UNIVERSITY OF CALIFORNIA,
IRVINE

Molecular Mechanisms and Conservation of Pre-mRNA Splicing

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Sciences

by

Maliheh Movassat

Dissertation Committee:
Professor Klemens J. Hertel, Ph.D., Chair
Professor Bert L. Semler, Ph.D.
Professor Yongsheng Shi, Ph.D.

2018

DEDICATION

To my husband, Amir E. Golji

For everything.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Many thanks go to my PI and mentor Dr. Klemens (Dicki) Hertel, for giving me a lab-home when I had none and for the countless opportunities he placed my way. His guidance and support gave me the courage to succeed and for that, I am eternally grateful.

I would like to thank the members of my committee, Dr. Bert L. Semler and Dr. Yongsheng Shi. I am truly grateful for their scientific guidance, helpful suggestions and advice throughout the course of my graduate career.

I would also like to thank the many members of the Hertel lab, past and present that made this journey possible. Will, thank you for training me when I first joined the lab and for your constant positivity and enthusiasm towards life. Anke, your camaraderie and "mad" bioinformatic skills are something I will never forget. Angela, your friendship, persistence that I learn R, and our countless conversations about all things under the sun kept my spirits up and helped me get through everything grad school threw our way. I will forever miss sitting behind you. Wendy, your humor, our great many talks, and your unwavering friendship have been so encouraging and supportive. Francisco and Hossein, the energy, encouragement and laughs you brought to the lab were indispensable. Elmira, thank you for all your hard work and enthusiasm to help move projects forward.

This manuscript and all the work therein would not be possible without funding from the National Institutes of Health, and in addition, the use of my previously published material which is included in this document with the permission of *RNA Biology, Methods in Molecular Biology - Springer* and *Molecular Cell*.

To my parents, Masoud and Johanna, thank you for all that you did to get me to where I am today. You were my examples of what hard work can achieve. Your sacrifices, support, and continual confidence gave me strength to achieve my goals. To my brother, Meisam, thank you for always knowing how to make me laugh and for being my biggest fan.

Thank you to my amazing husband, Amir, from the bottom of my heart. You were my strength and greatest advocate. Thank you for your shoulder when I needed it and for your patience and love through it all. Most importantly, to the One who made it all be, thank you.

CURRICULUM VITAE

Maliheh Movassat

EDUCATION

University of California, Irvine, CA 2018
Doctor of Philosophy, Biomedical Sciences

San Jose State University, San Jose, CA 2006
Master of Biotechnology (MBT)

Santa Clara University, Santa Clara, CA 2004
Bachelor of Science, Biology; Minor, Art History

RESEARCH/PROFESSIONAL EXPERIENCE

Ph.D. Research, Dr. Klemens J. Hertel, PhD 2013 - 2018
Department of Microbiology and Molecular Genetics
University of California, Irvine

Graduate Student Researcher, Dr. John P. Fruehauf, MD, PhD 2010 - 2012
Chao Family Comprehensive Cancer Center
Department of Biological Chemistry
University of California, Irvine

Research Associate II 2006 - 2008
Microfluidics and Assay Development
Caliper Life Sciences, Mountain View, CA

Research Associate I 2005 - 2006
Protein Engineering Department
Rinat Neuroscience (Pfizer), South San Francisco, CA

Research Assistant 2005
Conservation Genetics Laboratory
Department of Biological Sciences, San Jose State University, San Jose, CA

TEACHING EXPERIENCE

Adjunct Faculty 2017
Santa Ana Community College
Department of Biology

Bio Bootcamp Tutor 2013 - 2014
University of California, Irvine

Teaching Assistant 2011 - 2012
University of California, Irvine

SELECT ABSTRACTS/PRESENTATIONS

1. **M. Movassat.** *Exon Size and Sequence Conservation Improves Identification of Splice Altering Nucleotides.* Chicago, IL, USA: ISMB 2018. **Poster**
2. **M. Movassat**. *Gene Regulation by Alternative Splicing.* Guest lecturer, ‘Brews and Brains’, Irvine, CA, 2016. **Talk.**
3. **M. Movassat**, J. Flesher, K. J. Hertel. *Splicing repression results in changes to U1 snRNP complex integrity at the 5’ splice site*. Cold Spring Harbor, NY, USA: Eukaryotic mRNA Processing, Cold Spring Harbor Laboratory Meeting 2015. **Talk.**
4. **M. Movassat**, T. L. Crabb, A. Busch, C. Yao, Y. Shi, K. J. Hertel. *Coupling between alternative polyadenylation and alternative splicing is limited to terminal introns.* Quebec, Canada: RNA Society Meeting 2014. **Poster.**
5. **M. Movassat**, T. L. Crabb, A. Busch, C. Yao, Y. Shi, K. J. Hertel. *Coupling between alternative polyadenylation and alternative splicing is limited to terminal introns.* San Diego, CA, USA: American Society for Biochemistry and Molecular Biology Meeting 2014. **Talk and poster.**

PUBLICATIONS

1. **M. Movassat**, H. Shenasa, K. J. Hertel. Preparation of Splicing Competent Nuclear Extract from Mammalian Cells and In Vitro Pre-mRNA Splicing Assay. Methods Mol Biol. 2017; 1648:11-26.
2. B. E. Aubol, G. Wu, M. M. Keshwani, **M. Movassat**, L. Fattet, K. J. Hertel, X. D. Fu, and J. A. Adams. Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. Mol Cell. 2016; 63(2):218-28.
3. **M. Movassat**, T. L. Crabb, A. Busch, C. Yao, D. Reynolds, Y. Shi, K. J. Hertel. Coupling between alternative polyadenylation and alternative splicing is limited to terminal introns. RNA Biol. 2016; 13(7):646-55.
4. **M. Movassat**, W. F. Mueller, K. J. Hertel. *In vitro* assay of pre-mRNA splicing in mammalian nuclear extract. Methods Mol Biol. 2014; 1126:151-60.
5. **M. Movassa**t. A Blanket of Peace. Issues in Ethics, Markkula Center for Applied Ethics, Winter 2003. Santa Clara Magazine, Spring 2005.

ABSTRACT OF THE DISSERTATION

Molecular Mechanisms and Conservation of Pre-mRNA Splicing

By

Maliheh Movassat

Doctor of Philosophy in Biomedical Sciences

University of California, Irvine, 2018

Professor Klemens J. Hertel, Ph.D., Chair

The eukaryotic genome is a large and complex network of molecules that work together to create diversity across many different species. Pre-mRNA splicing is a vital step in the processing of RNA into functionally diverse proteins. Splicing has a vast history and understanding the combination of many different factors and elements is fundamental to studying gene expression.

Gene evolution and the evolutionary pressures that encode a set of exonic sequences maintain efficient pre-mRNA splicing and ultimately dictate the selection of amino acids that define a protein. Exonic sequences have been regulated in a way to demonstrate the co-existence of coding and splicing pressures. Through the design of an exon conservation database, evolutionary conservation patterns were identified that influenced the final sequence of an exon. This information led to important predictions about splicing patterns in human disease. The database allowed for the identification of essential architectural parameters of the human genome. In addition, analysis of nucleotide variations at the wobble position identified splice altering SNPs and how these SNPs influenced exon inclusion.

Regulation of alternatively spliced exons requires a coordinated effort by many cis and trans-acting factors. Understanding how these factors work together is important for the mechanism of alternative splicing regulation. SR proteins and hnRNPs have previously demonstrated a position-dependent method of regulation. It was shown that U1 snRNP, at activating or repressive conditions, displayed dynamic changes in its compositional integrity. This demonstrated that U1 snRNP integrity is therefore modulated by the presence of position-dependent interactions with splicing regulatory factors, further suggestive of U1 snRNP as a molecular gatekeeper for splicing initiation.

Polyadenylation is a fundamental step in the 3' end processing of mRNA. Alternative polyadenylation is another contributor to genomic diversity. The coordinated efforts between splicing and polyadenylation have been demonstrated for terminal exons, however, understanding the influence these two processes have on upstream exons was unknown. Genome-wide analyses allowed for the identification of an important role that a polyadenylation factor (CstF64) has on alternative splicing. In addition, the coupling that was seen between alternative polyadenylation and alternative splicing was in fact limited to terminal exons.

CHAPTER 1

Introduction

Eukaryotic gene expression and proteomic diversity is dependent on the correct removal of introns and joining of exons in a process orchestrated by the spliceosome called splicing. The mechanisms that dictate splicing and their regulation are fundamentally important in the study of gene expression. This chapter will focus on examining the basic principles that control splice site recognition, the interplay between a myriad of factors and sequence elements, the coupling between distinct stages of gene expression, and the contribution of evolution in guiding the generation of differentially spliced mRNA isoforms.

Pre-mRNA Splicing

The splicing of pre-mRNA involves the recognition of distinct sequence elements located within the exon and intron and the stepwise assembly of components of the major spliceosome [1]. The spliceosome is composed of five uridine rich small nuclear ribonuclear proteins (UsnRNPs): U1, U2, U4, U5 and U6 snRNPs, as well as other non-snRNP protein factors. Each of these snRNPs is comprised of a small stable RNA bound by protein components and other less stably associated splicing factors, for a total of over 300 spliceosome associated proteins [2–4].

The first step in spliceosomal assembly is the recognition of the 5' and 3' splice sites (ss). The 5'ss or the splice donor site is defined by a nine nucleotide (nt) consensus sequence, YAG/GURAGU (where Y is a pyrimidine, R is a purine, and "/" denotes the exon/intron boundary and actual splice site) to which U1snRNP makes direct base pairing interactions. The 3'ss, also known as the splice acceptor site, is located at the intron/exon junction. It is defined by three sequence elements located upstream of the intron-exon junction. These loosely defined sequence elements include the 3'ss YAG/N sequence, the branchpoint sequence (BPS) with a consensus sequence of YNYURAY (where Y is C or U nucleotide), and the polypyrimidine tract (PPT) which varies in length and is characterized by a great number of pyrimidines (C or U nucleotides) [5] (Figure 1.1). U2AF35 interacts loosely with the 3'ss and U2AF65 binds the PPT. The interaction of these components and the binding of U1snRNP to the 5'ss initiates the formation of E complex or the early spliceosomal complex in an ATP independent manner. Subsequent ATP dependent steps lead to the stabilization of the E complex and the binding of the SF1 protein component of U2 snRNP to the branchpoint sequence, generating the A complex. Recruitment and rearrangement of the U4/U5•U6 tri-snRNP generates the pre-catalytic B complex. Structural rearrangements lead to the dissociation of U1 and U4, giving rise to the activated B complex. Further rearrangements lead to the C complex formation, through base pairing of U6 with U2 and the 5'ss. This permits the first transesterification reaction to occur, whereby the 5'ss phosphate is attacked by the 2'OH of the branchpoint adenosine, resulting in ligation of the 5' end of the intron to the branchpoint adenosine. Subsequent ATP dependent rearrangement then generates the second catalytic step whereby the 3'OH of the 5' exon attaches to the 3'ss leading to the completion of exon ligation and excision of the lariat structure creating a final spliced product [6] (Figure 1.2).

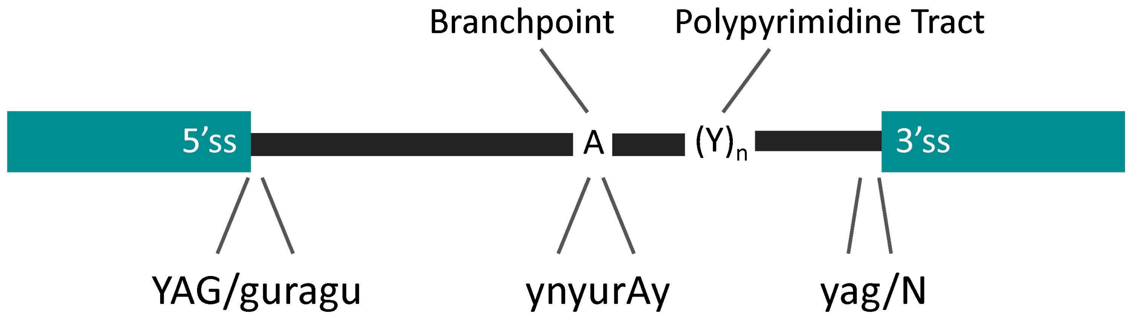


Figure 1.1. Splicing Recognition Sequence Elements.
Schematic of exon/intron (5'ss) and intron/exon (3'ss) junction and the associated necessary sequence elements for spliceosomal recognition. Y refers to a pyrimidine (C or U nucleotide), R refers to a purine (A or G nucleotide), N refers to any nucleotide, and "/" denotes a junction. Introns are represented as thick black lines and exons as boxes.

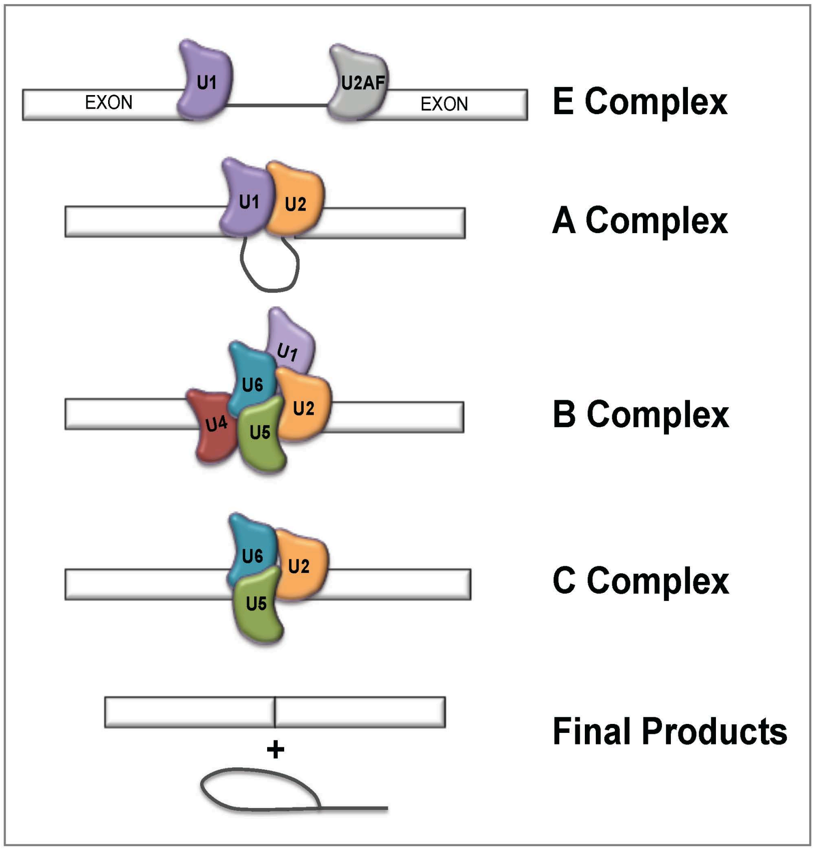


Figure 1.2. Model of Spliceosome Complex Formation.
Splicing occurs through recruitment of many splicing factors including U1, U2, U4, U5 and U6 snRNPs assembling in a stepwise fashion on the pre-mRNA. Splicing complex formation occurs in the order of E -> A -> B -> C to generate a final spliced product. Lariat formation is a by-product of this reaction.

Alternative splicing (AS) is a process that generates multiple isoforms from a single pre-mRNA transcript. AS results from the spliceosome's usage of different combinations of splice sites to impart an even greater influence on gene expression. There are roughly 21,000 human protein coding genes [7], a number that fluctuates as advances in understanding the genome are made. However, over 90,000 unique proteins have been documented. Given the vast number of proteins and the limited number of genes, proteomic diversity within the human genome has been predicted to be achieved through the mechanism of AS. It is known that ~95% of human genes undergo AS [8, 9]. AS can be grouped into multiple categories: alternative 5' splice site selection (5'ss), alternative 3' splice site selection (3'ss), cassette or skipped exons (SE), retained introns (RI), and mutually exclusive exons (MXE). Certain estimates have placed alternative 5'ss and 3'ss at roughly 25% of all AS events with cassette exons making up the largest proportion at about 50-60% [10, 11]. Exons that are alternatively spliced generally contain splice site sequences that vary significantly from the consensus sequence, which is suggestive of a lower affinity for the spliceosome [12–15]. Furthermore, the prevalence of pseudo splice sites within exons makes reliable distinction of canonical splice sites challenging without the aid of additional signals such as those from regulatory elements [16, 17]. Therefore, the presence of other regulatory factors that contribute to the overall recognition of exons that undergo AS is important. AS, therefore, depends on the variable assembly of the spliceosome across the pre-mRNA, a process that is aided by the transcript's interaction with regulatory factors that cooperate with spliceosomal components in defining exons. AS results in greater proteomic complexity and genetic diversity, however, when disrupted it can cause human disease [18–22]. Thus, pre-mRNA splicing is an essential step of gene expression.

Regulation of Splicing

Many factors contribute to the definition and recognition of an exon and how efficiently a pre-mRNA spliced. Splicing of internal exons relies on the recognition of its 5'ss and 3'ss. Efficient recognition of splice sites by the spliceosome is facilitated through combinatorial contributions of several key parameters, some of which include the splice site strength, splicing regulatory elements (SRE), and the exon/intron architecture [23] (Figure 1.3). Each of these factors contributes to the overall affinity of spliceosomal components for the exon and ultimately exon inclusion levels.

Splice site strength determines the ability of U1 snRNP and U2AF to efficiently recognize the exon. The degree of sequence complementarity between U1 snRNA and the 5'ss as well as the length and pyrimidine content of the PPT at the 3'ss determines the binding affinity between splice sites and spliceosomal components, thus dictating the level of exon recognition [24]. A strong 5'ss is defined by high sequence complementarity with U1 snRNA and a weak 5'ss is defined by low sequence complementarity [25]. This simple binding difference allows for various ways of deriving 5' splice site scores [26, 27]. The strength of the 3′ splice site requires the recognition of three important factors: the BPS, PPT, and the AG dinucleotide. The PPT has shown the greatest sequence variability, thus, strong or weak splice sites are mostly classified by the composition and length of the PPT [28–30]. The importance of these splice site scores can be seen through the use of the combined 5'ss and 3'ss scores, where constitutive and alternative exons can be more efficiently distinguished from each other [31, 32].

