luhrmann R, Jankowsky E, Chasin LA, Nilsen TW. 2008. Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. *Cell* **135**: 1224-1236.

Zamore PD, Green MR. 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A* **86**: 9243-9247.

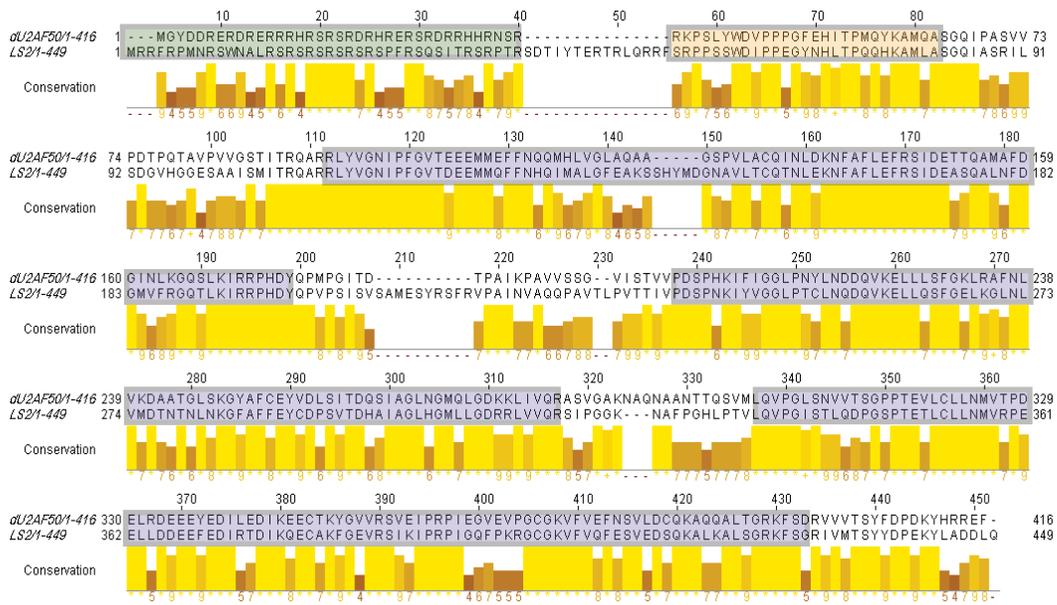
Zamore PD, Patton JG, Green MR. 1992. Cloning and domain analysis of the mammalian splicing factor U2AF. *Nature* **355**: 609-614.

Zorio DA, and Blumenthal T. 1999. Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* **402**: 835-838.

Chapter 4: Evolution of a tissue-specific splicing network

Figure 1. dU2AF⁵⁰/LS2 sequence alignment. Primary sequence alignment of dU2AF⁵⁰ and LS2. Yellow bars represent degrees of conservation with higher yellow bars representing more conservative substitutions. The RS domain of both proteins is highlighted in green, the dU2AF³⁸ interaction domain in beige, and the 3 RNA recognition motifs in purple.

Chapter 4: Evolution of a tissue-specific splicing network



Chapter 4: Evolution of a tissue-specific splicing network

Figure 2. LS2 arose and diverged in function from dU2AF⁵⁰ in *Drosophila*. **(A)** Phylogenetic tree of LS2, dU2AF⁵⁰, and orthologs from honey bee, mosquito, and human. The non-*Drosophila* gene sequences are the single, most similar genes to both dU2AF⁵⁰ and LS2 in each of the 3 outgroup genomes. Clades with >90 credibility value are denoted with a black circle. **(B)** Venn diagram showing the overlap of splice junctions affected by dU2AF⁵⁰, dU2AF³⁸, and LS2 RNAi knockdown. **(C)** A scatter plot of splice junctions affected by dU2AF⁵⁰ and LS2 RNAi knockdown. Axes represent the log₂ change of splice junction intensity in response to RNAi of the indicated protein.

Chapter 4: Evolution of a tissue-specific splicing network

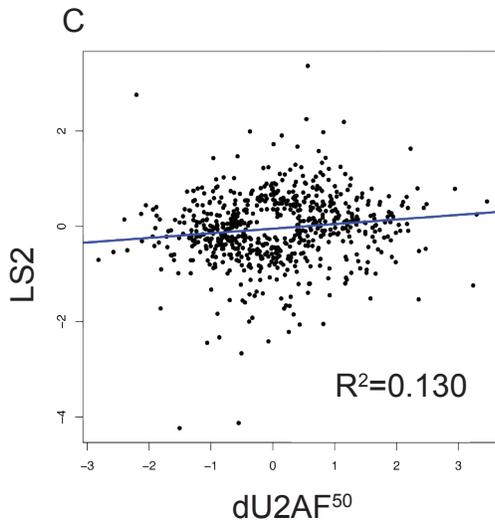
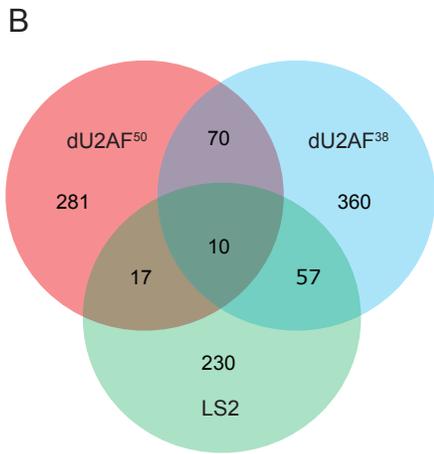
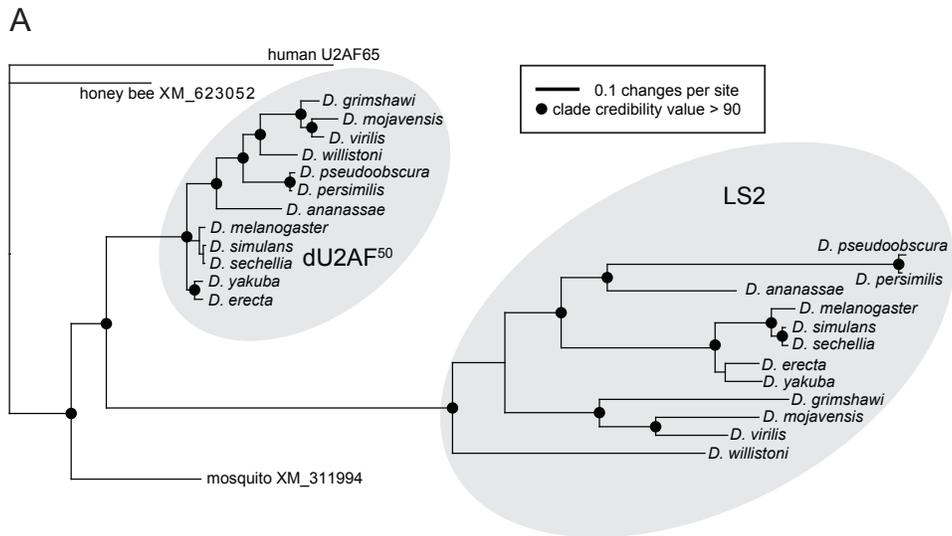


Figure 3. dU2AF⁵⁰/LS2 splice junction microarray results. **(A)** Efficiency of LS2 RNAi measured by RT-PCR of LS2 (Duplicate biological replicates; lanes 3 and 4) and an unrelated negative control gene, CG5005 (Ctrl; Duplicate biological replicates; lanes 1 and 2). **(B)** Total numbers of splice junction probes and genes affected by RNAi of dU2AF⁵⁰, dU2AF³⁸, and LS2. **(C)** Similar to Figure 1C, a scatter plot of change in splice junction intensity in response to RNAi knockdown of two different proteins. Y-axis values represent the log₂ change in intensity in response to knockdown of the indicated protein. **(D)** Validation of selected splicing changes from the microarrays by RT-PCR. Y-axis values represent the log₂ change in isoform ratio in response to treatment with control dsRNA or LS2 dsRNA. Error bars represent standard deviations from 3 technical replicates of 2 biological replicates.

Chapter 4: Evolution of a tissue-specific splicing network

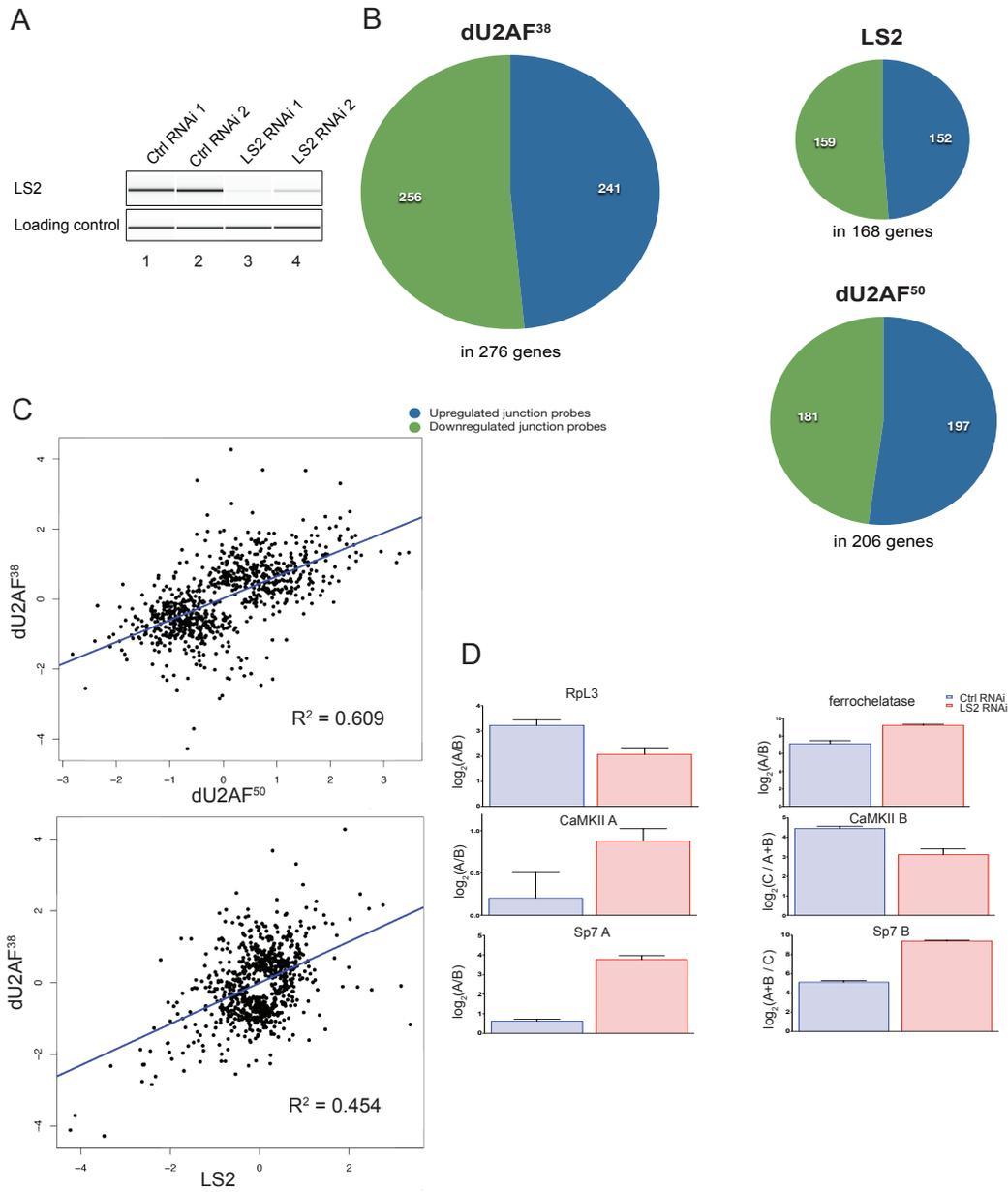


Figure 4. LS2 has diverged in RNA binding sequence specificity from dU2AF⁵⁰ and its binding motif is enriched in its target transcripts. (A) SELEX-derived PSSM (position-specific scoring matrix) of the RNA sequence recognized by the LS2 protein. (B) and (C) Electrophoretic gel mobility shift assays using purified recombinant GST-tagged LS2 protein and a synthetic RNA containing the SELEX-derived G-rich recognition motif (B) or a mutant RNA in which all the important guanosine residues (see motif) had been mutated to cytosine (C). Protein concentrations ranged from 305 pM (lane 1) to 10 μ M (lane 16) in 2-fold increments. Lane 17, no protein control. (D) Phosphorimager quantification of the results in A and B. Similar experiments were done using GST-tagged truncated versions of LS2 lacking the N-terminal RS domain (data not shown). (E) Enrichment of the LS2 recognition motif in its affected target transcripts. Each point represents the fraction of genes that contain an LS2 recognition motif scoring at the x-axis value or higher.

Chapter 4: Evolution of a tissue-specific splicing network

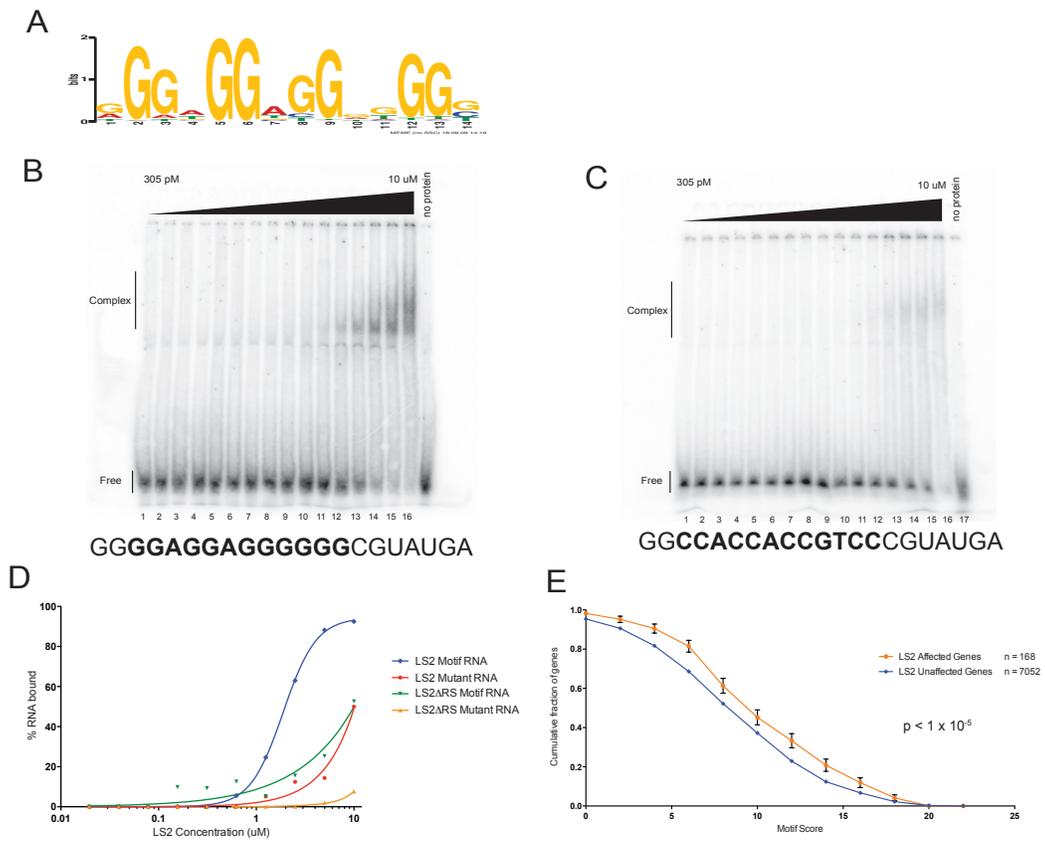
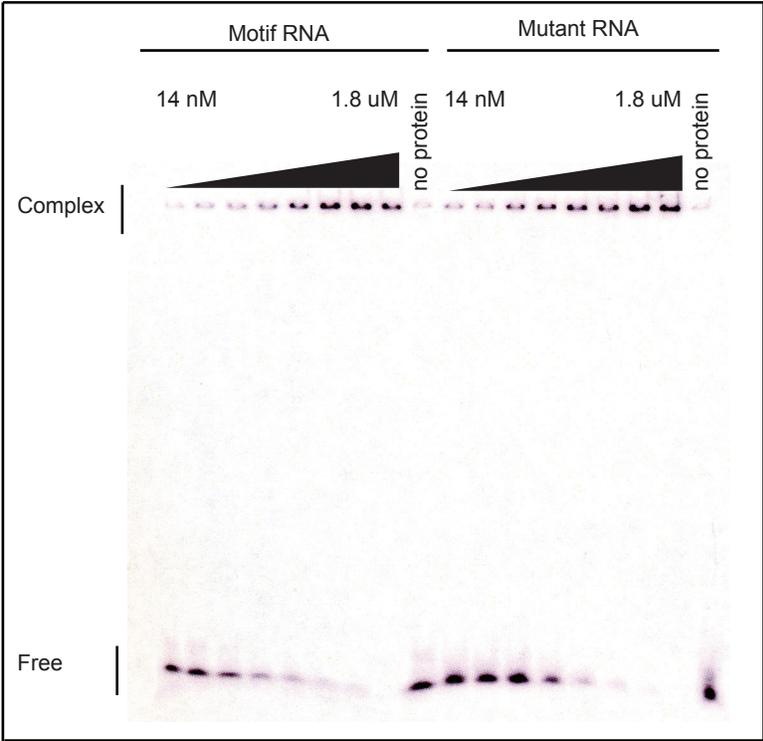


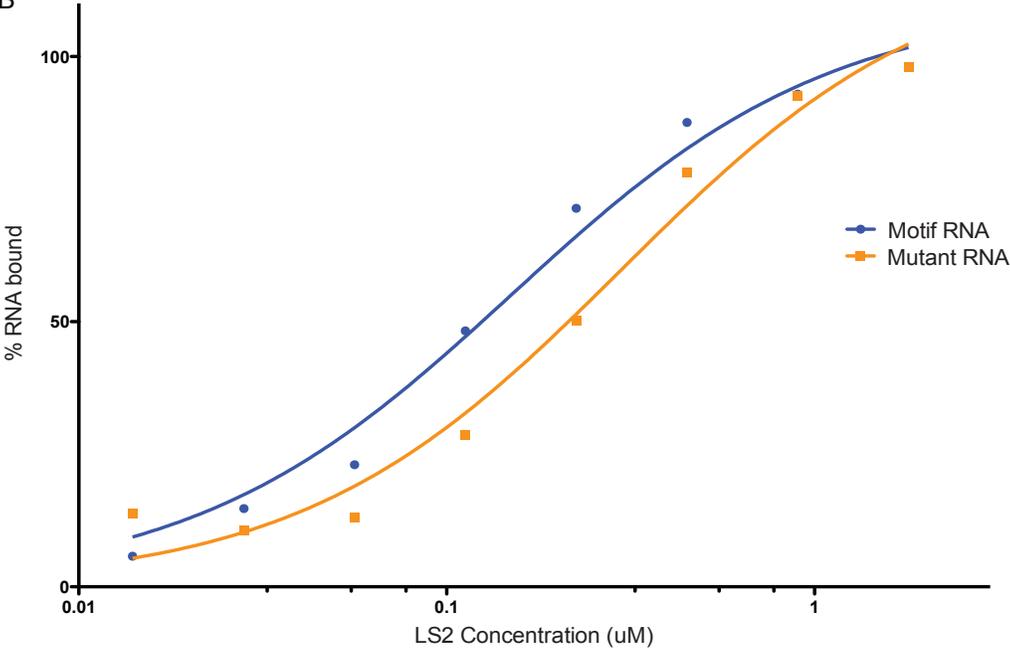
Figure 5. RNA binding activity of LS2 / dU2AF³⁸ heterodimer. (A) RNA probes were prepared as in figure 2. RNA probes containing G-rich or mutant sequences (see figure 2 for RNA sequences) were incubated with purified recombinant LS2 / dU2AF38 heterodimer of concentrations ranging from 1800 nM to 14 nM in 2-fold serial dilutions. (B) Phosphorimager quantitation results of A. Quantitation was performed as in figure 2. For details of the quantitation, see supplemental methods. The apparent K_d of the heterodimer for the G-rich RNA was 150 nM and for the mutant RNA was 297 nM.

Chapter 4: Evolution of a tissue-specific splicing network

A



B



Chapter 4: Evolution of a tissue-specific splicing network

Figure 6. LS2 physically interacts with dU2AF³⁸. (A) Immunoblots from GST-pull-down experiments using purified recombinant GST-tagged dU2AF⁵⁰ (lanes 1 and 2), LS2 (lanes 3 and 4), or an LS2 truncation lacking the putative dU2AF³⁸ interaction domain (lanes 5 and 6) and *E. coli* lysates containing His-tagged dU2AF³⁸ (see Supplementary Fig. 1). Eluates from the pull-downs were then immunoblotted using anti-dU2AF⁵⁰ and anti-LS2 (top panel), or anti-dU2AF³⁸ (bottom panel) antibodies. (B) Immunoblot analysis from co-immunoprecipitation experiments performed with epitope-tagged LS2 expressed in S2 cells. Polyoma (GLU-GLU) epitope-tagged LS2 was immunoprecipitated from S2 extracts in the presence (+; lanes 1 and 2) or absence (-; lanes 3 and 4) of RNase A and the precipitates were immunoblotted using anti-LS2 (top panel) or dU2AF³⁸ (bottom panel) antibodies. In all cases the immunopurified proteins were compared to input lysate lanes (Input, lanes 1, 3 and 5). To show specificity, similar experiments were done using FLAG-tagged LS2 (negative control; lanes 5 and 6).