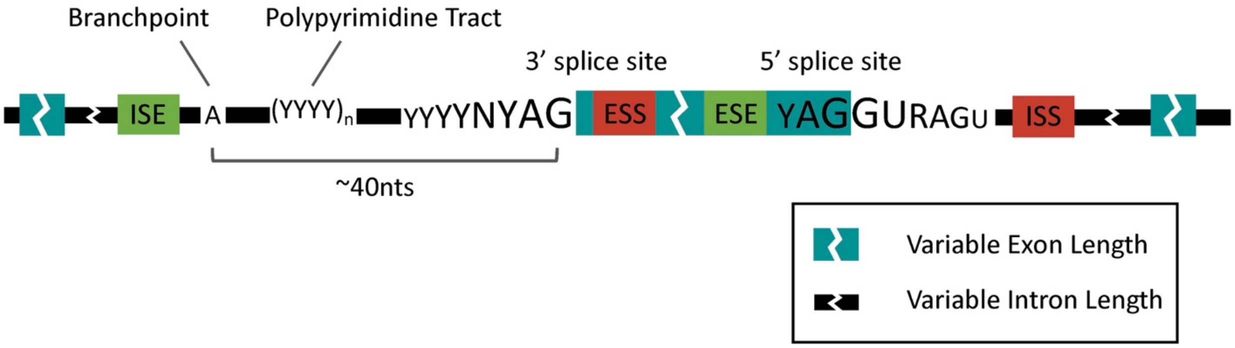


Figure 1.3. Splicing Regulatory Elements.
Schematic of splicing regulatory elements that control splice site recognition and regulation. Splice site sequences for both the 3’ss and 5’ss, the branch point sequence, polypyrimidine tract as well as locations of intronic and exonic enhancer or silencer sequences are depicted.

SREs are cis-acting sequence elements that serve as binding sites for splicing regulatory proteins that either increase or decrease spliceosomal recruitment. They can be defined by four categories: exonic or intronic splicing enhancers (ESEs or ISEs), and exonic or intronic splicing silencers (ISEs or ISSs). Two main classes of regulatory proteins that interact with SREs are SR proteins, classical splicing enhancers, and hnRNPs, classical splicing repressors.

SR proteins are a family of serine/arginine (SR)-rich proteins that contain an RS domain and at least one RNA recognition motif (RRM), which is essential for RNA binding. The RS domain has been shown to enable protein-protein and protein-RNA interactions [33, 34] and to mediate the recruitment of the spliceosome and modulate splice site recognition [1, 35–38]. SR proteins play a critical role in constitutive and AS through interactions with ESE sequences [39–41]. Additionally, some SR proteins have also been shown to promote E complex [34, 42].

Heterogenous nuclear ribonucleoproteins (hnRNPs) are another family of splicing factors that also contain an RRM. They are known to influence many aspects of an RNA, including its stability, its transcription, its translational regulation [43], and its AS [44–49]. Historically, these two families of splicing regulators have been defined by their classical roles: SR proteins were known activators and hnRNPs were known repressors of splicing. However, SR proteins can also demonstrate a repressive role on the 3'ss by binding to the BPS and preventing the recruitment of U2 snRNP [33, 50]. SRSF9, an SR protein, was also shown to mediate the skipping of exon 7 of hnRNP A1 [51]. HnRNP F, hnRNP H, and the hnRNP-like protein TIA-1 have also mediated 5'ss activation from a downstream intronic position in several pre-mRNAs [52–54].

The impact a SRE has on pre-mRNA splicing can be context-dependent. It has been shown that an ESE sequence to which an SR protein binds can act as an ISS when present in an intron [55], and a GGG motif, a known ESS, can also be an ISEs [56]. Moreover, almost all the common splicing regulatory proteins tested by Erkelenz et al. activated splicing depending on their binding position relative to a regulated 5'ss; SR proteins always activated from an exonic location and hnRNPs always activated from an intronic location. Interestingly, SR proteins repress splicing when located intronically and hnRNPs repress splicing when located exonically [57]. Overall, these studies established that both classes of splicing regulators have the ability to promote or repress splicing, antagonistic activities that simply depend on whether the splicing regulator binds upstream or downstream of a regulated splice site. Thus, SR proteins and hnRNPs are functionally interchangeable and their ability to regulate splicing depends on the location of their exonic or intronic binding location. However, the mechanisms mediating the position-dependent regulation by SR proteins and hnRNPs have yet to be elucidated. Understanding how a splicing regulator activates or represses splicing would expand our knowledge of the regulatory mechanisms that dictate AS.

A third key parameter that influences splice site recognition is the exon/intron architecture. This parameter is defined by the length of exons and introns and the mode of splicing across them. The majority of human exons are between 50-250nts long, with an average length of 120nts [58]. Human intron sizes have a greater range, with an average 3,400nts length [59]. Splice sites are generally recognized across an optimum length [1, 60] and it is the length of an intron that has been shown to dictate the mode of splicing. There are two proposed mechanisms of splice site recognition by the spliceosome. The first model states that the spliceosome recognizes the 5'ss and 3'ss simultaneously across an intron when an intron is <250nts [61, 62]. This model was termed "intron definition" [58]. When an intron is significantly longer, initial splice site recognitions occurs across exons, a process referred to as "exon definition" [58, 63]. Given that the vast majority of exons in the human genome are short and introns are long, splice sites within the human genome are generally recognized across an exon [1, 64]. In lower eukaryotes such as Drosophila, where introns are shorter and exons are longer, intron definition is the frequent mode of splice site recognition [58, 62]. Given the observed differences in "intron-defined" or "exon-defined" splicing efficiencies, these modes of spliceosome assembly have allowed predictions regarding the frequency of AS. For instance, human and Drosophila exons flanked by long introns are more likely to be skipped than exons flanked by short introns [62].

While the presence of a 5'ss and a 3'ss define internal exons, first and last exons represent a different class of exons that require other signals for their recognition. First exons lack an upstream exon and 3'ss while last exons lack a downstream exon and 5'ss. The 5' cap and its associated nuclear proteins assist in the recognition of the first exon [65]. Last exons have been shown to be recognized through 3'ss definition and the polyadenylation machinery that assembles at the downstream AAUAAA polyadenylation motif [66, 67]. Thus, the cap binding complex and the polyadenylation machinery are critical components for terminal exon recognition.

Evolutionary Implications on Splicing

Orthologous genes are genes in different species that evolved from a common ancestral gene. AS events that are conserved in orthologous genes can be seen as evidence of functionally significant events. Thus, functionally important exons are believed to be evolutionarily conserved and those that are non-functional may be removed by purifying or negative selection. A number of sequence and conservation feature differences have been identified when comparing alternatively and constitutively spliced exons. These differences include sequence conservation in constitutive exons, shorter lengths of alternatively spliced exons, greater sequence conservation between orthologous exons, and the preservation of the reading frame (labeled as symmetrical exons that are divisible by 3) [68–72]. Understanding the functional importance of evolutionary conservation in defining characteristics of the genome is of great significance, and this pursuit can be facilitated through the exploration of evolutionary changes on AS.

Two models have been suggested for the evolution of AS. The first model states that weak splice sites of internal exons cause the splicing machinery to skip internal exons, allowing for the generation of new transcript variants without actually altering the full-length transcript. This model is supported by findings that alternatively spliced exons display weaker splice sites than their constitutive counterparts [15]. Furthermore, lower conservation of alternatively spliced cassette exons is indicative of recent or rapid evolutionary changes [73]. The second model suggests that the evolution of splicing regulatory factors may force constitutive exons to become alternatively spliced. A prediction of this model is that a larger number of splicing regulators (SR proteins and/or hnRNPs) exist in species with more AS. Plants and metazoans, for instance, undergo extensive AS and the largest number of SR protein splicing regulators has been identified in *Arabidopsis thaliana*. At the other extreme, no SR proteins have been identified in *Saccharomyces cerevisiae*, which lacks AS [74]. While the splice site model seems more in line with evolutionary adaptation, it is possible that these two models co-exist and cooperate in influencing genome evolution.

The evolutionary impact on AS can also be discerned between contrasting exons that undergo intron or exon definition. Intron definition is regarded as an ancient mechanism, undergoing evolutionarily selection to remain short. Exon definition, on the other hand most likely evolved later, as it is the main mechanism in higher eukaryotes [75]. In addition, the prevalence of AS in higher eukaryotes is greater than in lower eukaryotes, demonstrating its lineage within the phylogenetic tree [75, 76]. As previously shown, maintaining an optimal exon size within the human genome is vital for efficient splicing and, therefore, important to the evolution of a gene, especially since exon length seems to have decreased over time [75]. Furthermore, several studies have demonstrated that AS increases from invertebrates to vertebrate species, suggesting that a driving force for evolution could be the generation of new alternatively spliced exons [76].

Understanding the origin of exons can also lead to greater insight into gene evolution. Several mechanisms have been proposed for the creation of an exon, two of which are exon shuffling and exonization. Exon shuffling is the duplication of an exon in a gene that leads to the creation of a new exon-intron structure, or the generation of a new gene through the introduction of a new exon from a different gene. There are several mechanisms that explain this mode of exon generation, one of which is intron-mediated recombination. Of particular importance to intronic recombination is the intron phase, the position that an intron is located in relative to a codon. Introns can disrupt the reading frame of a gene through insertion between two codons (phase 0 intron), after the first nucleotide of a codon (phase 1 intron) or after the second nucleotide of a codon (phase 2 intron) [77]. As genome complexity expands through the increased number of introns and repetitive elements, so does the chance of exon shuffling. Furthermore, it seems that the most influential role that exon shuffling has had on genome evolution is through the formation of modular proteins, which has been associated with the generation of multicellular organisms [75].

The second model of exon creation is exonization, a method of generating an exon through the introduction of genomic sequences from transposable elements such as Alu elements. Alu elements have the potential to convert into alternative exons paving the way for accelerated evolution [78]. They are short interspersed elements (SINEs) covering 11% of the human genome [79] with some predictions stating that there are roughly about 400 protein coding genes that contain these fragments, many of which are located within the intron of protein coding genes [80, 81]. Alu elements are around 300bp in length and contain nine potential 5′ss and fourteen potential 3′ss [82]. Four of these reside on the plus strand and 19 on the minus strand. This leads to the notion that the orientation of an Alu element could lead to its exonization, particularly when that orientation is different from the direction of transcription. In addition, Alu exons are more likely to be alternatively spliced [82], although mutations that lead to constitutive splicing of these Alu exons can cause human genetic disorders [83]. However, most constitutively spliced Alu exons have been inserted into the untranslated (UTR) regions of a gene and, therefore, do not affect the final protein product [84]. Moreover, alternatively spliced exons have shown a greater frequency of Alu elements [82]. In fact, Alu elements are only found in primates and primate-like mammals known as prosimians [85]. This highlights the existence of evolutionary pressures that prevent Alu exons from becoming constitutively spliced in coding regions and their important role in enhancing the evolution of primates.

The location of SREs within the pre-mRNA has also been shown to display a non-random arrangement, suggesting that evolution has shaped their exonic or intronic presence

[27, 86, 87]. There is a higher density of activating splicing regulatory binding sites in exons located specifically near an exonic splice site. Similarly, a higher density of repressing splicing regulatory binding sites are located intronically near splice sites [87]. These observations illustrate an enrichment of sequence conservation near splice sites, specifically for exons undergoing AS [71, 87]. Consequently, any pre-mRNA mutations within these regions could alter the relationship between cis- and trans-acting factors, ultimately resulting in different protein products. Thus, the exonic and intronic binding site arrangement for regulatory factors must be under appreciable evolutionary pressure to conserve optimal splicing signals within the pre-mRNA.

Considering the above discussion, it is anticipated that exonic sequences share evolutionary pressures to encode the sequence of amino acids that dictate functional proteins and to maintain efficient pre-mRNA splicing. Therefore, exonic sequences are tuned in a way for coding and splicing pressures to co-exist. Yet, how can overlapping evolutionary pressures be enforced into coding sequences? The third codon position, or the wobble position, is the most variable of any codon position. It is possible, therefore, that existing splicing pressures are more concentrated on wobble or other coding variable positions across the exon. This is because the wobble position would permit the greatest variability given the degeneracy of the genetic code. Indeed, synonymous mutations have been identified as disruptors of both splicing [88–90] and translation [91, 92]. Synonymous mutations are silent mutations, changes to the DNA sequence that do not impact the encoded amino acid due to the degenerate nature of the genetic code [93]. Furthermore, previous work has characterized the contribution of splicing evolutionary pressures through the identification of synonymous mutations that change exonic inclusion levels [94]. Strategies were proposed to deconvolute splicing from coding pressures. Alignment of exonic sequences between species was only permitted if the query species maintained the same exon length when compared to the human exon length. Thus, for any given exon, a unique set of species alignments was created to improve correlations between pre-mRNA splicing pressures and variable codon positions. This exon size-filtered alignment approach could in theory be used to identify nucleotides that have evolved to mediate efficient exon ligation.

Polyadenylation

Splicing is part of a larger landscape of RNA processing that together regulates eukaryotic gene expression. The interplay between splicing and 3' polyadenylation and their concerted actions results in an increase in the coding potential and modulation of the outcome of gene expression. Therefore, understanding the interrelationship between these two important mRNA processing events is of great significance.

Polyadenylation is an essential step in the 3' end maturation of mRNA. The polyadenylation complex consists of CPSF (cleavage and polyadenylation specificity factor), CstF (cleavage stimulation factor), CFI and CFII (cleavage factors I and II), the single subunit poly(A) polymerase (PAP) as well as other accessory proteins. Cleavage relies on the endoribonuclease activity of CPSF73 and recognition of the poly(A) hexamer signal (AAUAAA) by WDR33 and CPSF30, all subunits of CPSF [95, 96]. CstF64, a subunit of CstF, binds to a GU rich region 10-30 nucleotides downstream of the poly(A) signal, followed by subsequent cleavage and polymerization of the poly(A) tail by the PAP [97].

Alternative polyadenylation (APA), the use of more than one poly(A) signal has recently gained consideration as another central regulator of gene expression and has been identified in more than 70% of mammalian protein coding genes [98–100]. The ability to choose a poly(A) site depends on the strength of the hexameric sequence, as well as the surrounding *cis*-elements [101, 102]. There are two general classes of APA, those that have alternative poly(A) sites located in internal introns or exons, from which an APA event will produce a different protein isoform, or APA events that occur in the 3’ UTR resulting in transcripts with different 3’ UTR lengths, but code for the same protein (Figure 1.4). Changing the length of the 3’ UTR can also affect the stability, localization, transport and translation of the protein, thereby adding another regulatory factor to APA [22, 103]. APA is also known to alter the non-coding 3’ UTR which may affect the interaction with miRNAs, in turn, affecting the translation and expression of the final sequence [104–106].

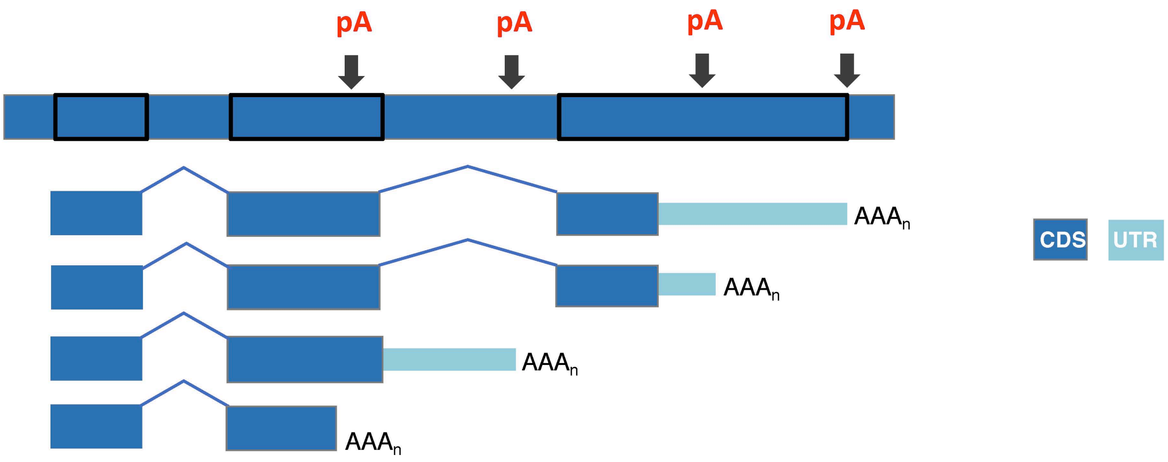


Figure 1.4. Model of APA Events.
APA generates alternative mRNA/protein isoforms that influence gene expression. There are many types of APA events, but in general can be grouped into two classes: 1) those that alter the 3' UTR affecting mRNA stability or 2) changes in the coding sequence (CDS) leading to changes in AS, in turn, generating different protein isoforms. Exons are denoted as dark blue boxes, UTR regions are denoted as bright blue boxes, and the poly(A) sites are denoted as pA.

Depending on the location of the poly(A) site relative to the stop codon, an APA event is defined as proximal or distal. Distal poly(A) signals generally have a conserved canonical signal, while proximal poly(A) sites have less canonical (non-conserved) sequences [21, 107]. Of the major proteins involved in polyadenylation, CstF and CF1m have been implicated in APA, along with several other factors [21, 108–111]. CF1m is a heterodimer composed of a small 25 KDa subunit (CF1m25) and a variable larger subunit composed of 59, 68 or 72 KDa. CF1m has been shown to stimulate cleavage and poly(A) addition through interaction with upstream RNA sequences (UGUAN). It has also demonstrated an ability to inhibit the binding of CPSF to the RNA and suppress poly(A) site cleavage [112]. Both CPSF and CF1m were also identified in purified spliceosomes [3, 113] Therefore, CF1m may inhibit binding of CPSF to the RNA through its interaction with pre-mRNA splicing machinery. Moreover, the large subunit of CFIm has shown structural similarity to SR proteins [111]. This suggests a link between both APA and splicing [3, 112]. Knockdown of CF1m25 has been shown to result in a shift from the typical distal poly(A) site in the terminal exon to a more upstream poly(A) signal [112, 114]. This indicates a possible function for CF1m25 in recognizing upstream poly(A) signals and inhibiting cleavage through interactions with splicing factors [114].

Another polyadenylation factor that plays an important role in poly(A) site recognition is CstF64. This protein binds directly to the RNA and interacts with the U/GU rich elements downstream of the poly(A) signal. CstF64 contains an N-terminal ribonucleoprotein-type RNA binding domain (RBD) and interacts with the CPSF subunit via CstF77 [115] and symplekin, for roles in cell growth, stability as well as polyadenylation site specificity [116]. CstF64 has also been shown to play a role as a regulator of APA [110], for example during B cell differentiation, whereby an increase in CstF64 levels leads to a shift from a distal site to a weaker upstream proximal polyadenylation site during IgM switching from a membrane bound to a secreted form [117].

Polyadenylation, just like other mRNA processing events is not isolated, but in fact is coupled with splicing, particularly in mammals at the last or terminal exon. Exon definition, based on the Berget model, states that U1 and U2 snRNP recognize the downstream 5'ss and the upstream 3'ss of an internal exon, which allows for correct positioning of the other snRNP's across the intron [58, 63]. However, the first and last exons do not fall into this model of exon definition as previously mentioned. The terminal exon is recognized through spliceosomal recognition of its 3'ss and interactions between splicing components and the polyadenylation machinery [66, 67]. This recognition occurs through direct interactions between U2AF and CF1m [118] or the PAP [119] as well as through direct interactions with CPSF [120] and the U2 snRNP component splicing factor 3b (SF3b) (Figure 1.5). U1-A, a U1 snRNP component, also demonstrates polyadenylation stimulation through interaction with CPSF160 [121]. In addition, U1 snRNP prevents premature cleavage and polyadenylation through the binding of U1 snRNA to cryptic poly(A) sites [122] and through direct interactions between U1-70K and the PAP [123, 124]. U1 snRNP also inhibits polyadenylation through interactions between its 5' end and the poly(A) signal preventing poly(A) addition and leading to the degradation of the pre-mRNA [125]. Coordinated efforts between the PAP and the nuclear poly(A) binding protein PABPN1 have also demonstrated the importance for terminal intron removal [126, 127].

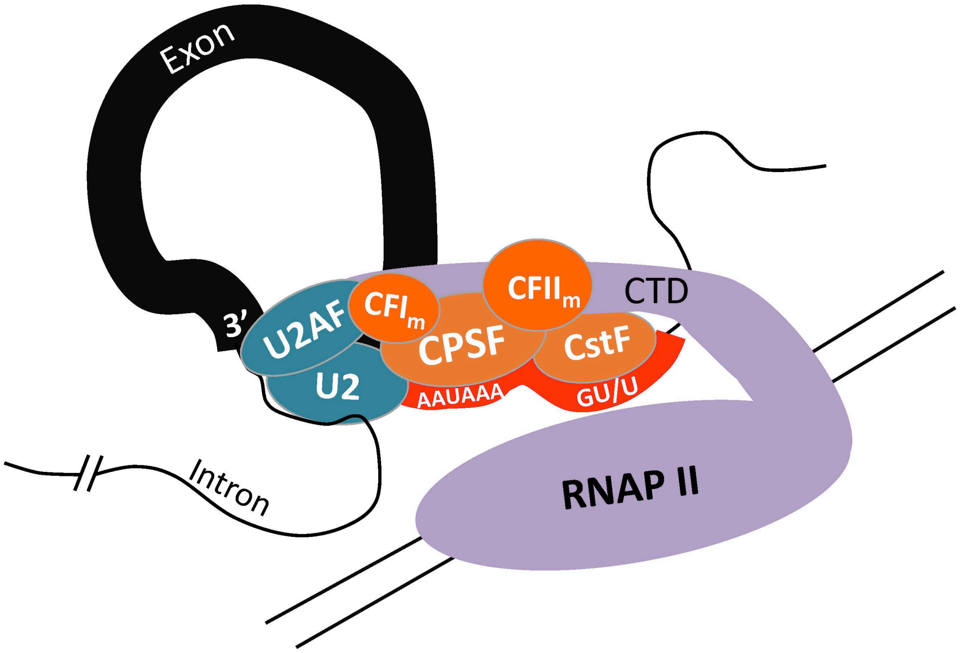


Figure 1.5. Splicing and Polyadenylation Factors at the Terminal Exon. Schematic representation of a simplified model of the coupling between polyadenylation and splicing at the terminal exon. Terminal intron removal requires the use of a functional poly(A) signal and 3'ss due to the lack of a 5'ss. Polyadenylation factors shown in orange, splicing factors shown in blue and key poly(A) sequence elements in red.

Additionally, the C-terminal domain (CTD) of RNA polymerase II maintains the spatial and temporal organization between splicing and polyadenylation factors [128–131]. This complex network of protein interactions at the terminal end demonstrate a vital role for the coordinated regulation and enhancement of splicing and polyadenylation.

The mechanisms of interaction and the degree of coupling between APA and AS are still under investigation. It has been established that cooperation between splicing and polyadenylation occurs at the terminal exon, however, it is not known whether upstream AS dictates APA or vice versa. The impact of such a coordinated process on gene expression and proteomic diversity could be significant. If APA is capable of altering the coding capacity of transcripts by influencing which exons undergo AS, it could dictate which mRNA isoform is expressed. Identifying such mRNA diversification mechanisms would further our understanding of how co-transcriptional processing and regulation occurs.

Splicing defects have been shown to play a role in many human diseases [35, 132], and splicing mutations have been implicated in numerous types of cancer [133–136]. Similarly, APA has been demonstrated to contribute to disease and cancer [18–20, 22]. It is therefore important to decipher the mechanisms of pre-mRNA splicing, the mechanisms of polyadenylation, the cooperation between splicing and polyadenylation, and the relationship between AS and APA.

Summary

Pre-mRNA processing is a complex process that requires the combined efforts of multiple factors, elements, and processing events. This chapter outlined the collaborative association of many cis- and trans-acting elements that guide the processing from pre-mRNA to mRNA. Yet, many aspects of this processing are not well understood. The following work aims to provide novel insights into pre-mRNA processing mechanisms by exploring splicing regulatory factor activities, by evaluating novel evolutionary conservation approaches to predict splicing outcomes, and by testing the extent of functional coordination between the splicing and polyadenylation machineries.

Chapter 2 describes the generation of an exon conservation database using key parameters of the human genome to extrapolate evolutionary conservation at the species level through an analysis of exon size and sequence conservation. The goal of this chapter is to use evolutionary conservation patterns to identify what influences the final sequence of an exon and to use this information to predict splicing patterns in human disease. Our database permitted the identification of fundamentally important architectural parameters of the human genome. While protein coding pressures control the nucleotide composition at fixed codon positions, analyses of nucleotide variations at wobble positions increased the probability of identifying splice altering nucleotides.

Chapter 3 focuses on understanding the position-dependent activity of the splicing regulatory proteins SRSF7 and TIA-1 and their mechanism of regulation. In activating or repressive conditions, U1 snRNP components display altered complex dynamics, demonstrating that the U1 snRNP integrity is modulated by position-dependent interactions with splicing activators and repressors.

Chapter 4 discusses the cooperation between APA and upstream AS events. Through genome-wide analyses, the mechanistic coupling between APA and pre-mRNA splicing was shown to be limited to terminal exon definition. The results also identified an intriguing role for the polyadenylation factor CstF64 in mediating AS through its effects on UTR selection of known splicing regulators such as hnRNP A2/B1.

CHAPTER 2

Exon Size and Sequence Conservation Improves Identification of Splice Altering Nucleotides

Summary

Pre-mRNA splicing is an essential step of gene expression that is regulated through multiple trans-acting splicing factors. These regulators interact with the pre-mRNA at intronic and exonic positions. Given that most exons are protein coding, the evolution of exons must be modulated by a combination of selective coding and splicing pressures. It has previously been demonstrated that selective splicing pressures are more easily deconvoluted when phylogenetic comparisons are made in the framework of exons of identical size. Our hypothesis is that exon size-filtered sequence alignments may improve the identification of nucleotides that have evolved to mediate efficient exon ligation. To test this hypothesis, an exon size database was created that filters 76 vertebrate sequence alignments based on exon size conservation. In addition to other physical parameters, such as splice site strength, gene position or flanking intron length, this database permits the identification of exons that are not only sequence conserved, but also size conserved. Highly size-conserved exons are always sequence conserved. However, sequence conservation does not necessitate exon size conservation. Our analysis identified exons that are unique to humans/primates; exons that may be considered evolutionarily young. A published dataset of ~5000 exonic SNPs that are associated with disease was also analyzed to test the hypothesis that exon size-filtered sequence comparisons increase the detection of splice-

altering nucleotides. Improved splice predictions could be achieved when mutations are at the third codon position, especially when a mutation decreases exon inclusion efficiency. The results demonstrate that coding pressures dominate the nucleotide composition at invariable codon positions and that the exon-size filtered sequence alignment approach permits the identification of splice-altering nucleotides at wobble positions.

Introduction

Splicing is a vital step in gene expression that relies on the correct recognition of exons and removal of introns in a process coordinated by the spliceosome. There are many factors that contribute to the identification of an exon and how efficiently it is spliced. Regulation of splicing and, therefore, alternative splicing requires many trans-acting factors, such as SR and hnRNP proteins, but also cis-acting elements, such as regulatory sequences within the intron or exon necessary for correct splice-site recognition [23]. Some of these cis-acting factors include splice site (ss) strength at the 5' or 3' end of an exon as well as splicing regulatory elements (SRE) to which trans-acting factors bind [137, 138]. Splice-site strength is based on sequence complementarity of U1 snRNA with the 5'ss and a longer sequence motif of ~ 23 bases for the 3'ss, which makes up the binding site of U2AF. The Maximum Entropy Score (MES) is a computationally derived value assigned to splice site sequences based on the modeling of short sequence motifs around the 5'ss or 3'ss using the maximum-entropy principle [26]. The numerical scoring permits the designation of strong (MES > 8) or weak (MES < 6-7) splice sites, [32] with an average 5'ss score for constitutive exons of MES = 8.3 and MES = 8.8 for the 3'ss [73]. Any mutations within the pre-mRNA can alter how cis- and trans-acting factors work together for efficient splicing, in turn influencing the final protein product. These pre-mRNA positions must then be under evolutionary pressures to maintain optimal splicing signals needed for intron removal.

In addition to correct splicing of the pre-mRNA, every amino acid and its sequence are also essential. Therefore, when evaluating the make-up of any coding sequence, two sets of evolutionary pressures are needed to maintain a functional product: pressure to splice correctly and pressure to code for a functional protein. Given the importance of these two evolutionary sequence pressures, they would have to coexist. Due to the degenerate nature of the genetic code [93], the third position of a codon (wobble position) permits variability within the human genome. It is therefore possible that necessary splicing pressures could be overrepresented within the wobble position. Uncoupling evolutionary coding and splicing pressures would be difficult without considering the role of synonymous mutations. Synonymous mutations are silent mutations that don't alter the protein sequence but are not necessarily silent with respect to how an exon is spliced or how a protein is folded [92]. Synonymous mutations have been indicated in the modulation of splicing, translation [91, 92], and mRNA stability [139]. Previous studies from the Hertel lab have systematically characterized the contribution of splicing evolutionary pressures through the identification of synonymous mutations that alter splicing efficiency through changes in exonic inclusion/exclusion levels [94]. The results from these studies demonstrated an important limitation within PhyloP [140], a widely used computational approach to determine base-wise conservation. PhyloP does not delineate whether a certain nucleotide has a positive or negative influence on splicing, in part because it does not take into consideration altered splicing pressures expected to be imposed by exons of different sizes. Taking into account the exon/intron architecture as an evolutionary feature was shown to have a significant impact on deconvoluting splicing from coding pressures [94]. Based on these findings, species should be aligned by those containing the same exon length prior to computing PhyloP scores. For each human exon, a different set of species may be used for sequences alignment and conservation score caculations (Figure 2.1). Based on these findings, we hypothesize that exon size-filtered sequence alignments may improve the identification of nucleotides that have evolved to mediate efficient exon ligation. To address this hypothesis, an exon size conservation database was generated that also reported on additional exon features such as the splice-site strength, the exon and intron lengths, the exon size variation, and the length and sequence conservation across 76 species. Using this database, it is possible to extrapolate evolutionary conservation at the species level. To test the hypothesis that exon size-filtered sequence comparisons increase the detection of splice altering nucleotides, a published dataset of ~5000 exonic disease associated SNP's was analyzed.

The exon conservation database permitted the identification of exons that are not only sequence conserved, but also size conserved, demonstrating that highly size-conserved exons are always sequence conserved. The analysis also allowed for identification of exons that are unique to humans/primates and that are evolutionarily young. Using the published dataset of ~5000 disease associated exonic SNPs, improved splice predictions could be achieved when nucleotide variations are located at the third codon position. These results demonstrate that coding pressures dictate the nucleotide composition at inflexible codon positions and that the conserved third codon positions largely uphold splicing pressures.

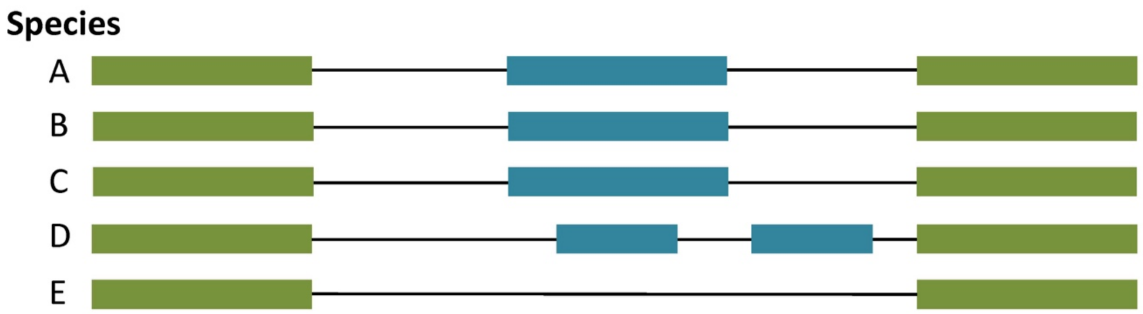


Figure 2.1. Schematic of Different Exonic Splicing Pressures.
Uncoupling of splicing pressures from coding pressures, species alignment requires exons of the same length. Different sized exons contain different splicing pressures. Only species with the same sized exons can be aligned for size conservation analysis, such as species A, B, and C. Species D and E are excluded from the conservation analysis. Exons are denoted as green and blue boxes, introns are denoted as thin black lines.

Results

Architecture of the human genome within the conservation database.

Previous findings identified the importance of exon size-filtered alignments in identifying exon conservation between human and other vertebrate species [94]. Using this approach, a database was generated aligning length-filtered exons across 76 species. Based on conserved exon size among species, sequence conservation scores were determined using PhyloP [140]. Other exon parameters were also identified, which included the exon position within the gene, exon length, the type of exon (first, internal, last, or single), flanking intron lengths and 5'ss and 3'ss scores. This information also allowed for investigations into the architecture of the human genome. Exon length and splice site score analyses were carried out for 184,796 exons (18,225 genes), representing the exon categories *first, internal, last* and *single exons* (Figure 2.2). A large proportion of *single exons,* which make up the smallest category of exon types are ~1000nts in length (Figure 2.3A). *First exons* exhibit strong 5'ss scores and are on average shorter exons than *last exons* (Figure 2.3B, 2.3C). *Last exons* generally harbor strong 3'ss scores with a larger distribution of exon lengths. As expected, *internal exons* (Figure 2.3D) make up the largest of the exon types (156,383 exons) with a very tight size distribution around 50-250nt and an average exon size of 120nts. This *internal exon* length distribution is consistent with previous findings that demonstrated that the optimal exon length for efficient splicing is between 50-250nts [141, 142]. In addition, *internal exons* have a tight distribution of 5'ss and 3'ss scores, with average scores around 8 MES, consistent with previous findings [32, 73].

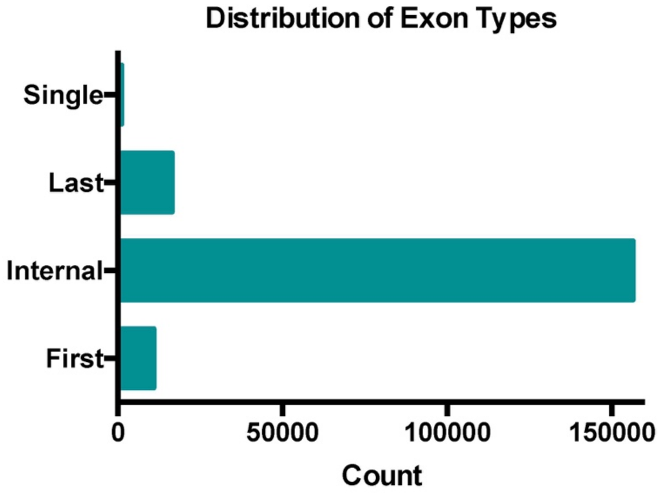


Figure 2.2. Distribution of Exon Types in Exon Conservation Database.
The exon conservation database is composed mainly of internal exons. The database totaled 184,796 (18,225 genes) that are categorized into first, internal, last and single exons.

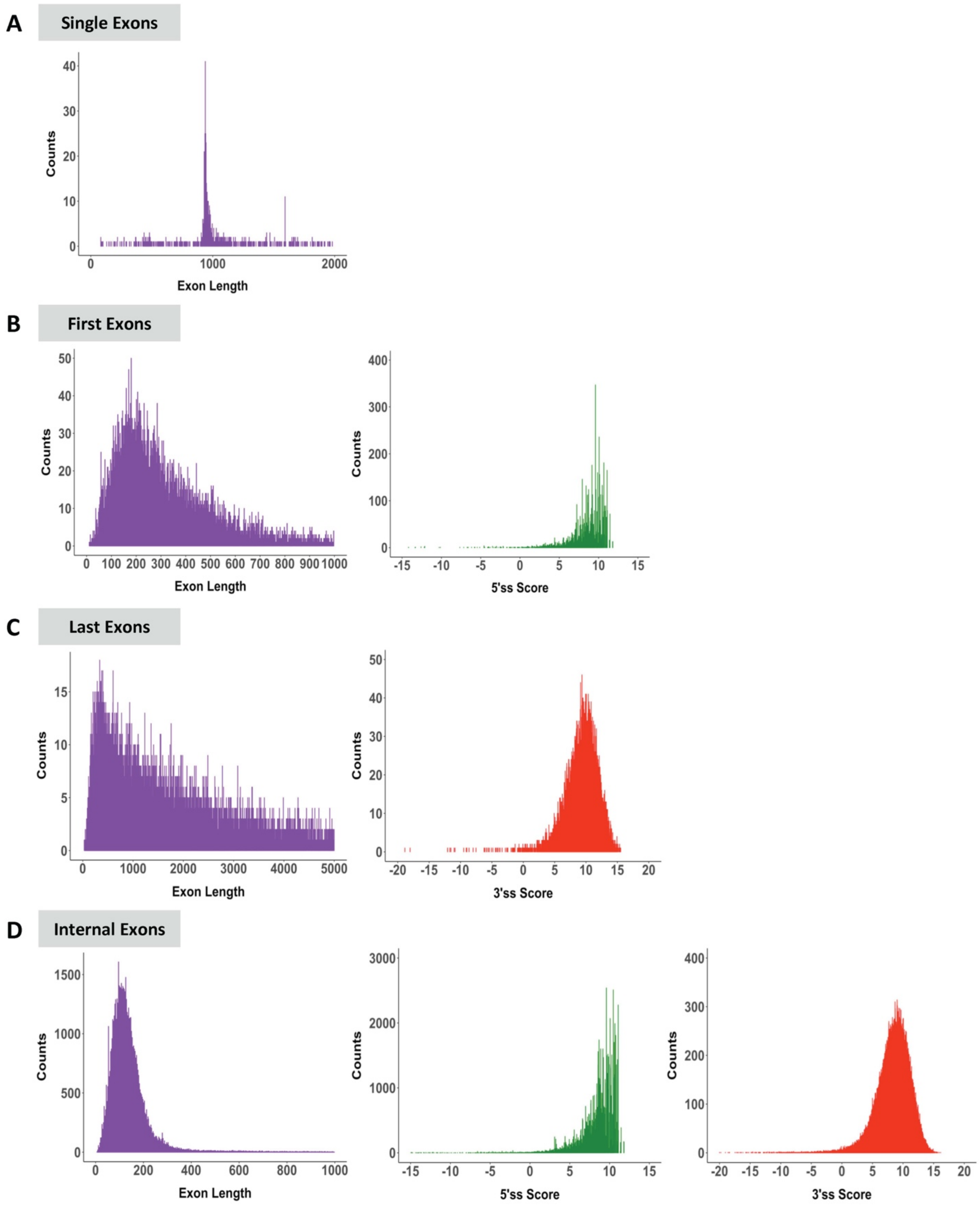


Figure 2.3. Architecture of the Human Genome.
Exons categorized into four distinct types contain varying exon length distributions, but maintain strong splice site scores (MES >8).