Chapter 4: Evolution of a tissue-specific splicing network

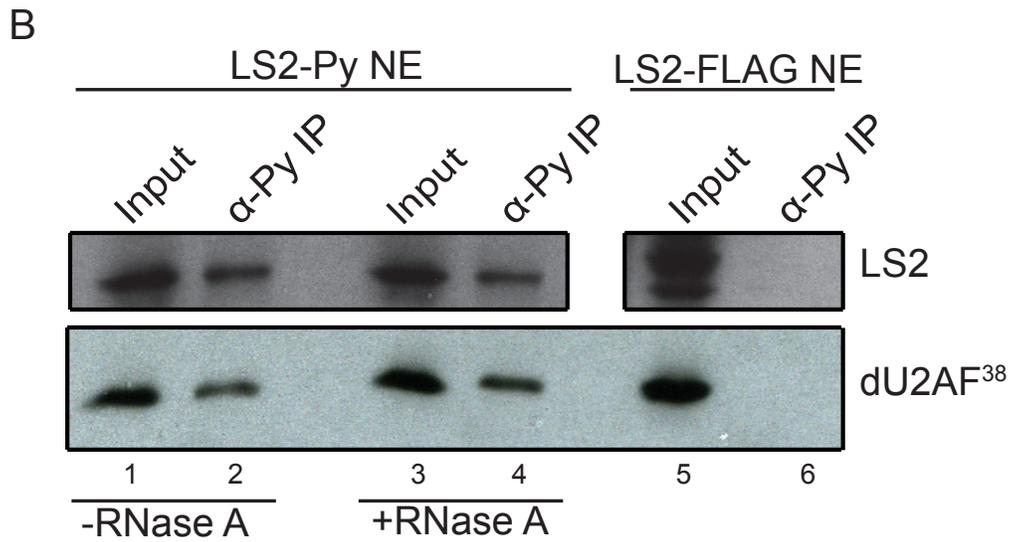
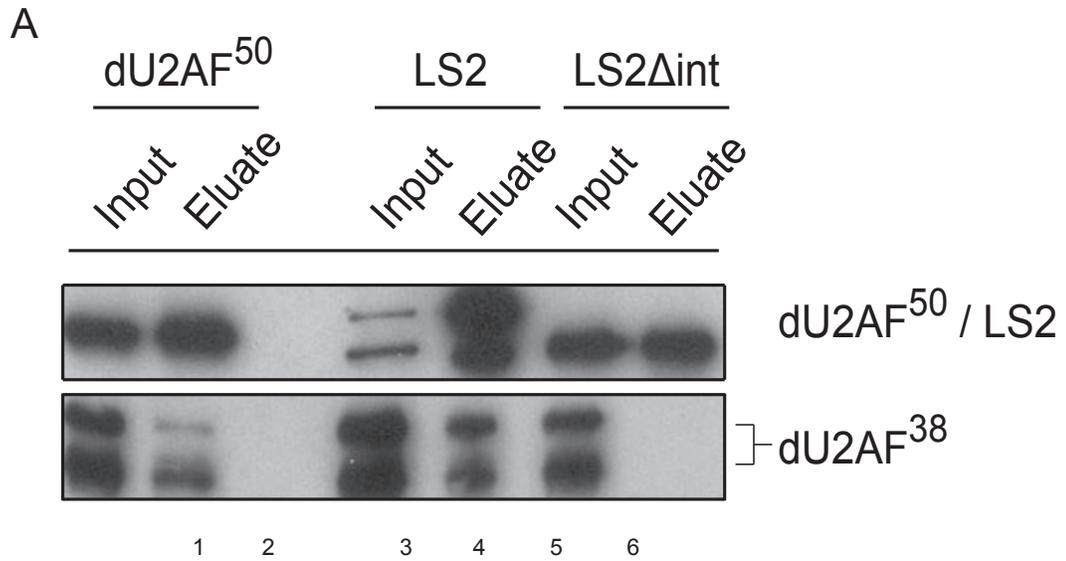


Figure 7. LS2 and its target transcripts are enriched in testes and regulate testes-related functions. (A) Immunoblot analysis of lysates from whole flies, heads, and testes probed with anti-PSI (loading control) and anti-LS2 antibodies. (B) Enriched gene ontology terms for the LS2-affected genes (Al-Shahrour et al. 2006). P values were calculated using a two-tailed Fisher's exact test. (C) Testes expression levels of all genes, genes expressed in S2 cells and genes affected by LS2. The y-axis is the mean expression level from four Affymetrix Dros2 expression arrays (Chintapalli et al. 2007). Whiskers represent the maximum and minimum values, boxes represent the 25th and 75th percentiles, crosses represent the mean value, and lines represent the median value. P values were calculated using a two-tailed t test.

Chapter 4: Evolution of a tissue-specific splicing network

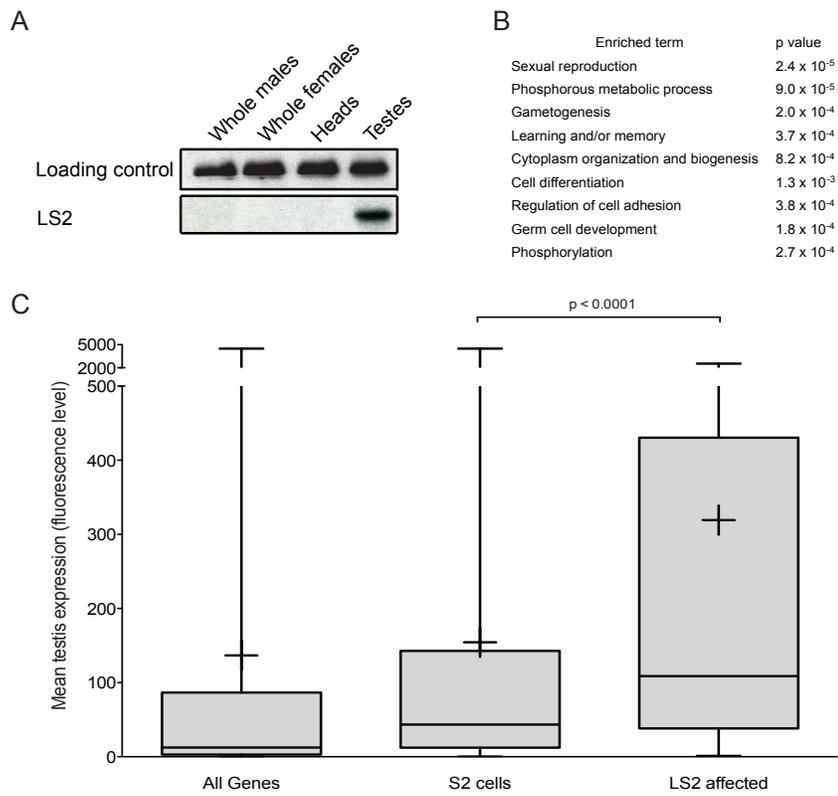


Figure 8. Gene ontology networks in LS2, dU2AF⁵⁰ and dU2AF³⁸ affected transcripts. Genes identified as splicing regulatory targets of LS2 (A), dU2AF⁵⁰ (B) and dU2AF³⁸ (C) were analyzed for enrichment of specific gene ontology networks using BABELOMICS (Al-Shahrour et al. 2006) (see Supplemental Table 1). Enriched terms are hierarchically organized such that the enriched biological terms get more general as they proceed up the graph. The statistical significance of each enrichment is indicated by color with orange labels indicating high statistical significance and beige labels indicated low statistical significance.

Chapter 4: Evolution of a tissue-specific splicing network

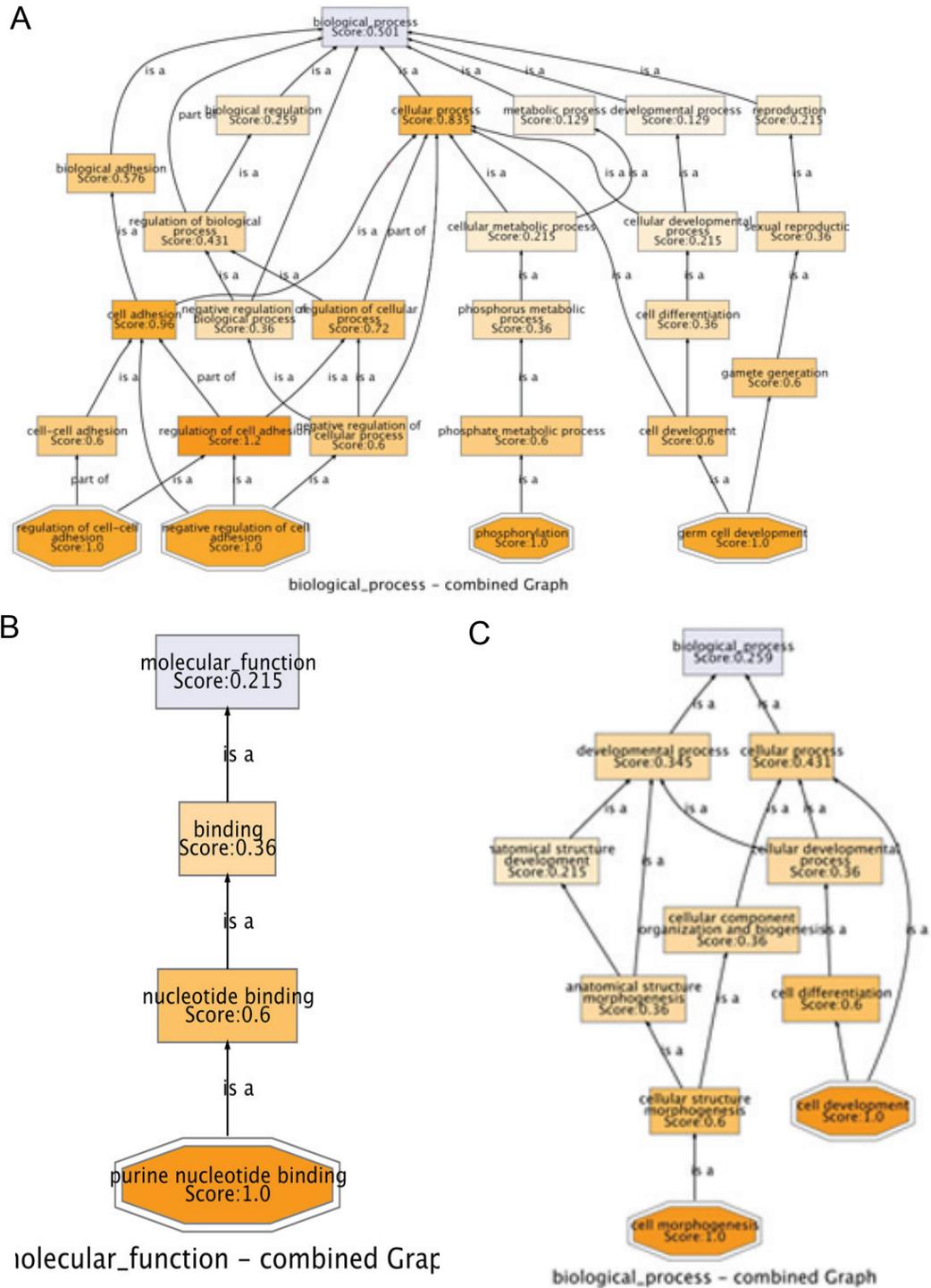
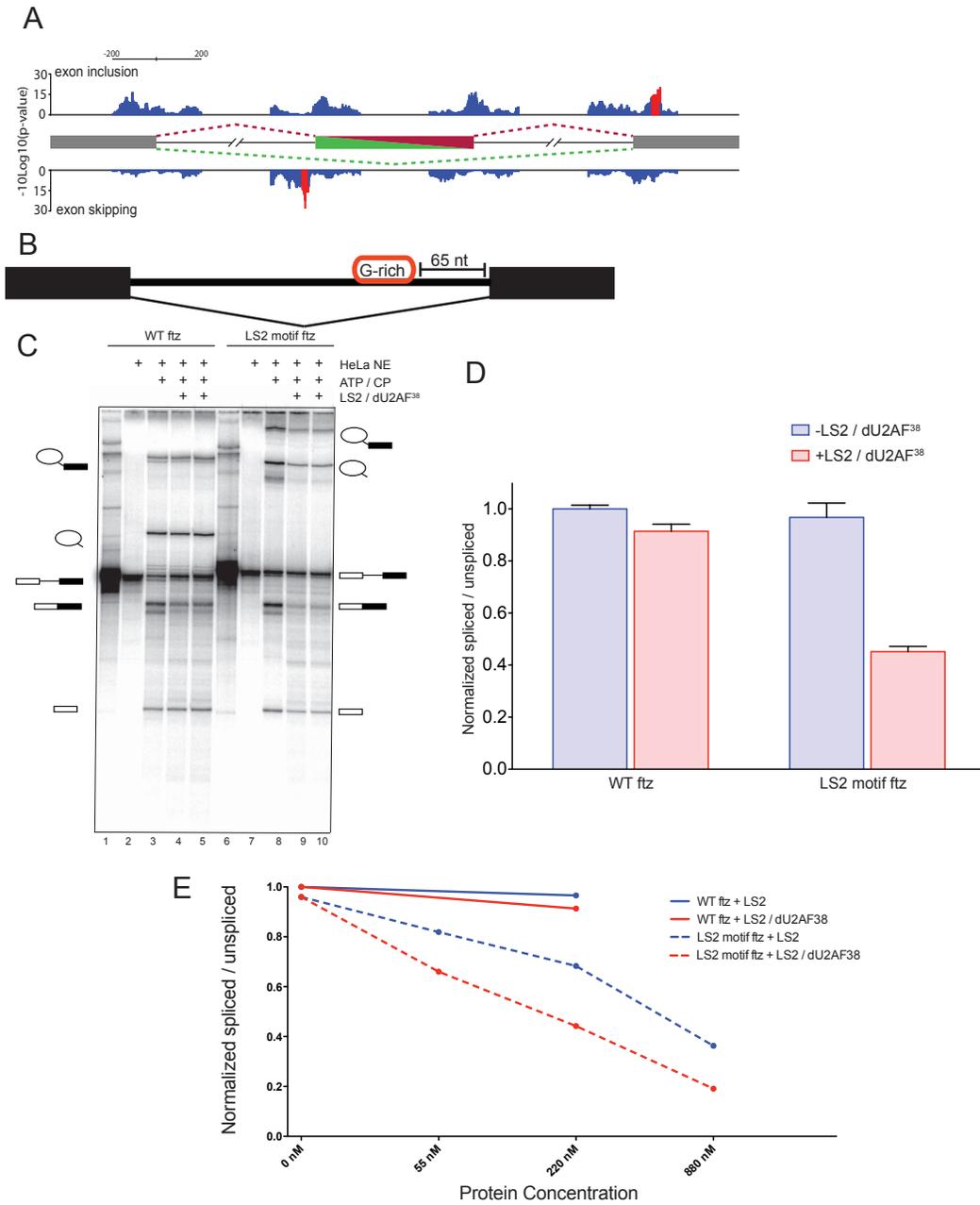