Distribution of length-conserved exons across 76 species.

The exon size or length conservation score, also defined as the "Ultra-In" score (see Methods), was obtained by comparing 76 vertebrate species to the human reference genome, and determining for each exon, the number of species that maintain the human exon size. This score ranges from 0, representing a unique exon size in human, to 76, which represents exon size conservation across all species evaluated. Low length conservation is defined as a score of <10 (10 or less species with exon length conservation), moderate length conservation is defined by a score between 10-40 and high length conservation as a score >40. An analysis of size conservation frequencies for all exons highlights the two general populations that emerge (Figure 2.4A). These two populations are those with low exon length conservation, observed only in human or in a small number of species, mainly primates (data not shown) and those with high exon length conservation across a larger number of species.

Striking differences are observed when categorizing exons by length, those that are short, <50nts (Figure 2.4B), average sized exons 50-250nts (Figure 2.4C), or long exons >250nts (Figure 2.4D). The distribution of length-conserved exons seen in (Figure 2.4A) is mainly reproduced by the 50-250nts exon size group (Figure 2.3D). It is also characteristic of the distribution across *internal exons*, the largest population of exons (data not shown).

However, the population of low length-conserved exons is no longer present in this exon size bin. Rather, this population is overrepresented in the exon size group that are >250nts in length (Figure 2.4D). The fact that the majority of exons between 50-250nts are highly length conserved suggests that optimal exon size is an important evolutionary conservation feature.

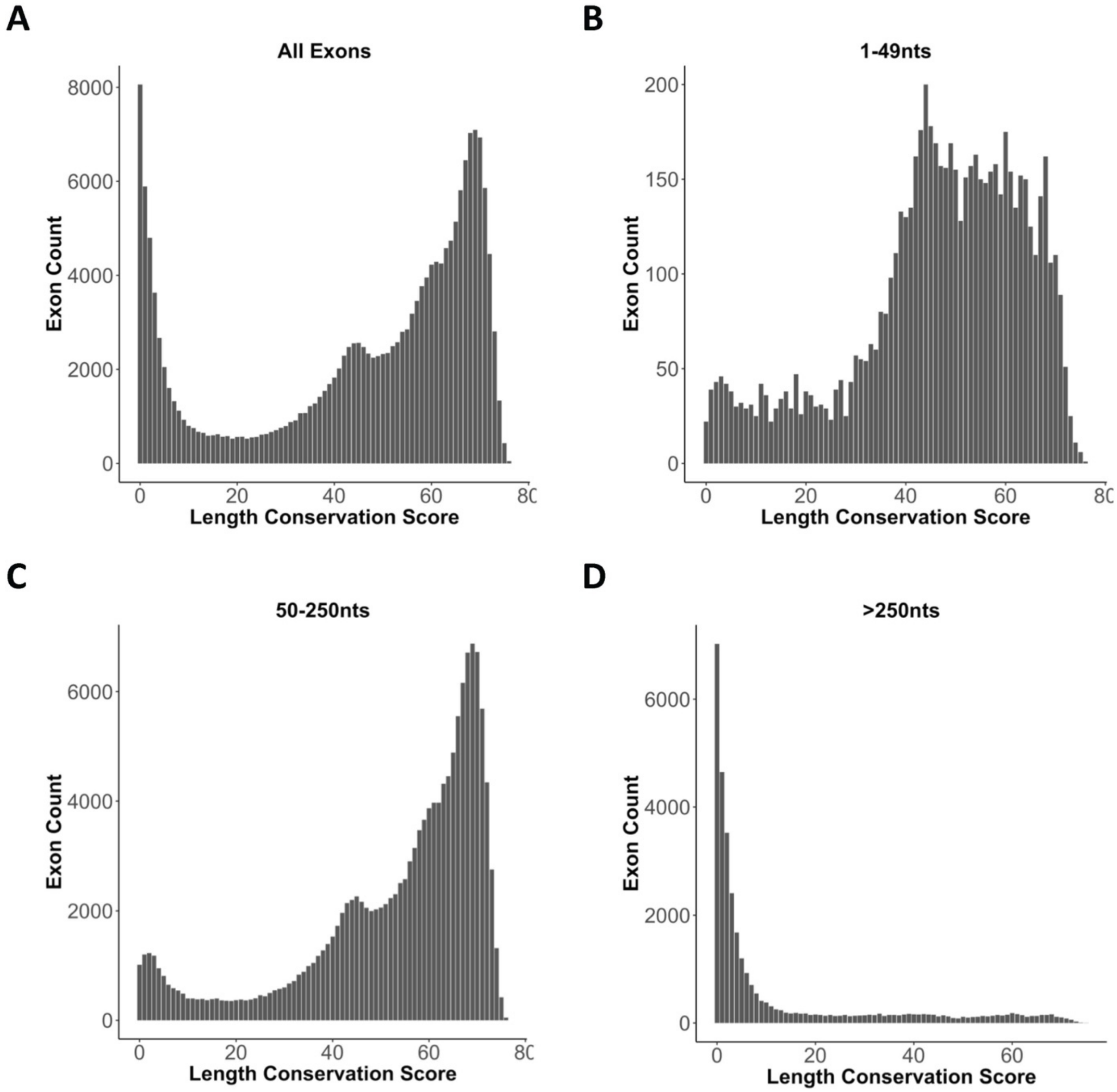


Figure 2.4. Distribution of Exons Across 76 Length Conserved Species.
Length conservation score represents the number of species with length conserved exons when compared to human. Frequency of all types of exons A) all exon lengths B) between 1-49 nucleotides in length C) between 50-250 nucleotides in length and D) longer than 250 nucleotides in length.

High length-conserved exons are always sequence conserved.

As demonstrated above, the majority of exons are size conserved (score >40). However, it is unknown to what degree exon length conservation and exon sequence conservation co-vary or evolve independently. To evaluate the correlation between sequence and architectural features, the average PhyloP score (sequence conservation score) and the length conservation score was determined for each *internal exon*. A tight correlation was identified between length and sequence conservation, R=0.83 (Figure 2.5). The largest population of exons is characterized by high length (>40) and high sequence conservation (PhyloP score > 3), while fewer exons are either not length and/or sequence conserved. In other words, exons that have high sequence conservation usually also demonstrate high length conservation, but not all high size-conserved exons are sequence conserved. This strong correlation represents a potential convergence between size and sequence over the evolutionary lineage of an exon in defining optimal length and sequence elements for efficient exon recognition.

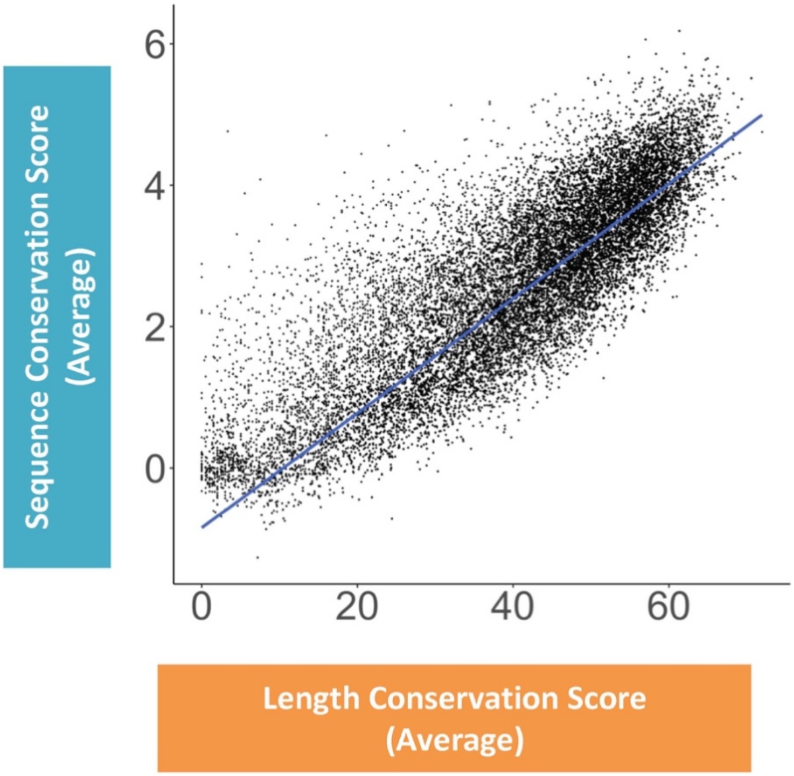


Figure 2.5. Correlation Between Length and Sequence Conservation.
Highly size-conserved exons are always sequence conserved. Average score of exons obtained for sequence and length conservation was correlated. Regression line represented by the blue line. R = 0.83.

Distribution of length-conserved species across various parameters.

To analyze the correlation between exon size conservation and common exon recognition signals *internal exons* were arbitrarily binned into size conservation groups based on their “Ultra-In” length conservation score. The average exon length is longer for exons that have poor length conservation (Figure 2.6A). In addition, the sum of splice site scores was observed to be lower for the low length-conserved group 0-10 (Figure 2.6B). Consistent with Figure 2.5, exon length and sequence conservation are highly correlated with minimal sequence conservation overlap between the two extreme categories 0-10 and 70-76 (Figure 2.6D). These observations support the conclusion that length and sequence conservation are highly correlated evolutionary features.

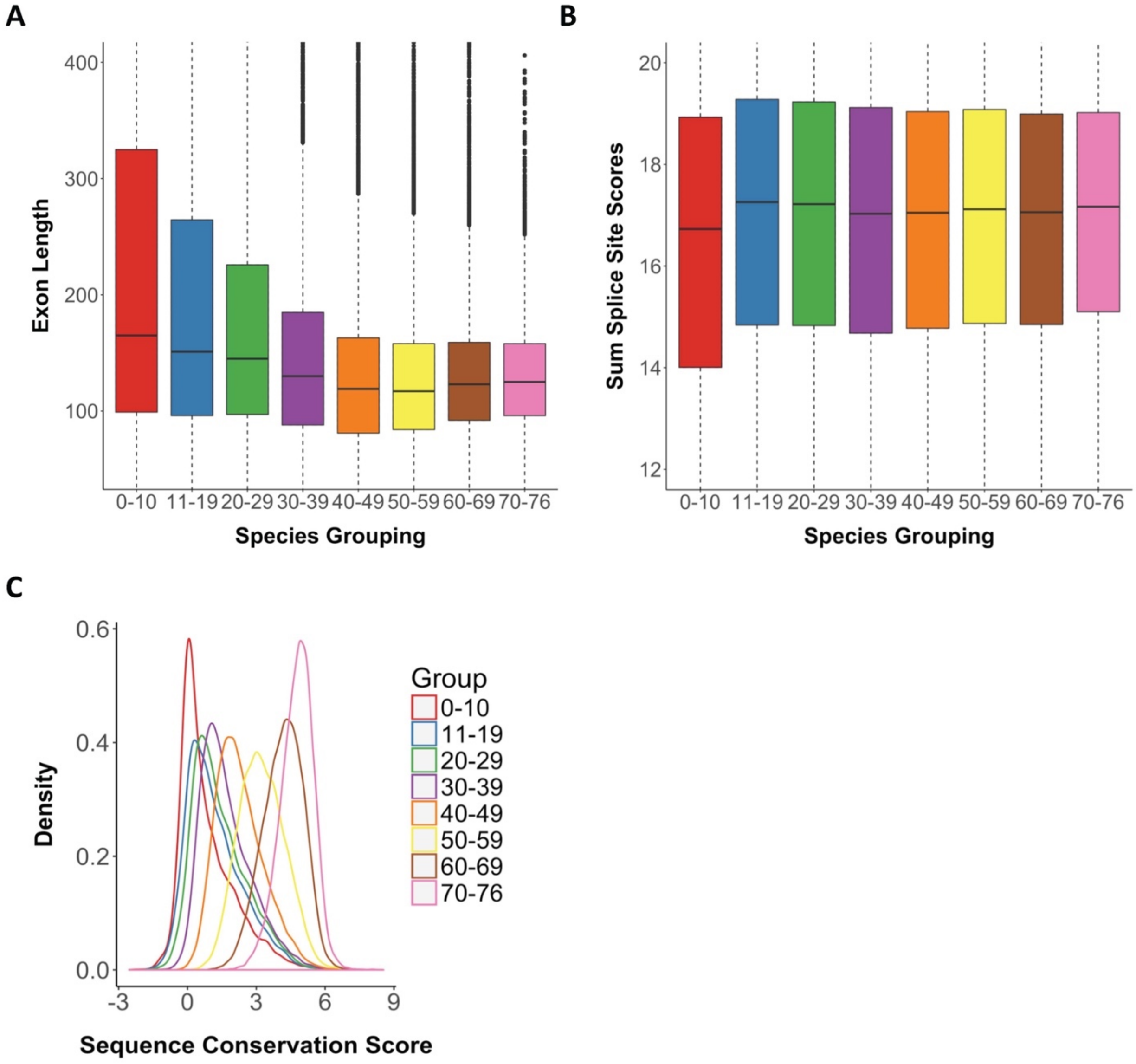


Figure 2.6. Distribution of Length-Conserved Species Across Various Parameters. Internal exons arbitrarily binned into groups of 10 (last group in a bin of 6) based on length conservation score and compared across A) exon length B) sum of the 5'ss and 3'ss scores C) density distribution of the sequence conservation score across the various binned groups. Outliers were generally excluded from boxplots for ease of visualization of the differences between the averages.

Primates display the greatest similarity in exon architecture with humans.

Using exon length conservation as a method of reassessing species closeness, species with similar exon lengths when compared to human were identified. The number of times an exon had the same length in each of the 76 species located within the 0-10 group relative to human was plotted. The assumption made in this analysis is that species that are the more closely related to humans should exhibit the highest representation of exon length conservation in the 0-10 category where minimal overall length conservation is seen. Peak similarity was observed to be the strongest for those primates with the greatest length conservation when compared to human (Figure 2.7) with a striking drop off in exon size similarity from marmoset to bushbaby. Marmosets are small monkeys, considered to be part of the new world monkeys, appearing roughly 30 million years ago. Bushbabies are generally the size of a squirrel and considered to be prosimians, primate-like mammals that appeared much earlier, roughly 60 million years ago [143]. The difference between the evolutionary appearance between these two mammals could explain the reduction in exon size variation that is seen, given that new world monkeys perhaps evolved from prosimians [143]. Interestingly, this exon size conservation analysis is highly consistent with what is known about the phylogeny of primates. Chimps are the most closely related primate to humans, followed by gorillas and orangutans and lesser apes such as the gibbon, the old world monkeys, the baboons, and new world monkeys such as the squirrel monkey and marmoset [143]. In summary, this analysis demonstrates that exon size conservation is a genomic feature that significantly contributes to evolutionary trends in mammals. Size conservation could therefore participate in evolutionary fitness of the species overall and can be used to retrace the lineage of species creation.

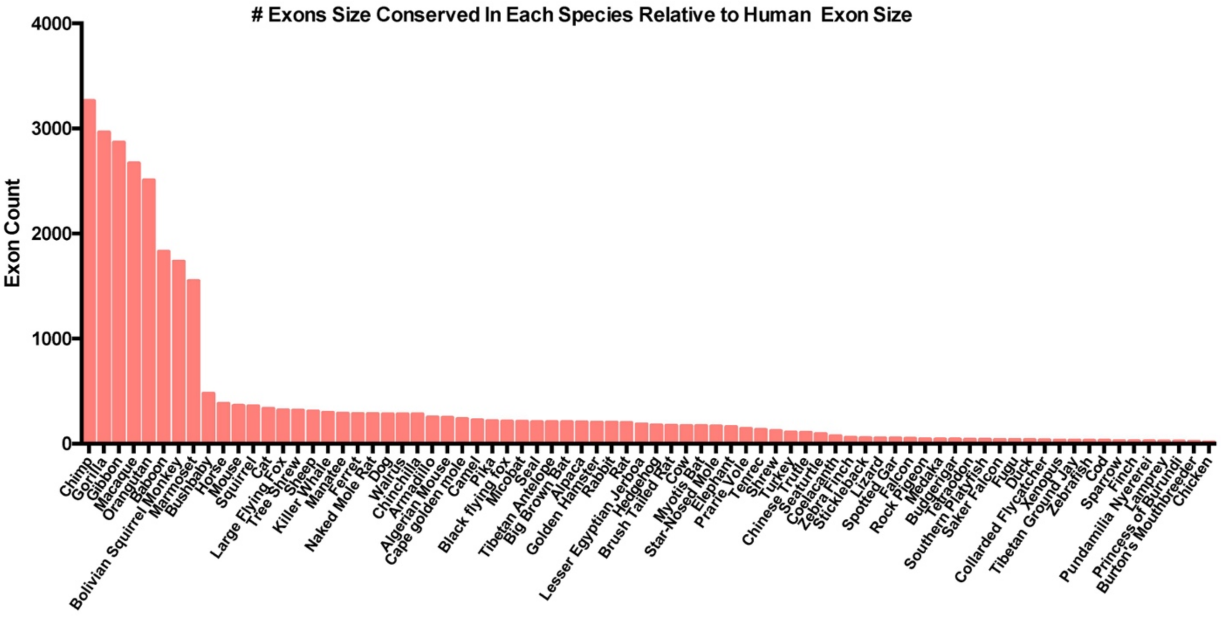
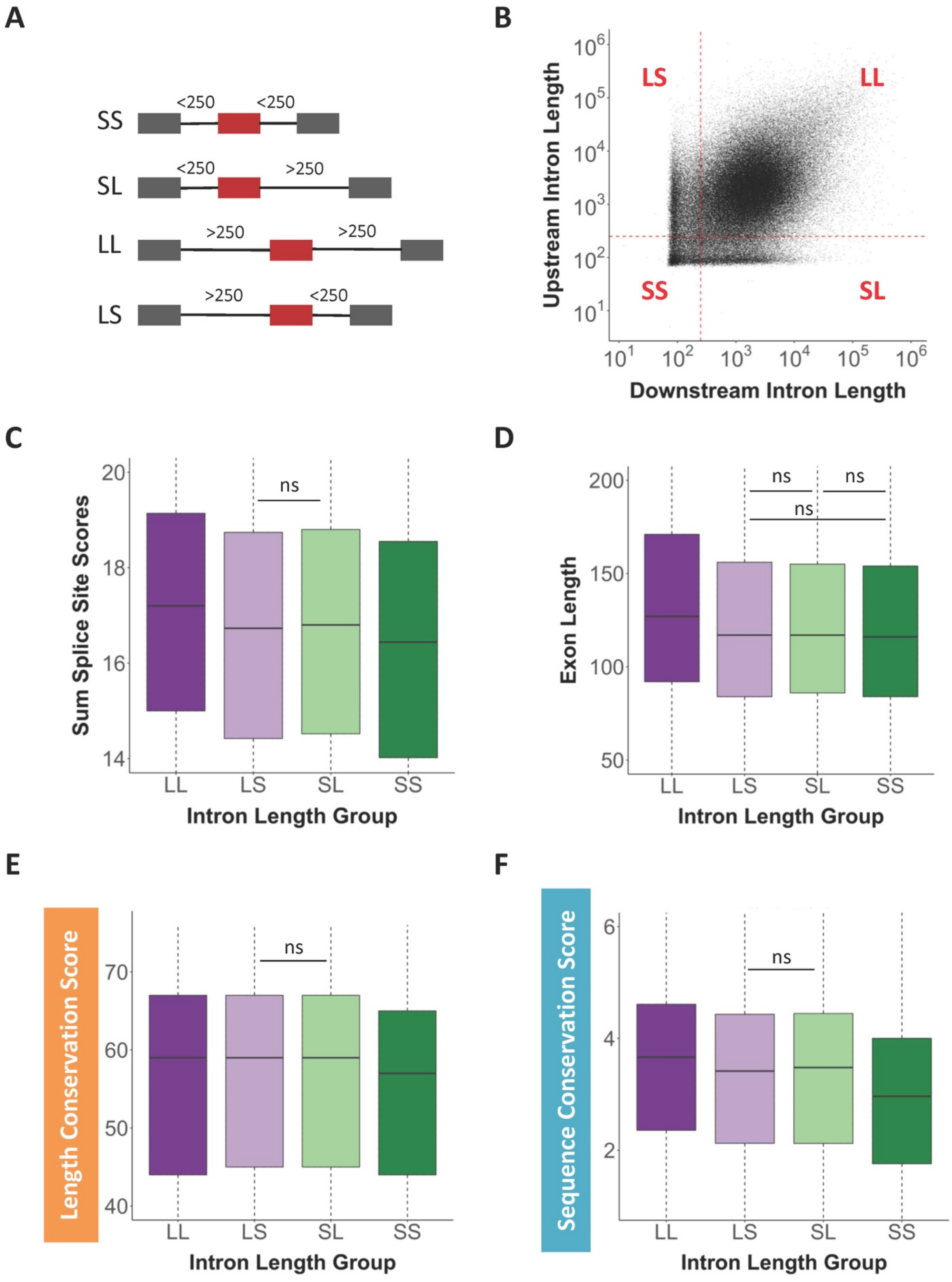


Figure 2.7. Comparison Between Human and Other Species Exon Size.
Exon size conserved primates demonstrate greatest similarity with humans when observing the low length conserved population of species (0-10 group). Species names on x-axis and count of exons per each species depicted on the y-axis.

Distribution of intron lengths flanking internal exons of different size conservation.

It is known that the length of introns flanking *internal exons* defines how an exon will be recognized and spliced [62], either through exon or intron definition. To understand the variation of flanking intron lengths around exons with high length and sequence conservation, human upstream and downstream intron lengths were recorded and analyzed. Four main intron length categories were designated based on the length transition between intron and exon definition. This length of an intron was defined by the transition from intron definition to an exon definition model of splice-site recognition, based on previous findings [62]. An *internal exon* can be flanked on either side by introns that are both short in length (<250nts), designated as SS ("short short") introns, or the upstream intron is short, and the downstream intron is long (>250nts) designated as SL ("short long") introns. *Internal exons* can also be flanked by long introns designated as LL ("long long") introns, or a long upstream and short downstream intron designated as LS ("long short") introns (Figure 2.8A). A correlation of internal exons and their flanking intron size nicely highlights the exon definition population of internal exons (LL) and the population of exons that are at least partially intron defined (LS, SL, SS). Interestingly, this genome view demonstrates underrepresentation of flanking intron sizes that are at the transition point between exon and intron definition, and it illustrates a remarkable demarcation for minimal intron size (~75nts) (Figure 2.8B). Using these four intron length groups, the sum of the 5'ss and 3'ss scores was correlated. On average, significantly stronger splice site scores are observed when flanking introns are long compared to splice site scores of *internal exons* flanked by SS introns (Figure 2.8C).

Figure 2.8. Distribution of Intron Lengths Flanking Internal Exons.
A) Cartoon depiction of the four, intron length defined exon groups. Exons that are flanked by SS (short short), SL (short long), LL (long long), or LS (long short) introns. Gray boxes represent flanking exons, red boxes represent internal exons, black lines represent introns of various lengths. Numbers indicate nucleotide lengths of introns. B) Correlation between upstream and downstream intron lengths flanking internal exons. Each quadrant is defined by the length of the upstream and downstream intron length respectively. Intron length, defined by the transition from intron to exon definition is delineated by red dotted vertical and horizontal lines. The four intron groups were correlated with C) sum of the 5'ss and 3'ss scores D) internal exon length E) length conservation score and E) sequence conservation score. All intron length groups have $p > 0.05$ unless marked with "ns" – not significant.


A
SS
<250
<250
SL
<250
>250
LL
>250
>250
LS
>250
<250
B
Upstream Intron Length
Downstream Intron Length
LS
LL
SS
SL
10^1
10^2
10^3
10^4
10^5
10^6
C
Sum Splice Site Scores
14
16
18
20
ns
LL
LS
SL
SS
Intron Length Group
D
Exon Length
50
100
150
200
ns
ns
ns
LL
LS
SL
SS
Intron Length Group
E
Length Conservation Score
40
50
60
70
ns
LL
LS
SL
SS
Intron Length Group
F
Sequence Conservation Score
2
4
6
ns
LL
LS
SL
SS
Intron Length Group

This observation suggests that exons undergoing exon definition require stronger splice sites on average than those that undergo intron definition. In addition, *internal exons* flanked by LL introns were on average significantly longer in length than those flanked by LS, SL, and SS introns (Figure 2.8D). The average length of these exons, however, was still within the optimal exon length window of 50-250nts. Regardless, this exon size difference could represent the need for LL exons to provide additional trans-acting factor binding sites to ensure efficient spliceosomal recognition.

When exon length conservation was evaluated for the four intron groups, exons flanked by LL introns demonstrated significantly higher size conservation than exons flanked by SS introns (Figure 2.8E). Similarly, exons flanked by LL introns are characterized by significantly higher sequence conservation when compared to exons flanked by SS introns (Figure 2.8F). Interestingly, the distribution of intron lengths downstream of *first exons* is noticeably broader than that observed for *internal exons* (Figure 2.9A). These results demonstrate that first introns are generally longer than subsequent introns. Moreover, a comparison between internal and last intron size did not reveal differences in size distribution (Figure 2.9B). These results show that *first exons* have greater length distribution of their downstream intron, as compared to *internal* and *last exons.*

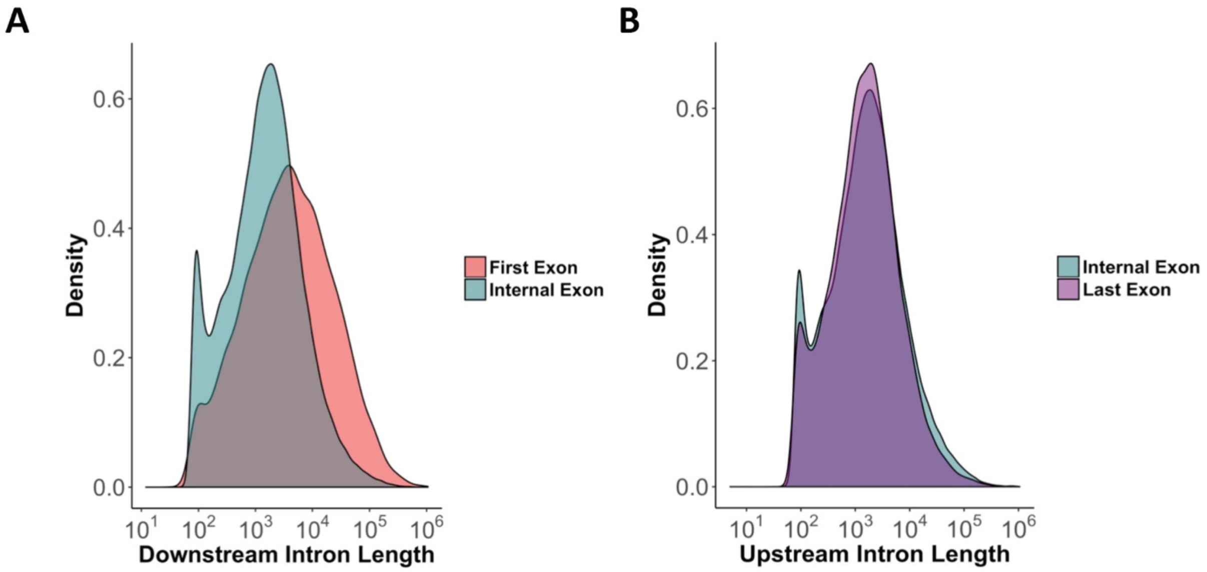


Figure 2.9. Intron Length Density Distribution in Relation to Exon Type.
Density distribution of downstream and upstream intron lengths relative to the position of the exon.

Distribution of splice site scores across internal exons.

Splice site scores also play an important role in efficient exon recognition. Previous work demonstrated that the transition between exon inclusion and exon exclusion is defined within a very narrow window of splice site strength designation (MES of 7 to 8) [32]. Based on previous findings, there is a very tight window of splice site usage with a stark transition between MES of 7 to 8 defining the usage of an alternative splice site [32]. Based on these findings, a MES of 8 was used as a cutoff for determining splice site groups. Exons were defined as those that either have a strong 3'ss and a strong 5'ss "strong strong" (SS), "strong weak" (SW), "weak strong", (WS) or "weak weak" (WW) splice site scores (Figure 2.10A). Using these splice site strength groups, significantly longer exon lengths were identified when flanking 3'ss and 5'ss were WW, as compared to the SS splice site group (Figure 2.10B). Additionally, high splice scores correlate with increased exon length conservation (Figure 2.10C) and nucleotide sequence conservation (Figure 2.10D). These observations suggest that the exon/intron architecture modulates the required pattern of splice site strengths across an exon. On average, exons with WW splice sites require longer lengths to aid in efficient recognition, presumably through increasing trans-acting factor binding. The direct correlation between length and sequence conservation is likely to reflect the importance of conserving splice site sequences for effective recognition and splicing of *internal exons*.

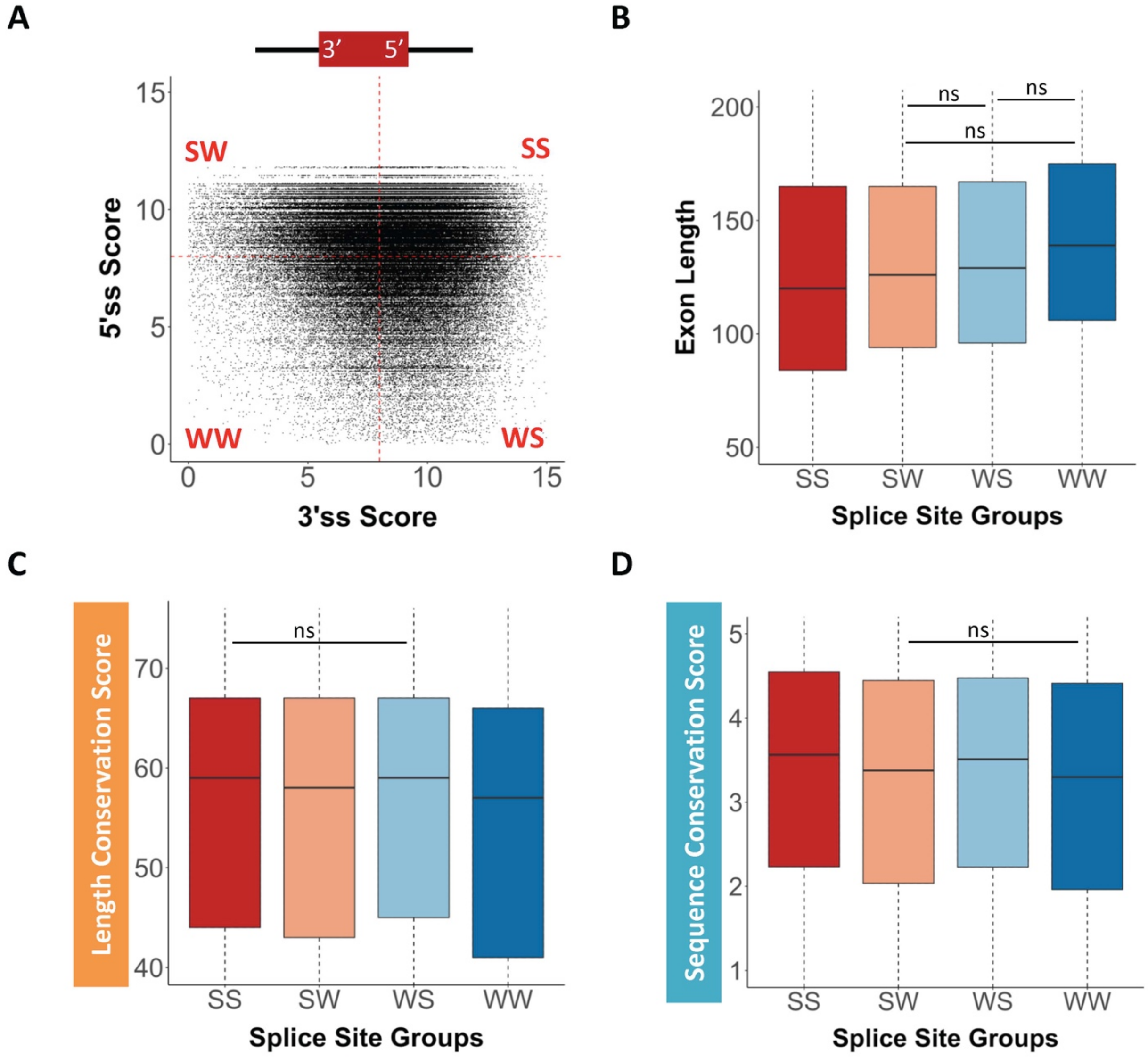


Figure 2.10. Distribution of Splice Site Scores Across Internal Exons.
A) Correlation between 5'ss and 3'ss scores of internal exons. Each quadrant is defined by the 5'ss and 3'ss score respectively: SW (short weak), SS (strong strong), WW (weak weak), and WS (weak strong). Average splice site score is delineated by red dotted vertical and horizontal lines. The four splice site groups were correlated with B) exon length E) length conservation score and E) sequence conservation score. All intron length groups have p > 0.05 unless marked by "ns" – not significant.

Using an exon conservation database to predict splicing from ~5000 disease-associated SNPs.

Previous studies have suggested that about 20% of disease-associated alleles alter splicing [144]. Recently, the splicing outcome of 5,132 disease-associated SNPs was reported using an exon trap model [145]. The authors generated wild-type (WT) and SNP versions of each analyzed exon and tested their effects on exon inclusion using a reporter assay. The published dataset provided the opportunity to test whether the exon size-filtered conservation approach described above could identify SNPs that could induce splicing changes. Given the discussion about the coevolution of coding and splicing features, such predictions would be even more significant for SNPs located at the wobble position. Predictively, high nucleotide conservation at wobble positions of size-conserved exons are most likely involved in mediating efficient splicing.

Using the splicing SNP dataset [145] as a starting point allowed for testing whether exon size-filtered sequence alignments could be used as a method for splicing prediction. The distribution of the data within the SNP database can be divided into two groups, SNPs that occur at the exon/intron boundary (or junction), and those that occur within the exon (non-junction). The junction and non-junction data can be further sub-divided by the location of the SNP within the context of a codon to differentiate between SNPs that lead to protein coding defects or SNPs that occur within the wobble position. Most of the 5,132 disease-associated SNPs analyzed represent nucleotide changes that lead to amino acid changes, given that the majority of SNPs evaluated occur within position one and two of the codon (Figure 2.11A). Furthermore, 796 SNPs are located at exon/intron junctions (as defined by U1 and U5 snRNP binding), with 367 SNPs located at the 5’ss and 429 located at the 3’ss. Of the 367 SNPs at the 5’ss, 283 alter splicing at the 5’ss - 236 lead to exon exclusion and 47 lead to no change or inclusion of the exon, essentially as anticipated from differences in splice site strength calculations (Figure 2.11B). Of the 429 SNPs at the 3'ss, 258 SNPs alter splicing - 164 SNPs lead to exon exclusion and 94 lead to no change or inclusion of the exon (Figure 2.11C), demonstrating that the majority of SNP cases at the junction negatively impact the 5'ss and 3'ss scores leading to less exon inclusion. This observation confirms that nucleotide alterations at splice sites impact splicing efficiency through changes in the complementarity between the pre-mRNA and spliceosomal factors. More importantly, this junction analysis demonstrates that the SNP splicing data generated through exon trap experimentation faithfully follows expectations that a mutation leading to altered splice site strengths results in altered exon inclusion levels.

To predict the impact a non-junction SNP has on exon splicing, the sequence conservation of each SNP was evaluated within an upstream and downstream 5nt window. Each SNP was then categorized depending on its reading frame location (first, second or wobble position) and whether the SNP was ever observed in other species with size-conserved exons. A SNP was defined as a *'mutation not important'* if that SNP was present alone in other exon size-conserved species, but no other nucleotide changes were seen within the flanking 5nts. These SNPs are considered not important for splicing due to the fact that they are seen within size-conserved exons in other species. The second category of SNPs was defined as *'mutation not observed'*, where a SNP is never seen across all the exon size-conserved species. Such nucleotide invariability suggests that the nucleotide identity at the SNP position is evolutionarily important and could be essential for pre-mRNA splicing.