Figure 9. LS2 inhibits splicing of the ftz intron in vitro. **(A)** Motif location and clustering in LS2-affected cassette exons. LS2-affected cassette exons were searched for LS2 recognition motifs in 50 nt overlapping windows. Motif-containing windows were called as significant (red bars) if they contained a significant motif enrichment (p value < 0.05) and were part of a stretch of at least five consecutive significant windows. LS2 recognition motifs associated with cassette exon inclusion are displayed on top while motifs associated with cassette exon skipping are on bottom. **(B)** Diagram of the modified ftz intron in vitro splicing substrate generated with and without a G-rich LS2 recognition motif inserted 65 nucleotides upstream of the 3' splice site. **(C)** In vitro splicing reaction of the wild type (lanes 1-5) and LS2 SELEX motif-containing (lanes 6-10) ftz substrates using HeLa cell nuclear extract in the presence or absence of purified recombinant GST-tagged LS2/dU2AF³⁸ heterodimer. The identity of each RNA species is shown schematically to the right and left of the panel. Creatine phosphate is abbreviated as CP. Reactions were carried out without HeLa nuclear extract (lanes 1 and 6), with nuclear extract but without ATP and CP (creatine phosphate) (lanes 2 and 7), with nuclear extract, with ATP and CP (lanes 3 and 8) and with nuclear extract, with ATP and CP and with 500 ng (lanes 4, 5, 9, and 10) recombinant LS2/dU2AF³⁸ heterodimer protein. **(D)** Phosphorimager quantification of the results in panel B. The y-axis represents the ratio of all splicing intermediate species to the unspliced pre-mRNA, with intensities for each species normalized to their length and all ratios normalized such that the value for WT ftz without added LS2/dU2AF³⁸ heterodimer is 1.0. Error bars represent standard deviations of four to six experiments. **(E)** In vitro splicing efficiency of WT ftz and LS2 motif ftz in the presence of varying amounts of purified recombinant LS2 and LS2/dU2AF³⁸ heterodimer. Quantification was performed as in D.

Chapter 4: Evolution of a tissue-specific splicing network



Chapter 4: Evolution of a tissue-specific splicing network

Figure 10. In vitro splicing assay using LS2 alone. In vitro splicing assays were performed as in figure 5 with the exception of the addition of purified recombinant LS2 instead of purified recombinant LS2 / dU2AF³⁸ heterodimer. The amount of LS2 added to each reaction was the same molar amount of LS2 / dU2AF³⁸ added to each reaction in figure 5 (final concentration of 220 nM).

Chapter 4: Evolution of a tissue-specific splicing network

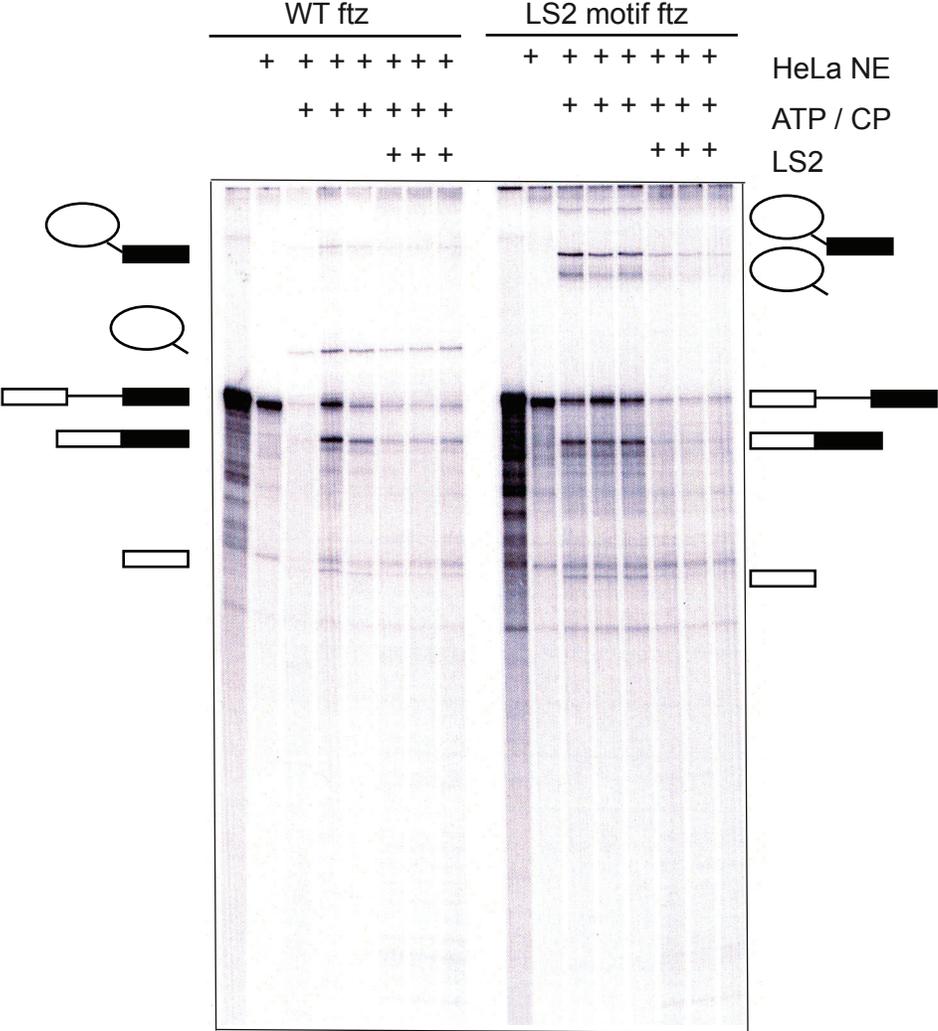


Figure 11. In vitro splicing assay using titrated amounts of LS2 and splicing of ftz G-replacement RNA. (A) In vitro splicing assays were performed as in figure 5. The RNA used is LS2 motif ftz (Figure 5B). Purified recombinant LS2 or LS2 / dU2AF³⁸ heterodimer was added to final concentrations of 55 nM, 220 nM, and 880 nM. Quantification was performed as in Figure 5D. **(B)** In vitro splicing of ftz G-replacement RNA. To test whether LS2 could activate splicing of an intron where the polypyrimidine tract had been replaced by LS2-binding motifs, we created such a construct using the ftz intron. Splicing of this construct was then monitored both in the presence and absence of purified recombinant LS2/dU2AF³⁸ heterodimer. Splicing of this intron at any appreciable level was never detected, either with or without LS2.

Chapter 4: Evolution of a tissue-specific splicing network

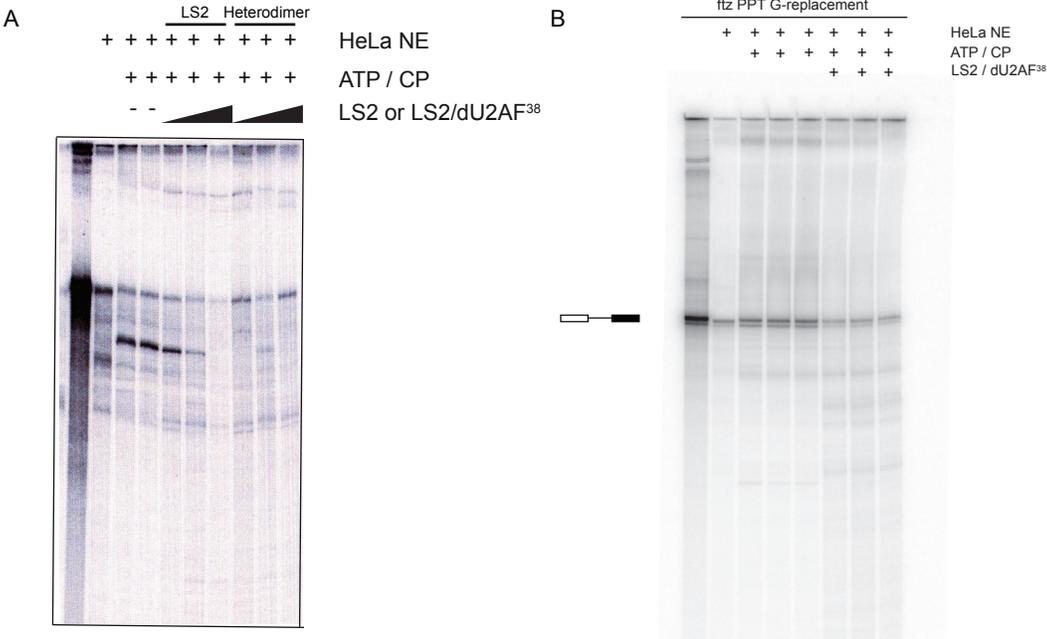


Figure 12. LS2 acts as a splicing repressor in vivo, is enriched at specific positions in its target transcripts and binds its predicted targets. (A) The minigene splicing reporter used with or without an LS2 G-rich binding motif inserted 60 nucleotides upstream of the cassette exon (white). **(B)** Effect on exon inclusion in S2 cells as measured by RT-PCR of RNA expression from the minigenes carrying the LS2 recognition motif and LS2 over-expression. Lane 1, splicing pattern without LS2 motif or LS2 overexpression; Lane 2, splicing pattern without LS2 motif but with LS2 overexpression, Lane 3, splicing pattern with LS2 motif, but without LS2 overexpression, Lane 4, splicing pattern with LS2 motif and LS2 overexpression. Lane 5, splicing pattern with neutral motif (see Supplementary methods) and without LS2 overexpression. Left side schematic, denotes inclusion (top) or exclusion (bottom) product. **(C)** Quantification of results in b. Error bars represent standard deviations from 3 independent biological replicates. **(D)** *Drosophila* S2 cells stably expressing epitope-tagged (Glu-Glu, also called Py) LS2 protein were lysed and LS2 was immunoprecipitated using either anti-Py antibodies (lane 5) or non-immune IgG (lane 4) and detected using anti-PSI antibodies (top panel) or anti-LS2 antibodies (bottom panel). Input protein is shown in lane 1 (5% of input). PSI protein was detected and used as a negative control for immunoprecipitation. Both immunoprecipitation pellets and flowthrough material for IgG (lanes 2 and 4) or anti-Py antibody (lanes 3 and 5) are shown. **(E)** Immunoprecipitation of LS2 nuclear RNP complexes followed by RT-PCR of predicted affected transcripts using equal amounts of immunoprecipitated or input RNA. These included two CaMKII isoforms, ferrochelatase, nonA, two RpL3 isoforms, and Sp7. cDNA amplification products specific for each gene was compared between input RNA, LS2-immunopurified and non-immune IgG immunopurified RNA samples.