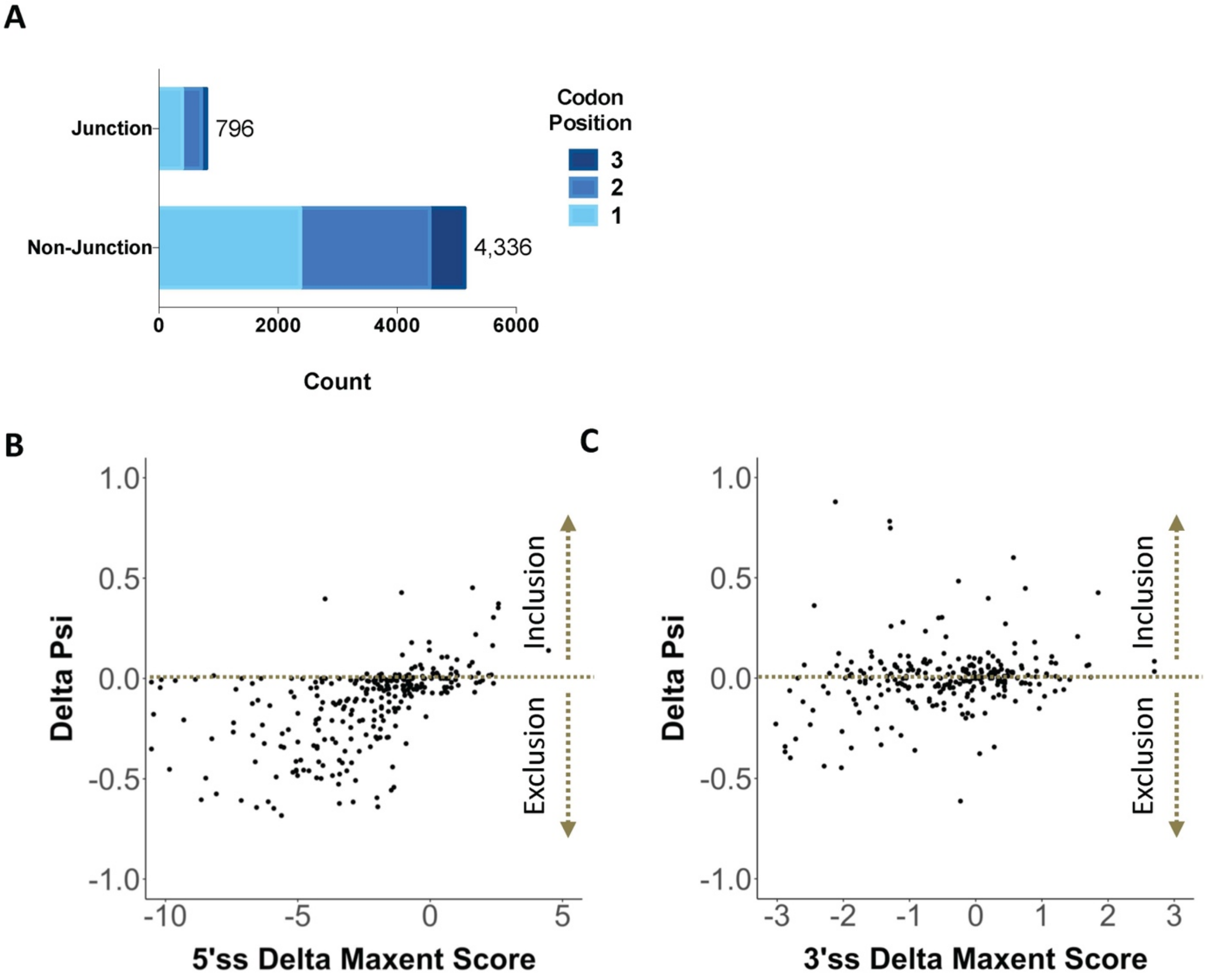


Figure 2.11. SNP Data Representation Within Exon Conservation Database.
A) representation of SNPs relative to codon position and location within an exon at junction and non-junction sites. Distribution of the delta maxent score vs delta psi for B) 5'ss and C) 3'ss.

The third category of SNPs are those that have *'SNPs with covariance'*. For this category, a single alteration was only seen in exon size-conserved species when other nucleotide variations were present within 5nts upstream or downstream of that SNP. These additional nucleotide changes could be important for splicing, as they may act to compensate for defects caused by a single mutation [94]. Of the 5,132 SNPs that were length-conserved, 4,336 were located at non-junction sites. Of these non-junction SNPs, 724 SNPs changed exon inclusion by more than 10%. Of those SNPs, 431 SNPs had a negative dpsi (dpsi-) value, indicating that the SNP reduced exon inclusion, and 293 had positive dpsi (dpsi+) values indicating that the SNP increased exon inclusion. As a control, SNPs that resulted in exon inclusion changes of less than 0.2% were used as a "no change" control group (dpsiC of 319 SNPs). For each of these splice effect groups (less inclusion, more inclusion, no change) the codon position of the queried SNP was determined (1st, 2nd, or wobble position) before each SNP was categorized at the evolutionary level as defined above. The major evolutionary category of SNPs at non-junction sites, regardless of codon position, was the *'mutation not observed'* category, suggesting that these SNP positions harbor overlapping splicing and coding pressures (Figure 2.12A-2.12C). The wobble position revealed the greatest variability between the three evolutionary SNP categories, with a higher percentage of *'mutation not important'* and *'SNPs with covariance'* categories when compared to their observed frequency at codon positions one and two.

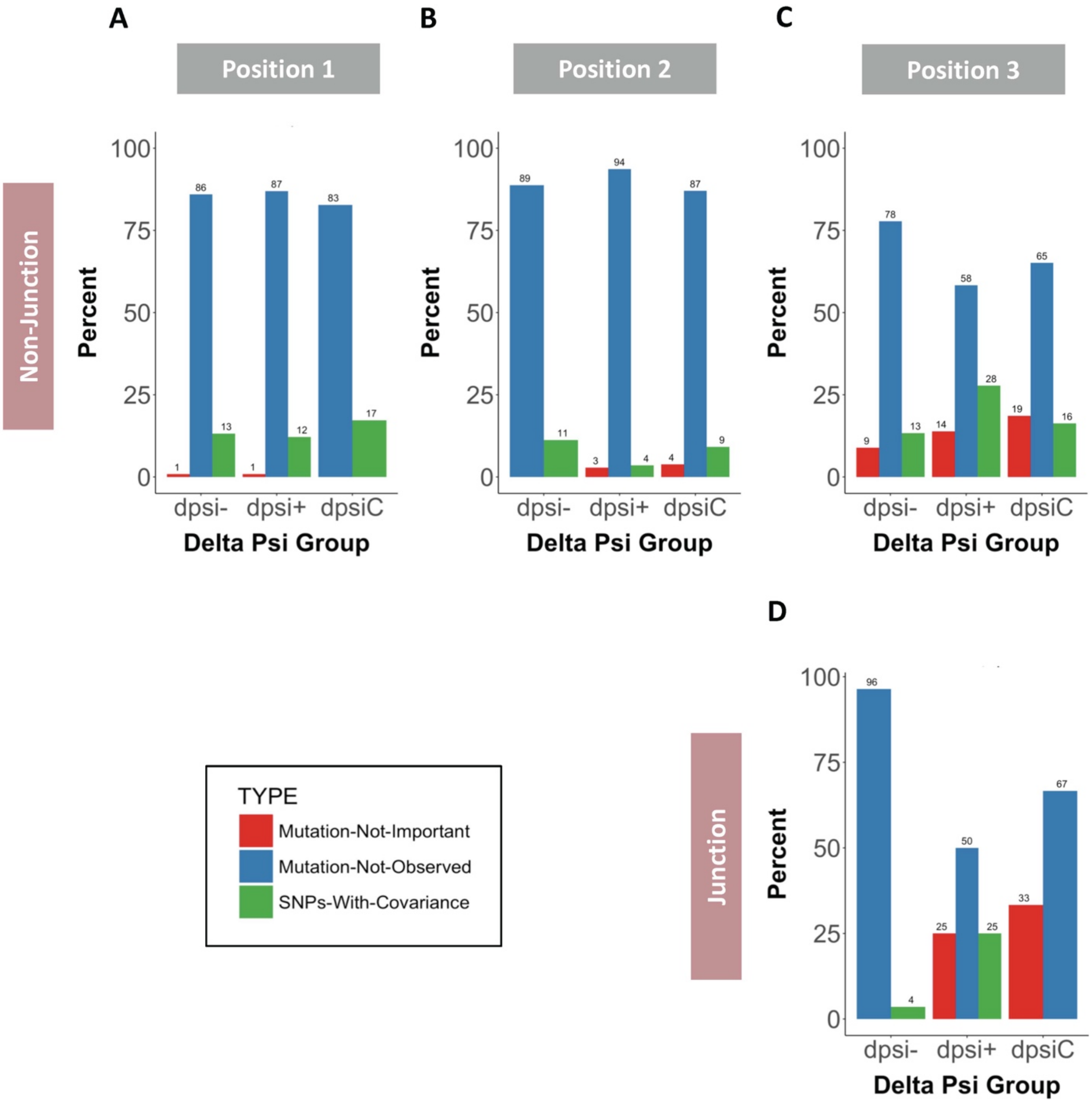


Figure 2.12. SNP Categories Relative to Codon Position.
SNP exonic positions at non-junction or junction sites compared across codon positions for each SNP “type”. Delta psi (dpsi) groups (-, + and C) represented on x-axis and y-axis depicts percent of each ”type” of SNP as a function of the total number of SNP’s for each designated dpsi group.

To test whether splice changing SNP positions are enriched for any of the three evolutionary categories, their relative representation was compared between those SNPs that reduce exon inclusion (dpsi-) and the control group (dpsiC) or those SNPs that increase exon inclusion (dpsi+) and the control group (dpsiC). SNP positions at the non-junction wobble site that reduce exon inclusion (dpsi-) were enriched for the '*mutation not observed*' category and selected against the '*mutation not important*' group (Figure 2.12C). These observations are consistent with the notion that splice altering nucleotide changes are selected against, especially when the nucleotide change results in exon skipping. Interestingly, the conservation features for SNP positions that result in increased exon inclusion are quite different. The most prominent enrichment is seen for the '*SNPs with covariance'* category (Figure 2.12C). It is possible that this category represents events of local compensatory nucleotide changes that re-establish efficient splicing (Figure 2.12C). Qualitatively identical, but quantitatively more striking selection trends are observed for junction SNPs at wobble positions (Figure 2.12D), reinforcing the interpretation that splice altering SNPs are identifiable using the exon size filtered phylogenetic conservation approach.

Using the published Fairbrother dataset, [145] it was possible to determine whether interrogated SNPs change the amino acid sequence, whether they create premature stop codons (PTC), or whether they are synonymous nucleotide alterations. The first unanticipated observation regarding the disease-associated SNP dataset was the fact that wobble position entries are underrepresented (Figure 2.13A), even for those SNPs that induce splicing changes of >10% (Figure 2.13B). Thus, it appears that disease-associated SNPs at wobble positions that alter splicing are selected against in the dataset.

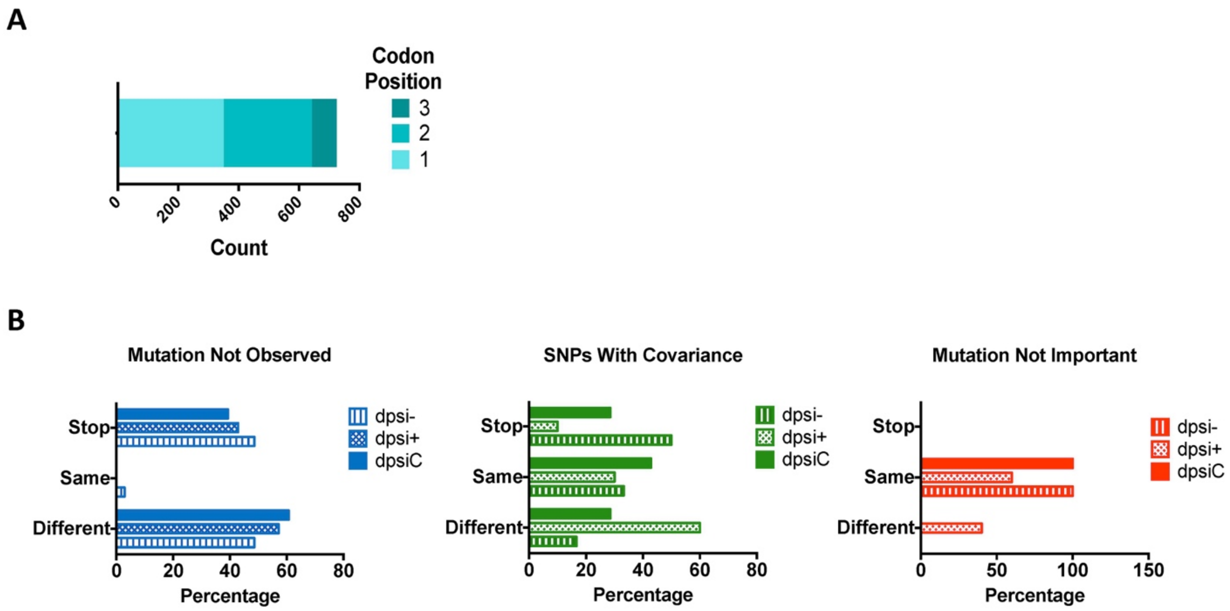


Figure 2.13. Amino Acid Change for Each SNP Category at Non-Junction Sites.
A) Count of SNPs that influence splicing at each codon position for non-junction sites with a 10% dpsi cutoff. B) Percentage of amino acids changing at the wobble position that lead to more or less exon inclusion as designated by a 10% dpsi cutoff. 'Stop' refers to changes in an amino acid sequence causing generation of a stop codon. 'Different' denotes sequences that lead to an amino acid change and 'Same' demotes sequences that do not lead to an amino acid change.

Further investigations into the potential functional consequence of the evaluated SNPs highlight unexpected trends (Figure 2.13B). Many of the *'mutation not observed'* SNPs either induce a PTC, or they change the encoded amino acid. Interestingly, SNPs that reduce exon inclusion display a greater proportion of stop codons that create nucleotide changes, suggesting that the loss of splicing could represent a molecular mechanism to avoid the incorporation of a PTC containing exon. This trend is also observed in the *'SNPs with covariance'* group. Remarkably, SNPs that increase exon inclusion in that group display an inverse relationship, strengthening the notion that splicing assists in the avoidance of PTC containing exons. In summary, the use of exon size-filtered sequence alignments allows for improved identification of splice-altering SNPs and, moreover, it assists in predicting the direction of splicing changes.

Discussion

The hypothesis that exon size-filtered species alignments improve the identification of nucleotides evolved to mediate efficient exon ligation was tested. The generation of the exon conservation database allowed for identification of fundamentally important architectural parameters of the human genome. The majority of *internal exons* are between 50-250nts long and most exons harbor strong 5'ss and 3'ss scores. It is important to note that some of the challenges of generating this exon conservation database were based on inconsistencies and errors within the annotation files of the species used for comparison. Initially, 100 species were used, however, due to poor genome annotations, 24 species were removed, leaving 76 species for analysis.

Displaying exon size conservation across 76 species allowed for the identification of two diverse exon populations (Figure 2.4A). One population of high length-conserved exons were mainly represented by *internal exons* of length 50-250nts. This represented the largest distribution of all *internal exons* within the human genome. A second population of low length-conserved exons is mainly represented by *last exons* that were longer in length (exons >250nts) (data not shown). It is possible that these exons may be evolutionarily young or have recently emerged, as they are highly conserved only between human and primates such as chimps, known to be our closest living ancestor [146]. In addition, longer length *internal exons* that make up a smaller population of exons compared to *last exons* are typically characterized by exon definition (more are flanked by longer introns compared to those exons that undergo intron definition) (data not shown). These findings demonstrate that optimal exon size is an evolutionarily conserved feature and *last exons* that are longer in length are generally poorly size conserved and may have significantly evolved only during the divergence from prosimians to primates.

A strong correlation between exon length and exon sequence conservation was identified (Figure 2.4), suggesting that exons have evolved to optimize sequence and length, presumably to satisfy coding pressures and splicing pressures alike. Interestingly, there are varying degrees of sequence and length conservation for each exon within a gene. One exon in a gene may have very high sequence and size conservation, while the flanking downstream exon may have very low sequence and size conservation. For example, CPXM1 is a gene that encodes a member of the membrane bound carboxypeptidase family, which are important for cell-cell interactions. Exon two of CPXM1 has a low PhyloP score of 0.53 and a very low length conservation score of 1. However, within the same gene, exon 12 has a high PhyloP score of 4.39 and a high length conservation score of 66. Another example is the gene CPSF, an important protein component of the polyadenylation machinery. Exon two has a high

PhyloP score of 4.3 but a low length conservation score of 10, whereas exon three has a low PhyloP score of 0.62 and a moderately high length conservation score of 39. This demonstrates great variation between sequence and length conservation of exons within the same gene. These and other comparable examples demonstrate that exons, rather than entire genes, are units of evolutionary selection, consistent with the idea that split genes increase the diversification potential of species.

Using an arbitrary cutoff of 10 for the number of species that demonstrated length conservation for a certain subset of exons allowed for the identification of species that display the closest exon size distribution to humans (Figure 2.6). This analysis revealed that within the 76 species used here, chimp is the most closely related species, as expected, followed by gorilla, gibbon, macaque, orangutan, baboon, squirrel monkey, and marmoset, thus faithfully re-tracing the evolutionary tree. These results suggest that exon size conservation is an important genome feature that contributes to overall similarities between species. Furthermore, it is possible that the gain of new exons can be traced through the exon size conservation database.

Understanding whether flanking intron lengths (Figure 2.7) or splice site score groupings (Figure 2.9) show any variation around exons with high length and sequence conservation revealed interesting features. Exons flanked by longer introns showed greater sequence and length conservation. It is known that splice-site recognition through exon definition is less efficient [62], so LL exons must have stronger splice sites and more optimal exon sizes to maintain efficient exon inclusion. Indeed, this increase in splice site strength is observed for LL exons (Figure 2.8C). Using similar arguments, exons flanked by short introns will be included more efficiently, so stronger splice site scores and more optimal exon lengths are less important exon features (Figure 2.8). These comparisons illustrate different ways to maintain important splicing pressures that help to define the exon/intron architecture of genes.

As for splice site groups, a significant trend is seen between SS splice sites and WW splice site exons. The stronger the average splice site score, the greater the length and sequence conservation. This observation is consistent with the notion that weaker splice sites (demonstrating less sequence and length conservation) undergo more alternative splicing [13, 15]. Thus, the exon conservation database can also be utilized to predict exons that display a high probability of undergoing alternative splicing. It has been shown that there is a higher incidence of alternative splicing in higher eukaryotes than lower eukaryotes, demonstrating the evolutionary lineage of alternative splicing within the phylogenetic tree [75, 76]. Several studies have also shown that alternative splicing increases from invertebrates to vertebrate species, suggesting that alternative splicing could be an evolutionary driving force for the generation of new exons [76]. From our observation, a consistent trend between length and sequence conservation is seen, where high length conservation for one group also demonstrates high sequence conservation for the other, and vice versa. Therefore, it is possible that these two conservation measures play a convergent role in the evolution of an exon. Exons that have been optimized for size and sequence will demonstrate high levels of conservation across species, whereas those that have newly emerged, through the process of alternative splicing, will select for important splicing parameters, eventually leading to high length and sequence conservation.

SNPs are the most abundant type of variation within DNA sequences [147], and many disease-associated SNPs have been suggested to play a role in splicing [144, 148].

Understanding the functional impact of SNPs in human diversity and disease and their influence on splicing may be key to understanding and improving treatments for various diseases. Using the Fairbrother SNP splicing dataset [145] and the exon conservation database, it was possible to examine whether the method of exon size-filtered alignments could be used as a splicing predictor. Although the majority of the Fairbrother SNP dataset represents protein coding mutations (Figure 2.11A), the use of this data allowed for interrogation of the impact that non-junction and junction SNPs have on exon inclusion. The greatest change on splice site scores was observed when a mutation was located at the 5'ss or 3'ss (Figure 2.11B), consistent with the notion that mutations within splice site sequences lead to splicing defects, further confirming the accuracy of the dataset being used. Although it is widely accepted that changes at splice sites from the canonical splice site sequence impacts splicing, it is not known how mutations at non-junction sites influence splicing and whether such splicing changes can be predicted. Using the SNP categories *'mutation not important', 'mutation not observed'*, and '*SNPs with covariance*', it was possible to differentiate between mutations that are more likely to have an impact on splicing and those that do not. Codon position played an important role in whether SNPs were allowable. The wobble position of the codon, known for allowable changes to the final designation of an amino acid, demonstrated the greatest variability between the three SNP categories (Figure 2.12). Compared to the control group, a greater percentage of SNPs at the wobble position represent nucleotide changes that are phylogenetically not observed across the exon size-conserved comparator species. These SNPs could impact splicing, in particular for SNPs that decrease exon inclusion, demonstrating the importance of these nucleotides. The next largest category of SNPs seen at the wobble position were those only seen when other SNPs were present – '*SNPs with covariance*'. Of these SNPs, those that increased exon inclusion signified a larger percentage than the control. This observation demonstrates the important nature of compensatory mutations in rescuing any defects that could be generated by the presence of isolate SNPs. For instance, single mutations at a regulatory binding site may prevent binding of regulatory proteins [149]. However, other mutations within 6nts upstream or downstream of that SNP may rescue the binding of that regulatory protein. It is also possible there may be synergistic or additive effects, where more than one SNP enhances the binding of a regulatory protein [94], increasing exon inclusion and improving splicing of that exon. From the data, an enrichment of cases is seen where a wobble position SNP, which creates a PTC, induces exon skipping. It is possible that preventing exon inclusion in these cases constitutes a compromise to rescue some partial function of the encoded protein, at the cost of losing an internal exon. Thus, the enrichment of exon skipping observed for *'mutation not observed'* SNPs is a path to avoid including PTC containing exons in the final transcripts.

Irrespective of the limitation of the evaluated SNP dataset, our approach increases the prospect of identifying splice altering SNPs at wobble positions. This analysis demonstrates that using exon size-filtered sequence alignments at the wobble position improves the predictive power of whether a SNP will influence exon inclusion.

Methods

Generation of Exon Conservation Database Using Ultra-Conserved Exons

Orthologous genes in different species vary in sequence, length, and the number of exons per gene. To find corresponding exons of the same gene in two or more species, it is necessary to compare the sequence of all possibly matching exons and not to rely solely on the exon number or its relative position in that gene.

To generate a database of human exons that contains information pertaining to exons that are conserved in both sequence and length, multiple sequence alignment (multiz100ways) of 99 species compared to human exons was used. This multiple alignment file was generated by the Multiz tool [150] downloaded from the UCSC Genome Browser. For each human exon, the genomic location of aligned sequences from other species to the sequence of that exon were extracted from this multiple alignment data. In each species, if that location overlapped with an annotated exon (covering at least 20% of the exon), that exon was listed as a matching exon to the human exon. If an exon matched with a similar sequence in another species but also maintained its structure (length difference ≤ 3nts between the human exon and the matching species exon), that exon was defined as ultra-conserved in that species. This generated a list of matching exons from all species with exon length information stored within the database. Some of the assemblies used in multiz100ways were older and the associated gene annotations were not found or trusted, therefore, the process was continued with only 76 species, including human. Additionally, only the canonical version of each gene was used from the human annotation (hg19 RefSeq).

Exon conservation database parameters. For each exon in the human genome listed in the exon conservation database, an "Ultra-In" number was generated by comparing the length of that exon and the matching exons in other species. This number, denoted as a score, represents the number of species in which a particular exon is ultra-conserved for length.

Basic characteristics of each exon were also listed in the exon conservation database. This included the following: genomic coordinates, exon length, relative exonic position in the transcript (first, internal, last, single), and upstream and downstream intron lengths (as extracted from the hg19 gene annotation). Additionally, sequence conservation scores and 5' and 3' splice site scores were also included. The sequence conservation score was obtained through PhyloP [140], generated by averaging the nucleotide-based values already available for multiz100way alignment, as phylop100way tracks on the UCSC Genome Browser. Splice site scores were obtained from MaxEnt scores [26] for 3' and 5' splice sites. The corresponding genomic sequence was extracted for each end of the exon and the scores were calculated using the web interface of MaxEntScan. In addition, 233 exons were filtered out due to poor annotations within the human genome itself. It is important to note that using PhyloP as a measure of sequence conservation has its limitations. PhyloP currently looks at base-wise conservation across 46 species only, whereas our size conservation score used in the exon conservation database is generated across 76 species. Future developments in PhyloP will aid in better comparisons, leading to a larger pool of species from which to compare and potentially tighter correlation between length and sequence conservation

Exonic SNP Conservation Database

Using the exon conservation database, those exons that included one or more mutations as catalogued [145] were separated and further analyzed to derive a second database: SNP conservation database. For each one of the SNP mutations, parameters within the database included the following: position of the mutation within the human exon as well as five nucleotides upstream and downstream of that position, "Ultra-In" score (from the multiple alignment data), the associated amino acid sequence for the wildtype and mutant SNP, and the position of that SNP within the context of a codon. Additionally, if the mutation was located at a junction, defined as a position located within 3nts of splice sites, the splice site scores (MaxEnt scores) were generated for both the wild-type and mutated version of 3' and 5' splice sites.

SNP Conservation Database Mutation Categories. Mutations within the SNP database were categorized into three groups based on their occurrences in the "Ultra-In" species and any variations observed in the neighboring nucleotides. The first category was defined as SNPs that are *'mutation not observed'*. These SNPs include mutations listed in [145] that did not occur in any "Ultra-In" species. In the second category, SNPs were defined as *'mutation not important'* when a SNP was observed in a subset of "Ultra-In" species but the neighboring nucleotides did not show any variation between species. This mutation was present in the "Ultra-In" species, however, no other mutations were nearby and the SNP containing exon still had the same length as the human exon. Hence the mutation did not influence splicing and this mutation was not important for splicing. The third group, *'SNPs with covariance'*, represents cases in which the annotated mutation was observed in some species, but at least one other variation was also seen in the ten nucleotides around that mutation in a subset of those species. The combination of this mutation and these nearby variations may not affect the splicing pattern; therefore, it is possible that these variations had a compensatory effect on the main mutation.

If the aligned sequence from other species compared to the eleven nucleotides (mutation and surrounding 10nt) in human contained three or more indels or "N", they were not considered in the final mutation categorization. Though the exons were already matched by sequence similarity, the difference that was not significant when comparing the whole exon could cover a noticeable portion of these 11 nucleotides. If the mutation was located close to a splicing junction, a sequence of “Z” was used to show the border of the exon in the codon containing the mutation and “X” was used in the associated amino acid column.

Delta maxent score was calculated for either the 5’ss or 3’ss as: MT maxent score – WT maxent score. The percent spliced in (psi) value was calculated as the ratio of Spliced/(Unspliced + Spliced). The delta psi (dpsi) value was calculated as: MT psi – WT psi. For all analysis, a 10% dpsi cutoff was used to generate the dpsi- and dpsi+ categories. For the control group, dpsiC, a 0.2% cutoff was used. This cutoff variable was necessary since a cutoff of 0 did not have sufficient data points.

Bedtools v2.25.0 was used to find the overlap between genomic regions when necessary in finding matching exons. Binary files, wigToBigwig and bigWigSummary, downloaded from the UCSC Genome Browser were used in conservation score calculation. All the other scripts were written in Python 2.7 and all analysis and graphs were generated using R v3.3.2.

CHAPTER 3

Splicing Repression Results in Changes to U1 SnRNP Complex Integrity at the 5' Splice Site

Summary

Previous categorization of the function of SR proteins and hnRNPs in splicing regulation was subject to a single role for each protein. SR proteins were commonly accepted as activators of splicing and hnRNPs were known repressors. However, the traditional classification of these regulatory proteins has been questioned by findings that demonstrated role reversal for both these protein families: SR proteins can also repress and hnRNPs can also activate. Furthermore, it was shown that repressor SR proteins and hnRNPs are able to recruit U1 snRNP to the 5' splice site; however, in this repressed state, U1 snRNP appeared to exist as a dead-end splicing complex. To understand the mechanism by which this dead-end complex is formed and maintained, the hypothesis that a repressor SR protein/hnRNP alters the affinity of U1 snRNP to the 5' splice site of the pre-mRNA was tested.

RNA competition assays were carried out to measure changes in the affinity of U1 snRNP components to the pre-mRNA. The findings suggest that U1 snRNP components change their strength of association within the complex, and that U1 snRNP integrity is modulated by position-dependent interactions with splicing activators and repressors.

Introduction

Eukaryotic gene expression is dependent on the correct removal of introns and the joining of exons in a process coordinated by the spliceosome called splicing. Splicing occurs in a step-wise fashion, forming four distinct complexes that differ in composition and order of appearance, E → A → B → C. These steps are catalyzed by the major spliceosome, which is composed of U1, U2, U4, U5, and U6 small nuclear ribonucleoproteins (snRNPs), and ~300 accessory proteins [3]. The first step in this assembly reaction is initiated by U1 snRNP. U1 snRNA acts as a scaffold to which 10 protein subunits bind. Seven of the proteins are common to all snRNPs and are called Sm proteins (Sm-B, E, F, G, D1, D2 and D3) and the remaining four proteins are unique to U1 snRNP: U1-70K, U1-C, and U1-A. The formation of E complex occurs when U1 snRNP binds the 5' splice site (ss) and U2AF binds the 3'ss. The 5'ss is defined by nine highly conserved nucleotides (nts) and the 3'ss is defined by two other distinct elements, the polypyrimidine tract and branch point sequence, located 40nts upstream [151]. Regulation of alternative splicing (AS) requires many trans-acting factors, such as SR proteins and the hnRNP family of splicing regulatory proteins, but also cis-acting elements, such as splice-site strength and binding sites for splicing regulatory proteins termed splicing regulatory elements (SREs) [23]. SREs, referred to as exonic and intronic splicing enhancer or exonic and intronic splicing silencer sequences, are binding sites for SR proteins and hnRNP splicing factors. When bound to such elements, these splicing regulators are believed to increase or decrease the recruitment of spliceosomal components [137, 138], however, detailed mechanisms by which SR proteins/hnRNPs regulate splicing are currently not well understood.

U1 snRNP is one of the most abundant snRNPs in eukaryotes. It has been shown that U1 snRNA participates in many polyadenylation and splicing-related activities. For instance, the transcriptome contains a multitude of pseudo poly(A) and pseudo splice sites, many of which are bound by U1 snRNA [124, 152]. This extensive U1 snRNP binding to the pre-mRNA necessitates mechanisms whereby spliceosomal assembly is only activated at exonic locations. Thus, the spliceosome must be able to differentiate between U1 snRNPs located at functional splice sites from U1 snRNPs bound to pseudo sites. Presumably, the exonic environment of U1 snRNP, which includes interactions with splicing factors such as SR proteins or hnRNPs, may be the key to spliceosomal activation. However, mechanisms that allow distinction between splicing and non-canonical U1 snRNP functions are not known.

Previous categorization of the function of SR proteins and hnRNPs in splicing regulation was subject to a solitary role for each protein that was unyielding to variation. SR proteins were commonly accepted as activators of splicing and hnRNPs were known repressors, however, the classical definition of these regulatory proteins has been challenged through findings that demonstrate dual roles or role reversal for both these proteins; SR proteins can repress and hnRNPs can activate [44–49, 55]. SR proteins have been well documented to play a role in activating splice site selection, for instance, exonically by forming a bridge between components bound to the 5'ss and 3'ss [34, 153]. However, SR proteins have also shown a role in repression of the 3’ss by binding to the branch point sequence and preventing recruitment of the U2 snRNP [33, 50]. TIA-1, hnRNP F, and hnRNP H have also been shown to mediate 5’ss activation from a downstream intronic position in several pre-mRNAs [52–54]. Recent findings have demonstrated dual functionality in the behavior of SR proteins and hnRNPs [31, 33, 48–51, 55, 137, 153, 154]. Almost all of the common splicing regulatory proteins tested can activate a regulated 5’ss depending on their location; SR proteins (such as SRSF7) always activate from an exonic location, hnRNPs (such as TIA-1, an hnRNP-like protein) activate from an intronic location [57]. Interestingly, both classes of splicing regulators can also inhibit splicing. A model was generated whereby SR proteins repress splicing when located intronically and hnRNPs repress splicing when located exonically (Figure 3.1).

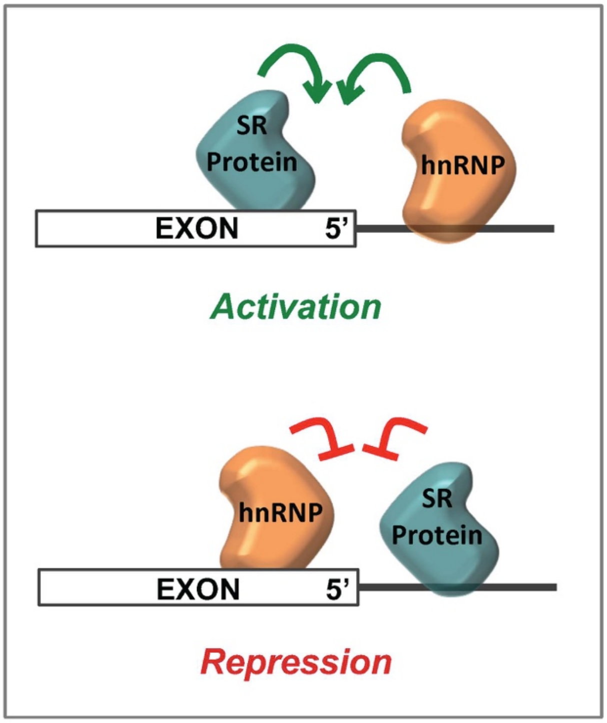


Figure 3.1. Model of Position-Dependent Regulation.
SR proteins and hnRNPs demonstrate similar regulatory behavior, but in mirror image of each other, dependent on the location of their SRE.

Unexpectedly, the repressive activities of SRSF7 and TIA-1 were shown to fully support the formation of E (early or commitment) complex (Figure 3.2B and [57]), however, higher order splicing complexes (A, B, C) were not present [57]. E-complex chase experiments demonstrated the formation of a dead-end splicing complex [57]. This observation led to the consideration that in the presence of a repressor, U1 snRNP may undergo structural rearrangements, resulting in changes in its affinity for the pre-mRNA or in the composition of bound U1 snRNP. Collectively, these findings demonstrate that splicing regulatory proteins behave in a highly position-dependent manner and the position of SREs is key to the activation or repression of the splicing mechanism.

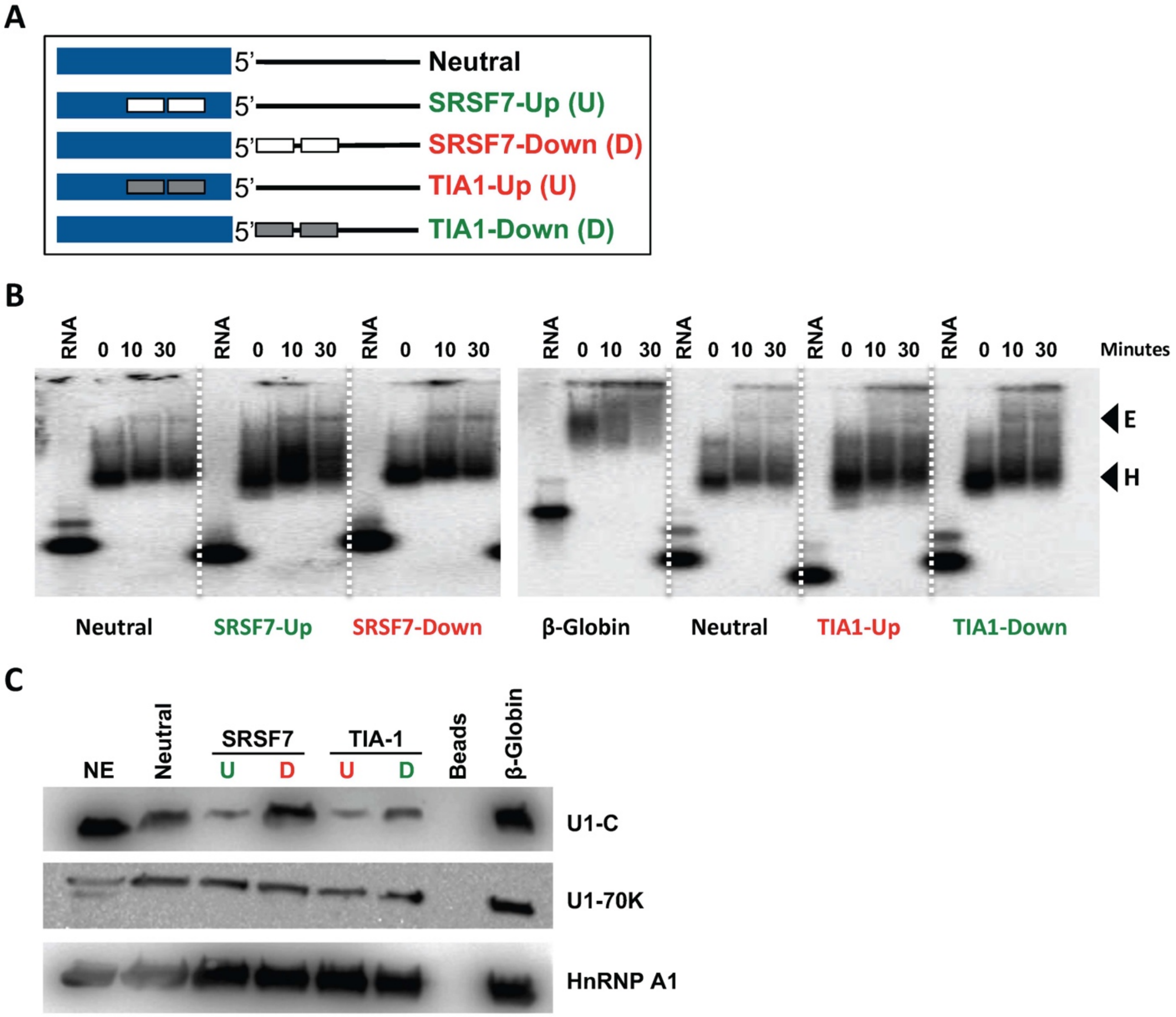


Figure 3.2. Half-Substrate RNA E Complex Formation.