Chapter 4: Evolution of a tissue-specific splicing network

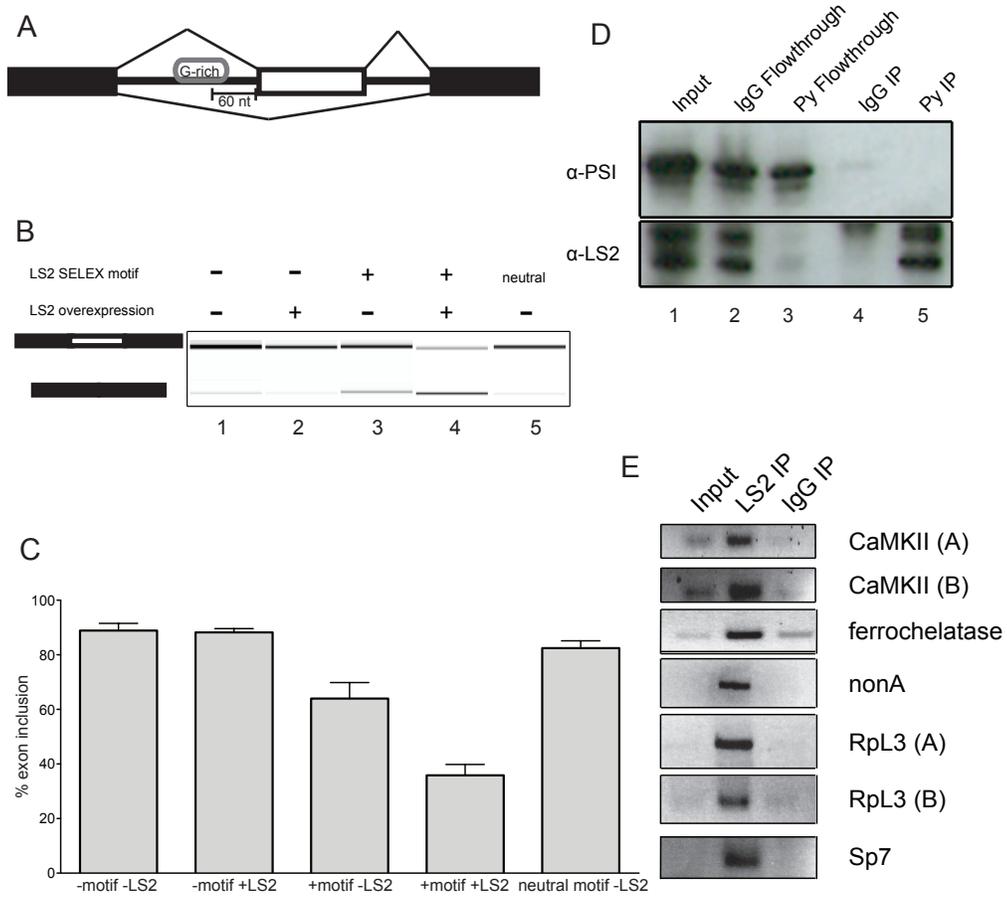
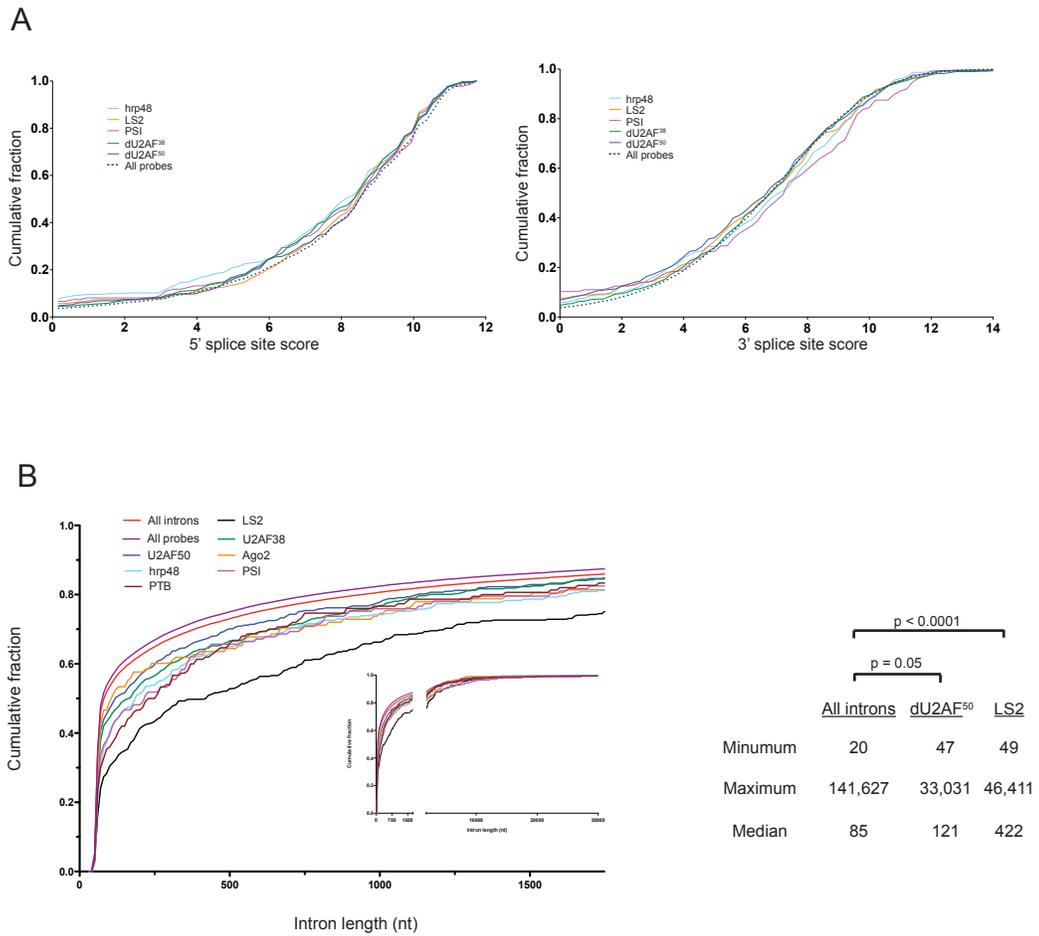


Figure 13. LS2 function is not constrained to splice sites and its affected introns are longer than expected. (A) Using MaxEntScore (Yeo and Burge, 2004), 5' or 3' splice site strengths of introns affected by hrp48 (Blanchette et al. 2009), PSI, dU2AF⁵⁰, dU2AF³⁸, and LS2 were calculated. As a baseline, the strengths of all splice sites interrogated by the arrays (all probes) were also also calculated. Each point represents the cumulative fraction of sites whose strength is the x-axis value or lower. There were no significant differences between the “LS2-affected” dataset and the “all probes” dataset. (B) Intron lengths of splice junctions affected by several splicing factors. Here, the “all introns” dataset represents all 57,124 introns annotated in the *D. melanogaster* genome by FlyBase. The “all probes” dataset again represents all introns interrogated by the arrays, approximately 82% of all introns. Each point represents the cumulative fraction of introns whose length is the x-axis value or lower. P values are calculated using a two-tailed nonparametric t test to correct for the non-Gaussian distribution of intron lengths.

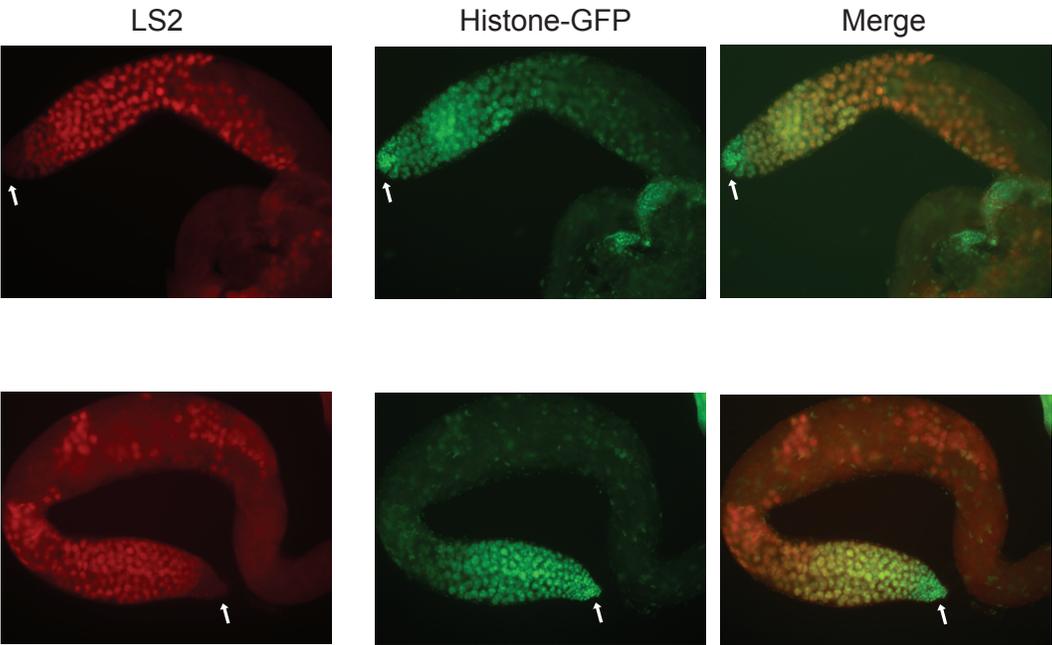
Chapter 4: Evolution of a tissue-specific splicing network



Chapter 4: Evolution of a tissue-specific splicing network

Figure 14. LS2 is expressed only in differentiated cells in the *Drosophila* testes. Immunofluorescence of 1-3 day old hand-dissected *Drosophila* testes. Flies were from a transgenic line expressing GFP-fused His2Av to visualize all cells. LS2 was visualized using an anti-LS2 primary and an Alexa Fluor 568 donkey anti-rabbit secondary (Invitrogen). Undifferentiated stem cells are located at the testis tip (white arrow).

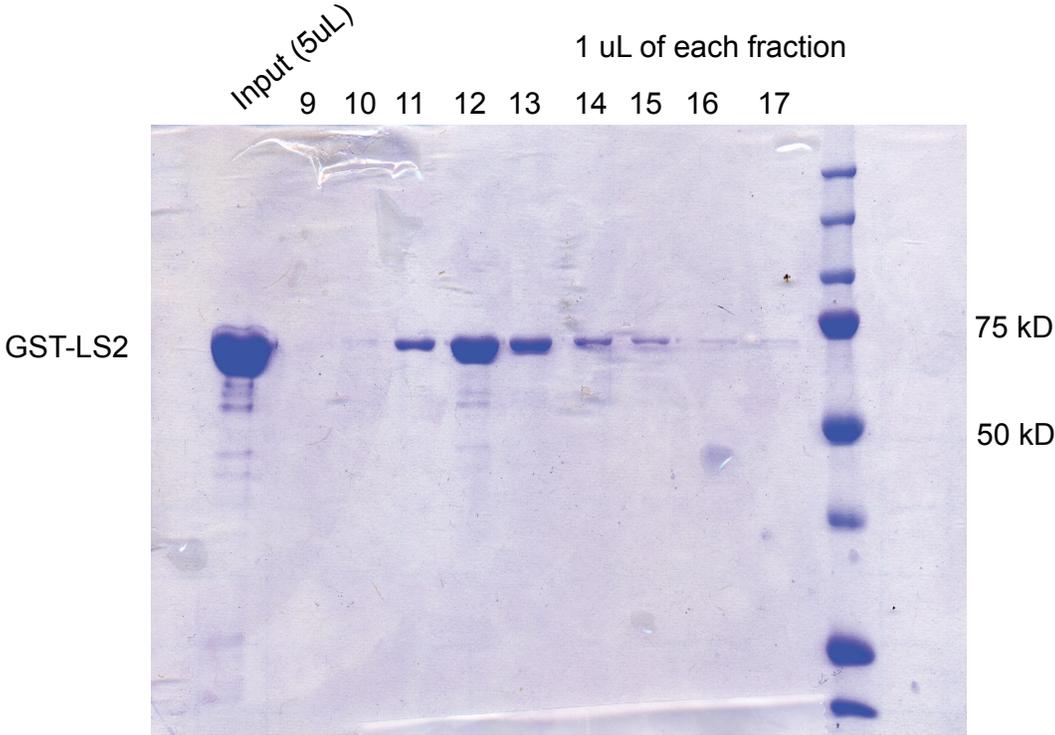
Chapter 4: Evolution of a tissue-specific splicing network



Chapter 4: Evolution of a tissue-specific splicing network

Figure 15. Recombinant LS2 purification using Poros HS. SDS-PAGE gel of Poros HS fractions. Recombinant GST-LS2 was purified first using glutathione agarose and then further purified using a 1 mL Poros HS column. Fractions were eluted over a KCl gradient from 0 to 1M. One microliter of each fraction was then run on an SDS-PAGE gel.

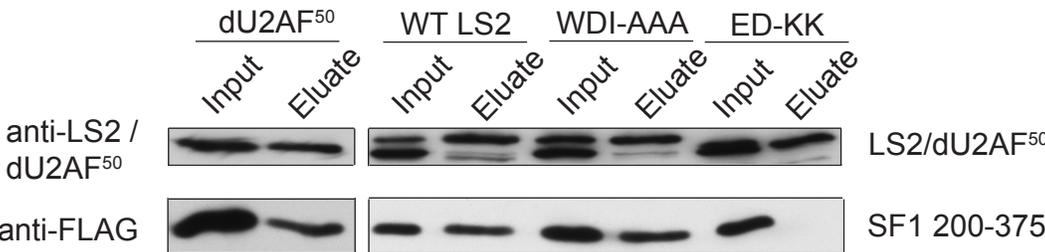
Chapter 4: Evolution of a tissue-specific splicing network



Chapter 4: Evolution of a tissue-specific splicing network

Figure 16. LS2 interacts with SF1. Western blot showing GST-tagged recombinant dU2AF50 and LS2 and FLAG-tagged recombinant SF1 200-375. WDI-AAA is a triple point mutant made in residues 62-64 of LS2 that is predicted to block interaction with dU2AF38. ED-KK is a double point mutant in residues 370 and 371 of LS2. The pulldown is visualized by immunoblotting for SF1 200-375, which is performed using an anti-FLAG antibody.

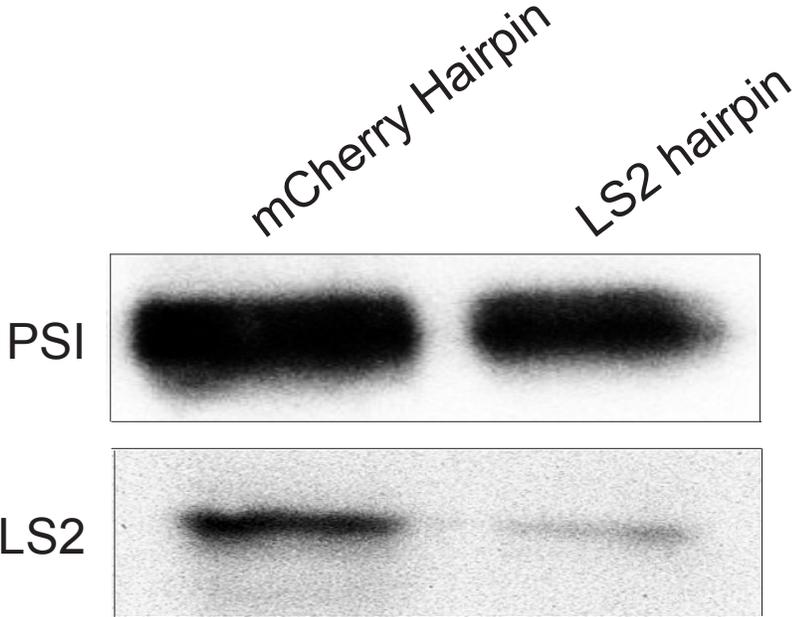
Chapter 4: Evolution of a tissue-specific splicing network



Chapter 4: Evolution of a tissue-specific splicing network

Figure 17. LS2 RNAi fly lines. Western blot showing LS2 levels in testes extracts in mCherry and LS2 RNAi lines. Both are driven by Act5C-Gal4. PSI is used as a loading control.

Chapter 4: Evolution of a tissue-specific splicing network



INTRODUCTION

This appendix is related to Chapter 2. Chapter 2 dealt with the phosphorylation of a splicing factor, PSI, that is a component of the splicing silencing complex that promotes retention of the third intron of the P element in *Drosophila* somatic cells. The experiments detailed here use site-specifically labeled RNA to determine the members of this complex and their relative positions. These experiments rely on the production of an RNA with a single radioactive phosphate placed at a known position. After the addition of purified proteins and/or extract, RNase protection patterns left by these proteins allow the determination of the location of bound proteins.

The production of site-specifically labeled RNA is laid out in detail in the experimental procedures section, but it is important to understand that this RNA is designed to contain only one radioactive phosphate, and that this phosphate is at a predetermined, known position (Figure 1A).

Using protection of site-specifically labeled RNA, we have mapped the binding sites of hrp48 and PSI to the P element RNA. Additionally, using Kc nuclear extract and a U1 antisense oligo, we have determined the contribution of U1 to this complex and its necessity for a complete occlusion of the true 5' splice site.

RESULTS

PSI and hrp48 protect the F1/F2 pseudo splice sites, but not the accurate 5' splice site

To examine in more detail how proteins and U1 snRNP interact with the P element silencer RNA, we used a site-specifically ³²P-labeled RNA in a nuclease protection assay (Maroney et al. 2000). Our initial studies were aimed at mapping the RNA binding sites for recombinant PSI and hrp48 proteins (Siebel et al. 1994; Siebel et al. 1995) to more accurately determine the binding sites on the P element IVS3 5' exon silencer for PSI and hrp48, we produced an RNA that was site-specifically labeled at the F1/F2 pseudo-5' splice sites, 24 nt upstream from the accurate 5' splice site (Figure 1B). With the ³²P-label at this location, both PSI and hrp48 were able to efficiently protect RNA regions containing the ³²P-label of 23 nt and 17 nt, respectively (Figure 1C). The lengths of these protected RNA fragments are also consistent with neither PSI nor hrp48 being able to prevent access to the accurate 5' splice site on their own, consistent with previous studies showing that U1 snRNP also binds this silencer element (Siebel et al. 1992; Labourier et al. 2002). The protected RNA fragment lengths of 23 nt and 17 nt must contain the label, which is 24 nt away from the accurate 5' splice site. If one assumes the label is at the extreme 5' end of the protection fragments, this puts the PSI and hrp48, at their closest, 1 nt and 7 nt away from the accurate 5' splice site, respectively. It is more likely, though, that the label is in the middle of the fragment, putting the proteins 7-10 nt farther away from the 5' splice site. The exact location could

Appendix A: Analysis of the P element splicing silencer complex by RNase protection

be determined by digesting the protected fragments with the original chimeric oligos used in the production of the site specifically labeled RNA with RNase H.

In accord with previous studies, when the label was placed at the accurate 5' splice site (Figure 1D), neither PSI nor hrp48 was able to confer any protection (Figure 1E). Thus, neither of these proteins, on their own, would likely be able to prevent the binding of U1 snRNP to the accurate 5' splice site, which is consistent with a larger RNP complex containing U1 snRNP along with PSI, hrp48 and other proteins bound to the P element splicing silencer.

To determine the contribution of U1 snRNP itself to the silencing complex, we repeated the protection experiment with the ³²P-label at the F1/F2 pseudo splice site and Kc nuclear extract (Figure 2A). Protection of the RNA with Kc nuclear extract resulted in a characteristic four RNA band protection pattern from 30 to 50 nt. This protection would likely be long enough to include the accurate 5' splice site. In order to establish the role that U1 snRNP was playing in this protection, we added a short 2'-O-methyl oligonucleotide that was antisense to *Drosophila* U1 snRNA to the reaction. This specific oligonucleotide was designed to base pair to the region at the 5' end of U1 snRNA that binds to 5' splice sites and block U1 snRNP binding. Addition of the 2'-O-methyl oligonucleotide caused the collapse of the four band pattern to a single, smaller band, implying that a smaller region of the RNA was now protected by proteins in the absence of U1 snRNP binding (Figure 2A, center 3 lanes). Increasing the amount of Kc nuclear extract in the reaction, and correspondingly more U1 snRNP, overcame the block due to the oligonucleotide and restored the characteristic four RNA band pattern (Figure 2A, right 3 lanes).

When the ³²P-label was placed at the accurate 5' splice site, the addition of Kc nuclear extract resulted in the protection of large (> 65 nt) RNA fragments (Figure 2B). Interestingly, these fragments were completely dependent upon the binding functionality of U1 snRNP, as addition of the U1 antisense oligonucleotides abolished their presence. This extended protection patterns is consistent with U1 snRNP binding to the F1/F2 sites and protecting an extended region including the accurate 5' splice site. Such extended protection of U1 snRNP binding to splicing silencers has been observed in pre-mRNAs containing in vitro-selected human silencer elements, some of which interact with the human homolog of hrp48, hnRNP A1. Taken together, these data are consistent with a U1 snRNP-dependent extended protection including the F1/F2 sites and the accurate 5' splice site. Additional studies using biochemically purified U1 snRNP, recombinant PSI, hrp48, hrp36, hrp38 and PABPC1 with the site-specific labeling nuclease protection assays should clarify the role of U1 snRNP and other RNA binding proteins in the assembly and organization of the P element splicing silencer complex.

DISCUSSION

Appendix A: Analysis of the P element splicing silencer complex by RNase protection

In these studies we have looked at the mechanism behind splicing silencing at the third intron of the P element transcript in *Drosophila*. From the RNase protection experiments using site-specifically labeled RNA (Figure 1), it was clear that while both PSI and hrp48 could protect regions around the F1/F2 silencer, neither could protect the accurate 5' splice site on their own. This implies that another factor is necessary to protect the region around the accurate 5' splice site and prevent U1 snRNP from binding.

RNase protection assays using Kc nuclear extract (Figure 2) demonstrated that this extra factor may be U1 snRNP itself. The addition of Kc nuclear extract resulted in the protection of a much larger region, up to 55 nt, around the F1/F2 silencer. This would likely be large enough to reach the accurate 5' splice site. U1 snRNP, however, was an integral part of this complex, as addition of an antisense DNA oligo targets against U1 snRNA reduced the length of the protected fragment to 33 nt, likely not large enough to reach the accurate 5' splice site. This inhibition of U1 snRNP binding could be overcome by the addition of more extract, and thus more U1 snRNP.

Further supporting the idea of U1 snRNP binding to the F1/F2 silencer being the critical step in blocking access to the accurate 5' splice site is the fact that while Kc nuclear extract efficiently protected large regions around the accurate 5' splice site (Figure 2B), this protection was completely dependent upon the ability of U1 snRNP to bind RNA as addition of the U1 antisense oligo abolished that protection. PSI and hrp48 are still able to bind the F1/F2 region under these conditions, but consistent with figure 1D, they are not able to protect the accurate 5' splice site on their own.

All of this is consistent with previous reports that showed that the A/B region of PSI is necessary for its interaction with PSI (Labourier et al. 2001). Furthermore, deletion of the A/B region releases some of the inhibition of splicing at the P element third intron (Labourier et al. 2001). This leads to a model, presented in figure 3, in which PSI and hrp48 bind the F1/F2 silencer, but are unable on their own to prevent binding of U1 snRNP to the accurate 5' splice site. Only upon recruitment of U1 snRNP to the F1/F2 site through an interaction with the A/B region of PSI is the resulting complex then big enough to occlude binding of another molecule of U1 snRNP to the accurate 5' splice site. The binding of U1 snRNP to the F1/F2 site requires stabilization through both interaction with PSI and its normal snRNA-mediated mode of RNA binding as the disruption of either (through the use of PSI Δ AB or the U1 antisense oligo) results in the loss of U1 recruitment to the F1/F2 site.

EXPERIMENTAL PROCEDURES

Purification of PSI

Appendix A: Analysis of the P element splicing silencer complex by RNase protection

PSI in pRSETa was grown in BL21 pLYS E cells to OD 0.4 at 37C, cooled to 16C, and induced with 0.5 mM IPTG overnight. The bacteria were spun down and resuspended in lysis buffer (20 mM HEPES pH 7.6, 10% glycerol, 1 M NaCL, 0.05% Tween, 0.5 mM DTT, 0.4 mM PMSF) and snap frozen in liquid nitrogen.

The lysate was then thawed, sonicated 4 times for 30 seconds each, and spun at 15000 RPM for 30 min to separate insoluble material. The soluble lysate was filtered and purified over a NiHiTrap 1 mL column using lysis buffer as the “A” buffer, and lysis buffer containing 500 mM imidazole as the “B” buffer. The His-tagged PSI was eluted from the column over 15 column volumes in 0.5 mL fractions. See figure 4A for an SDS-PAGE gel of the purification.

Purification of hrp48

Hrp48 was purified in the same manner as PSI. Importantly, the bacterial cells were cooled to 16C before induction to promote the solubility of hrp48. See figure 4B for an SDS-PAGE gel of the purification.

Production of site-specifically labeled RNA

This is a complex and rarely used protocol, so I have written it here in more detail than is normal for a section of this type.

I. Large scale RNA transcription

Mix the following in an eppendorf tube: 139 μ L water, 50 μ L 5X HYB buffer (1M HEPES-KOH pH 7.6, 100 mM MgCl₂, 200 mM DTT, 10 mM Spermidine-HCl, 0.28 mg/mL BSA), 50 μ L 5X NTP (25 mM each), 2.5 μ L RNase inhibitor (40 U/ μ L), 2.5 μ L inorganic phosphatase (1U / μ L), 2.5 μ L α^{32} P-UTP (800 Ci / mmol), 2.5 μ L DNA template (1 μ g/ μ L), 1.25 μ L T7 RNA polymerase (homemade stock). Incubate 90 min at 37C. Add 1.25 μ L T7 RNA polymerase. Continue incubation for another 90 min at 37C. Save 2.5 μ L of the reaction as input for RNA quantification. Add 5 μ L RQ1 DNase. Incubate 37C for 30 min. Purify RNA using Qiagen RNAeasy kit. Elute in 100 μ L. Quantify RNA using Cerenkov counting. Expect 1 to 2 nmol of RNA.

II. 2'-O-methyl RNA/DNA oligonucleotide directed RNase H cleavage

The chimeric oligo is complementary to the target RNA with 8-10 nt long of 2'-O-methyl DNA followed by 4 DNA nucleotides and 3 2'-O-methyl DNA nucleotides. Cleavage occurs right before the 5' most DNA nucleotide. Blue = 2'-O-methyl Red = DNA Green = your RNA

```
5'XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 3'  
  3'XXXXXXXXXXXXXXXXX 5'
```

Appendix A: Analysis of the P element splicing silencer complex by RNase protection

|
Cleavage site

Mix the following in an eppendorf tube to a final volume of 20 μL : 200 pmol RNA, 2.5 μL 10 X RNase H buffer A (200 mM Tris-HCl pH 7.5, 1M KCl, 100 mM MgCl_2), 1 nmol 2'-O-methyl RNA/DNA oligo. Heat at 95C for 5 min and cool down slowly over the course of 30 min to room temperature. Add 2.5 μL 10 X RNase H buffer B (10 mM DTT, 0.5 mg/mL acetylated BSA, 16 U/ μL RNase inhibitor) and 2.5 μL RNase H (USB). Incubate for 2 hrs at 30 C. Spike in another 1.25 μL RNase H. Continue incubation for another 2 hrs.

Add 30 μL of RNA sample buffer and run on a 15% acrylamide denaturing gel for 3 hours at 300 volts. Expose to film to identify cleavage products. Cut products out of gel and elute in 0.3M NaOAc, 0.2% SDS overnight at 4 C. Phenol chloroform extract and ethanol precipitate. Resuspend the 5' fragment in 10 μL water and store at -20 C.

III. Dephosphorylation and kinasing of 3' half

Resuspend the 3' fragment in 17 μL water. Add 2 μL 10X CIP buffer and 1 μL CIP (Boehringer). Incubate 1 hr at 30 C. Phenol/chloroform extract and ethanol precipitate. Resuspend in 12.5 μL water, 2 μL Optikinase buffer (USB), 2.5 μL $\gamma^{32}\text{P}$ ATP (7000 Ci / mmol) and 2 μL Optikinase (USB). Incubate 1 hr at 37 C. Remove unincorporated ATP with 2 consecutive G10 columns. Phenol/chloroform extract and ethanol precipitate. Resuspend in 5 μL water.

IV. Ligation of the 5' and 3' fragment

The splint DNA oligo used for the ligation covered between 15 and 17 nt on each side of the cleavage site.

Mix the following: 5 μL kinased 3' fragment, 10 μL cold 5' fragment, 1.5 μL 10X ligase buffer (NEB) 2.5 μL splint oligo (20 pmol/ μL). Heat at 75 C for 2 min and then incubate at room temperature for 5 minutes. Add 1 μL RNase inhibitor (40 U/ μL) and 1.5 μL T4 DNA ligase (NEB). Incubate 16 hrs at 16 C. Run as before on a 15% polyacrylamide denaturing gel, elute and resuspend in 10 μL water.

Nuclease Protection Assays

The following were mixed in an eppendorf tube to a final volume of 15 μL : 6 μL 2.5X Binding Buffer (12.5 mM creatine phosphate, 10 mM MgCl_2 , 50 mM HEPES-KOH pH 7.6, 1.25 mM DTT, 1U / μL RNase inhibitor, 200 mM potassium glutamate, 6.25% polyvinyl alcohol), 5 μL Kc nuclear extract OR 150 ng purified protein AND/OR 600 ng U1 antisense oligo, 4 μL site-specifically labeled RNA (~20000 cpm).

Appendix A: Analysis of the P element splicing silencer complex by RNase protection

The reaction was incubated at room temperature for 2 hours to allow complexes to assemble and then diluted to 200 μ L with dilution buffer (5 mM creatine phosphate, 4 mM MgCl₂, 20 mM HEPES-KOH pH 7.6, 0.5 mM DTT, 100 mM potassium glutamate, 2 mM CaCl₂). 300 U of micrococcal nuclease (USB) was added, and the reaction was incubated 10 minutes on ice. EDTA was then added to 4 mM, and the reaction was phenol/chloroform extracted and ethanol precipitated. The recovered RNA was run on a 15% denaturing polyacrylamide gel for 3 hr at 300 V. The gel was fixed in 10% methanol, 10% acetic acid, dried, and exposed to a phosphorimager screen overnight.

REFERENCES

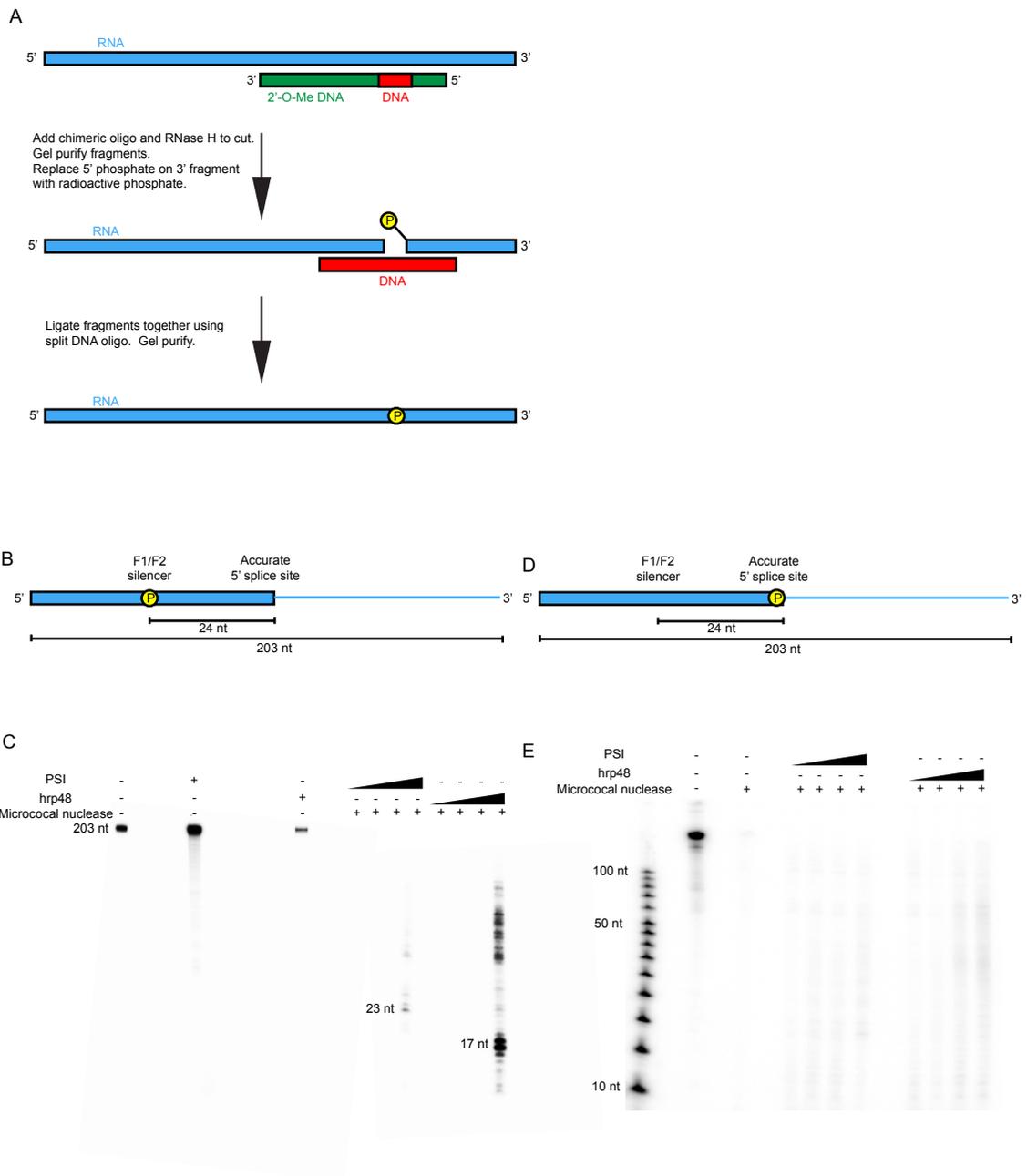
- Labourier E, Adams MD, Rio DC. 2001. Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. *Mol Cell* **8**: 363-373.
- Labourier E, Blanchette M, Feiger JW, Adams MD, Rio DC. 2002. The KH-type RNA-binding protein PSI is required for Drosophila viability, male fertility, and cellular mRNA processing. *Genes Dev* **16**: 72-84.
- Maroney PA, Romfo CM, Nilsen TW. 2000. Nuclease protection of RNAs containing site-specific labels: a rapid method for mapping RNA-protein interactions. *RNA* **6**: 1905-1909.
- Siebel CW, Admon A, Rio DC. 1995. Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes Dev* **9**: 269-283.
- Siebel CW, Fresco LD, Rio DC. 1992. The mechanism of somatic inhibition of Drosophila P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. *Genes Dev* **6**: 1386-1401.
- Siebel CW, Kanaar R, Rio DC. 1994. Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. *Genes Dev* **8**: 1713-1725.

Appendix A: Analysis of the P element splicing silencer complex by RNase protection

Figure 1. Protection of site-specifically labeled RNA with purified PSI and hrp48.

A) Production scheme of site-specifically labeled RNA. RNA is cut at particular sequence through the use of a chimeric DNA/2-O-Methyl oligo and RNase H. The 5' phosphate on the 3' fragment is replaced with a radioactive phosphate, and the two pieces are ligated back together. For a detailed description, see materials and methods. B) Schematic of RNA labeled site-specifically at the F1/F2 silencer. C) Protection patterns of F1/F2 labeled RNA produced by purified recombinant PSI and hrp48. D) Schematic of RNA labeled site-specifically at the accurate 5' splice site. E) Protection patterns of accurate 5' splice site labeled RNA produced by recombinant PSI and hrp48.

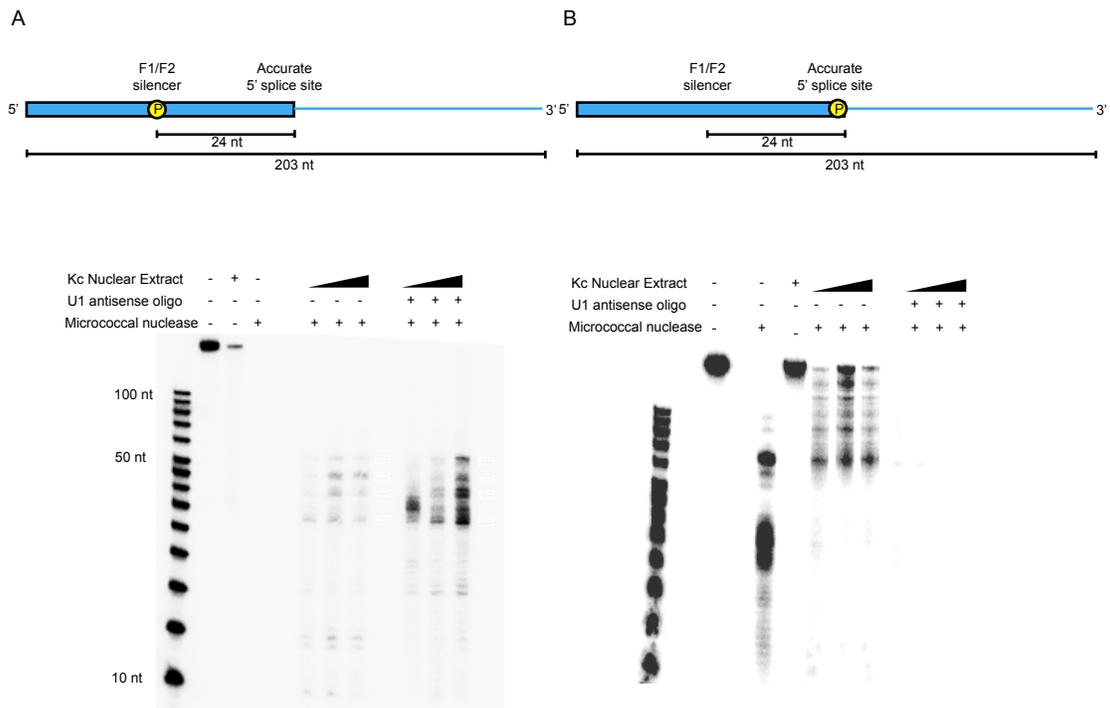
Appendix A: Analysis of the P element splicing silencer complex by RNase protection



Appendix A: Analysis of the P element splicing silencer complex by RNase protection

Figure 2. Protection of site-specifically labeled RNA with Kc nuclear extract. A) Schematic of RNA labeled site-specifically at the F1/F2 silencer. B) Protection patterns of F1/F2 labeled RNA produced by Kc nuclear extract with and without the U1 antisense oligo. D) Schematic of RNA labeled site-specifically at the accurate 5' splice site. E) Protection patterns of accurate 5' splice site labeled RNA produced by Kc nuclear extract with and without the U1 antisense oligo

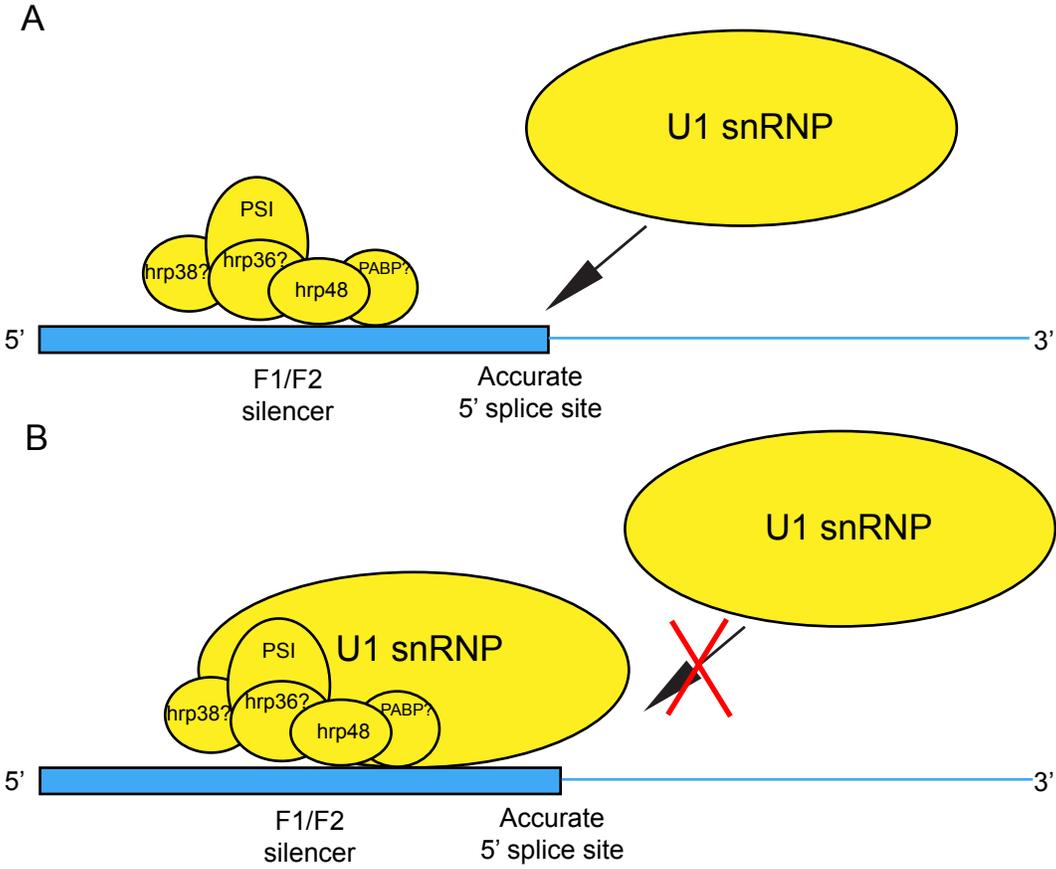
Appendix A: Analysis of the P element splicing silencer complex by RNase protection



Appendix A: Analysis of the P element splicing silencer complex by RNase protection

Figure 3. A model for the repression of splicing at the P element third intron. A) Several proteins, including PSI and hrp48, as well as possibly hrp36, poly-A binding protein, and hrp38, bind to the F1/F2 silencer. However, without recruiting U1 snRNP to the complex, they are unable to prevent the binding of U1 snRNP to the accurate 5' splice site. This would represent the case in the protection assays using purified hrp48 and PSI, as well as when PSI Δ AB is used in place of wildtype PSI. B) If the silencing complex is able to recruit U1 snRNP, the resulting complex is then large enough to occlude the binding of a second molecule of U1 snRNP to the accurate 5' splice site. In this way, splicing at the accurate 5' splice site is inhibited. Importantly, U1 snRNP likely uses its snRNA to contact the F1/F2 silencer since it is similar in sequence to a 5' splice site, and addition of the U1 antisense oligo seems to abolish its ability to bind the F1/F2 site.

Appendix A: Analysis of the P element splicing silencer complex by RNase protection

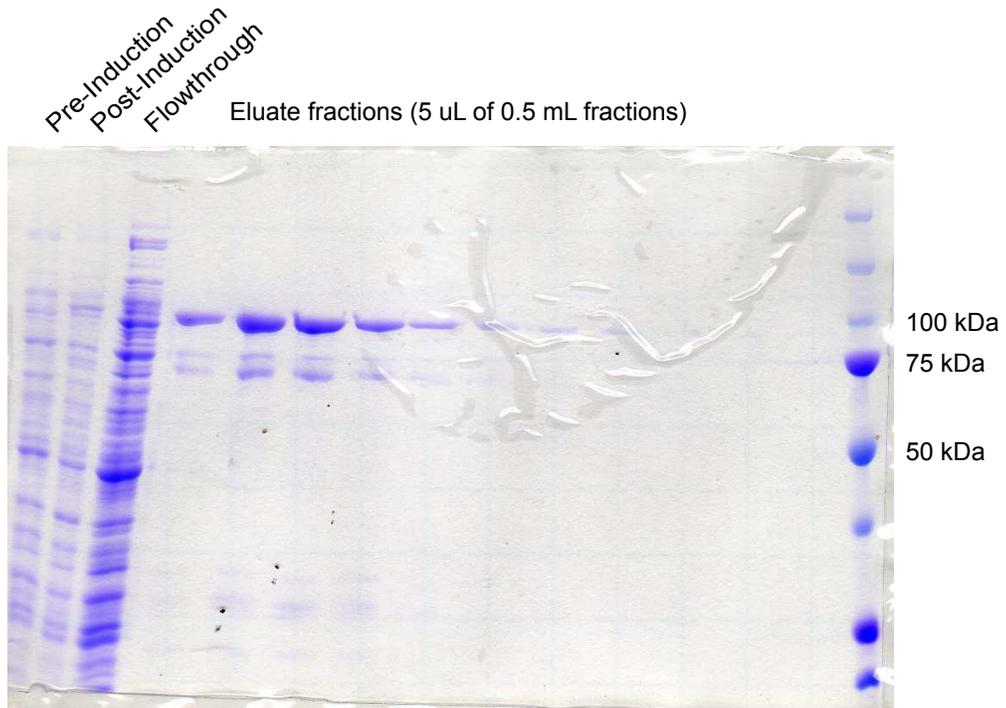


Appendix A: Analysis of the P element splicing silencer complex by RNase protection

Figure 4. Purification of recombinant PSI and hrp48. A) PSI purification. Coomassie-stained SDS-PAGE gel of fractions from a 1 mL Ni-HiTrap column. Each lane represents 5 μ L of 500 μ L fractions. PSI migrates at 97 kDa. B) Hrp48 purification. Coomassie-stained SDS-PAGE gel of fractions from a 1 mL Ni-HiTrap column. Each lane represents 5 μ L of 500 μ L fractions. Hrp48 migrates at 48 kDa.

Appendix A: Analysis of the P element splicing silencer complex by RNase protection

A



B