A) Schematic of half-substrate RNA constructs containing binding sites for SRSF7 or TIA-1 (denoted as white or gray boxes) used for experiments. Blue boxes are exons, black lines are introns. 5' splice site denoted between exon/intron junction. B) Native complex formation was performed on low melt agarose. RNA was added to ATP depleted nuclear extract mix for 0, 10 or 30 minutes. Half-substrate RNA that was not introduced to nuclear extract was loaded onto gel as a control. β-globin RNA was used as positive control. E complex and H complex denoted with black arrows. C) Formation of a dead-end E complex is not caused by lack of U1 snRNP recruitment to either activated or repressed positions. Western blot analysis was performed on half-substrate RNA with antibodies against U1-C and U1-70K; hnRNP A1 used as a loading control. NE denotes 30% nuclear extract loaded onto gel, U denotes upstream, D denotes downstream, bead sample contains no RNA and β-Globin RNA was used as a positive control.

Results

Recruitment of U1 snRNP is not affected by binding of SR proteins/hnRNP in repressive positions.

Based on the model of position-dependent splicing activation and repression (Figure 3.1) experiments were designed to understand why splicing stalls at E complex. *In vitro* transcribed half-substrate RNAs containing binding sites for SRSF7 or TIA-1 (in duplicate) either upstream or downstream of a strong 5'ss (Figure 3.2A) were used, and E complex formation was confirmed for each of the corresponding RNAs tested (Figure 3.2B). Although E complex formation was observed based on gel mobilities, it was not clear whether a functional U1 snRNP was recruited to the 5'ss in the presence of a repressor. To address this question, an RNA pull-down was performed through immobilization of the *in vitro* transcribed half-substrate RNAs to adipic acid dihydrazide agarose beads (Figure 3.2C). Western blot analysis of the pulled down RNA/protein complexes identified the presence of U1 snRNP complex at the 5'ss for both repressive bindings sites of SRSF7 and TIA-1, as demonstrated by the presence of U1-C and U1-70K, components of the U1 snRNP. Repressive SRE positions, therefore, do not interfere with the initial step of spliceosomal assembly but do interfere with further formation of a functional spliceosome and the transition to a catalytically active spliceosome, as was previously demonstrated, at least in the context of the evaluated 5'ss selection.

Splicing repression results in altered affinities of U1 snRNP components at the 5'ss.

To understand the mechanism by which a dead-end splicing complex was formed, the hypothesis that an enhancer or repressor SR protein/hnRNP alters the affinity of U1 snRNP to the pre-mRNA was tested. The Lynch lab has shown that hnRNP L and A1 repress splicing through extended base-pairing of U1 snRNP with the pre-mRNA, causing complex hyper-stabilization [155]. Although the test substrates used in our studies were theoretically not able to make extended base pairing with U1 snRNA, it is possible that the affinity of U1 snRNP changes through different mechanisms.

To test the hypothesis that U1 snRNP has a lower affinity for the 5' splice site in a repressed state compared to a non-repressed state, a dilution-chase reaction was performed using *in vitro* transcribed half-substrate RNAs containing SRE binding sites upstream or downstream of the 5'ss (Figure 3.2A). The goal of these experiments was to reduce the concentration of U1 snRNP by dilution to monitor U1 snRNP dissociation from pre-bound RNAs over time. Observed differences in the measured dissociation rates would then indicate differences in U1 snRNP binding potential. Binding equilibria were disrupted by dilution and U1 snRNP association with each RNA was monitored using RNA pull-down and western blot analysis. When SRSF7 binds in the upstream activating position U1-70K remained stably-bound throughout the time course of the experiment (Figure 3.3A), however, U1-C binding increased over time (Figure 3.3A). Interestingly, when SRSF7 binds the downstream repressive site, an increase in U1-70K association with the RNA is observed upon dilution, but U1-C levels remained constant (Figure 3.3B). The increase in U1-70K association with the repressed RNA and the increase in U1-C in the activated RNA suggests that the dilution may have titrated SRSF7 rather than U1 snRNP into sub-saturating levels, thereby potentially relieving U1 snRNP compositional restraints. While the preliminary dilution experiments did not provide the expected insights into differential U1 snRNP affinities, the data clearly demonstrated that the integrity of U1 snRNP is dynamic, as further demonstrated by changes in the ratio of U1-70K to U1-C (Figure 3.3A, 3.3B).

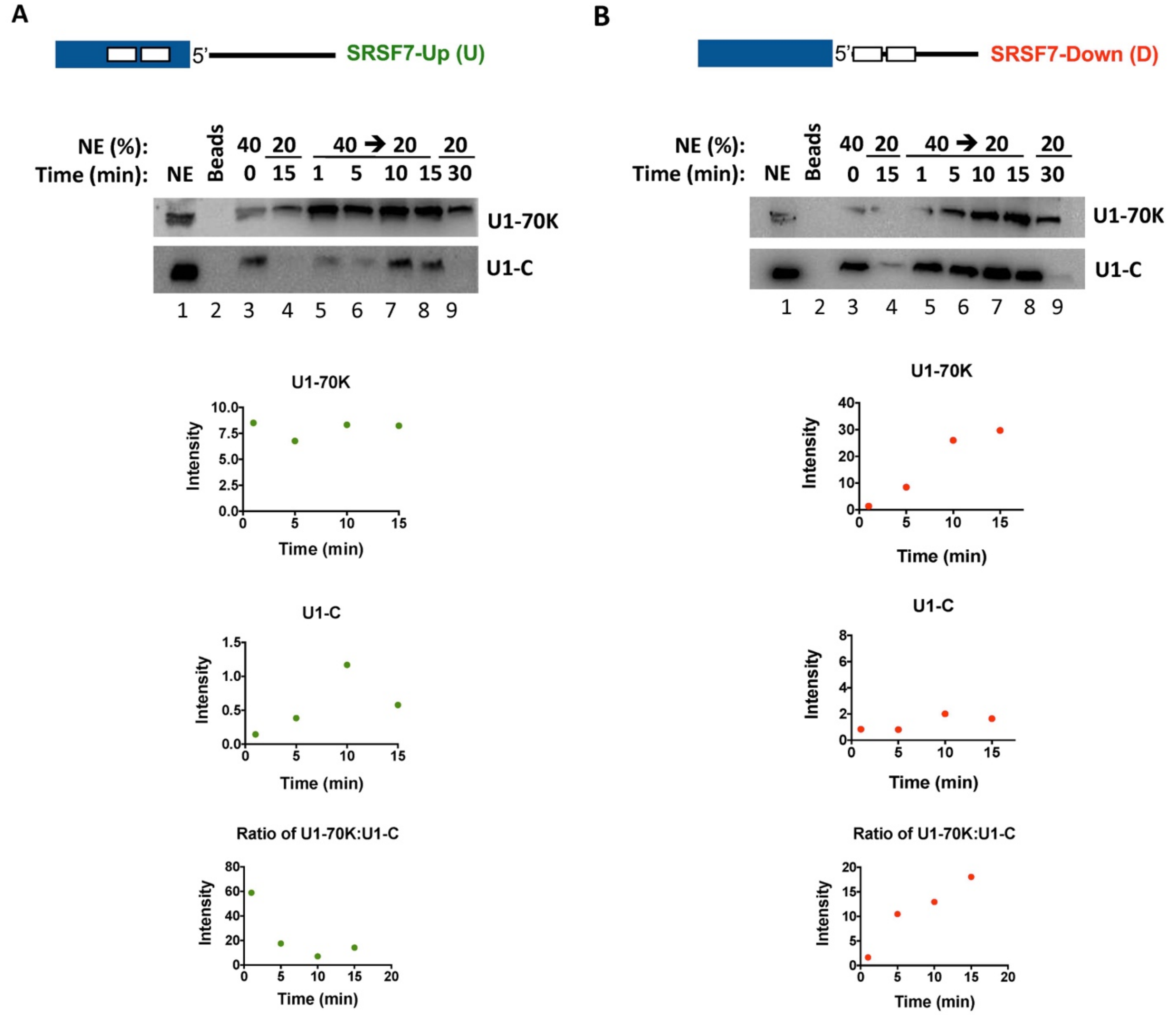


Figure 3.3. Dilution Chase Pull-Down on Half-Substrate RNA.

RNA pull-down followed by dilution chase experiment on A) SRSF7-Up RNA: activating position, B) SRSF7-Down RNA: repressive position. U1-70K and U1-C levels probed via Western Blot analysis. 40 → 20 refers to dilution of sample from 40% NE to 20% NE. Dilution chase controls depicted in lanes 2 and 9, NE denotes 30% nuclear extract loaded onto gel, and bead sample contains no half-substrate RNA. Band intensities of lanes 5-8 normalized to lane 1 and plotted for levels of U1-70K, U1-C and the ratio between U1-70K:U1-C for both A) SRSF7-Up and B) SRSF7-Down.

The goal of the dilution-chase was to measure rates of U1 snRNP dissociation from the test RNA. However, the manner in which the dilution-chase was setup did not allow for accurate binding affinities to be determined. This was due to the lack of knowledge of what was being titrated when the dilution was performed. The concentration of all components was reduced, but there was no way of knowing which RNA binding partner was limiting. Therefore, a different pulse-chase experiment was performed. Instead of diluting the entire reaction, a competition assay was performed through the addition of excess RNA as a chase. Highly concentrated and untagged RNA was added to the binding reaction, thus essentially maintaining the initial concentrations of U1 snRNP and splicing regulators during the chase. Through the use of an RNA that only harbors a 5'ss (chasing U1 snRNP), any detected dissociations of U1 snRNP components were then related to reduced availability of U1 snRNP or SRSF7. This is because the chase RNAs will compete for binding to any dissociated U1 snRNP or SRSF7 over time, preventing re-association with the tagged RNAs of interest. The competitor RNA used for the pulse-chase competition pull-down was a neutral sequence that does not contain binding sites for SRSF7 or TIA-1. This neutral RNA was introduced in high concentrations to compete for U1 snRNP dissociating from the RNA over time, sequestering U1 snRNP and preventing re-association to the RNA of interest. Using this approach, the hypothesis was tested that U1 snRNP has a lower affinity for the regulated 5'ss in the presence of a repressor.

U1 snRNP association with each RNA was monitored via RNA affinity pulse-chase, pull-down, and western blot analysis. Monitoring the binding of U1 snRNP to the neutral sequence over time demonstrated strong binding interactions for U1-70K, with a slow rate of dissociation over time, (0.04 min^{-1}) (Figure 3.4A).

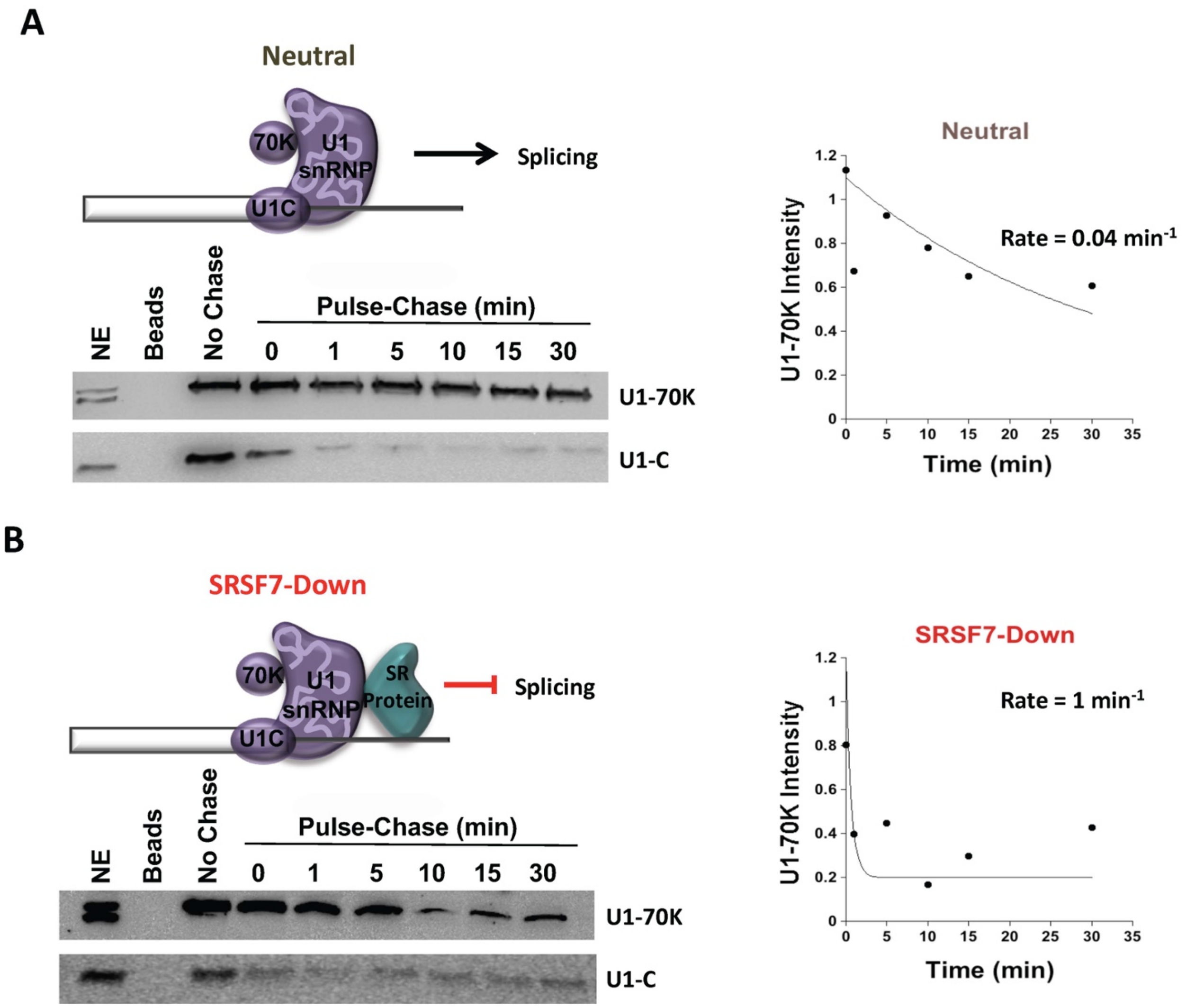
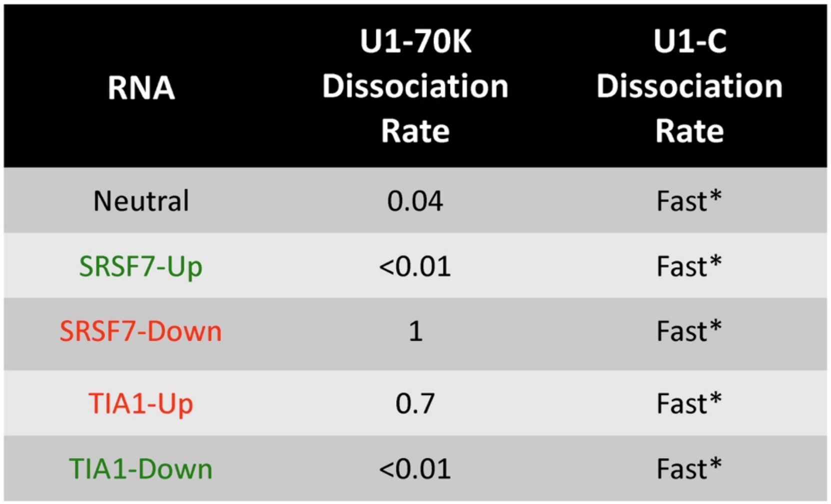


Figure 3.4. Competition Pulse-Chase Pull-Down on Half-Substrate RNA.
RNA Pulse-chase pulldown identifies changes in affinity of U1 snRNP components. A) Western blot analysis was performed on neutral half-substrate RNA or B) SRSF7-down half-substrate RNA, probed for U1-70K and U1-C. Pulse-chase was performed using high concentrations of neutral RNA with no binding site for SRSF7 or TIA-1 for 0-30 minutes. NE denotes 30% nuclear extract loaded onto gel, bead sample contains no half-substrate RNA, 'No Chase' sample denotes no pulse-chase RNA.

The same could not be said of U1-C which seems to dissociate more rapidly, not allowing for the determination of a measurable rate. When SRSF7 binds in the downstream repressive position, U1-70K dissociates quickly over time, with a rate of 1min^{-1} (Figure 3.4B). However, when SRSF7 is bound in the activating position, U1-70K remains stably bound with a slower rate of dissociation (Table 3.1). It is important to point out that the differences in the off rates seen from the pulse-chase pull-down could indicate that what is being measured is in the context of bound U1 snRNA, however, that has not been confirmed as of yet.

The observations from both sets of pulse-chase experiments indicate a general trend when comparing the consequences of activators and repressors. Regulatory proteins bound in activating positions exhibit much faster rates of dissociation and those bound in repressive positions have slower rates of dissociation (Table 3.1). U1-C dissociation rates were harder to calculate as their western signal was much weaker, and it appears that its rate of dissociation is faster than what can be measured through this method. U1-A levels were also probed and U1-A was not consistently present and dissociated quickly as well (data not shown). Additionally, through the pulse-chase experiment, there was a loss of E complex, confirming the results demonstrating the loss of U1 snRNP components (Figure 3.5).

Table 3.1. Dissociation Rates of U1-70K and U1-C.
U1-70K affinity changes for U1 snRNP in the presence of an activator or repressor SR protein/hnRNP. U1-70K rate of dissociation was calculated for all half-substrate RNAs. *Too fast to accurately measure through pipetting.



RNA	U1-70K Dissociation Rate	U1-C Dissociation Rate
Neutral	0.04	Fast*
SRSF7-Up	<0.01	Fast*
SRSF7-Down	1	Fast*
TIA1-Up	0.7	Fast*
TIA1-Down	<0.01	Fast*

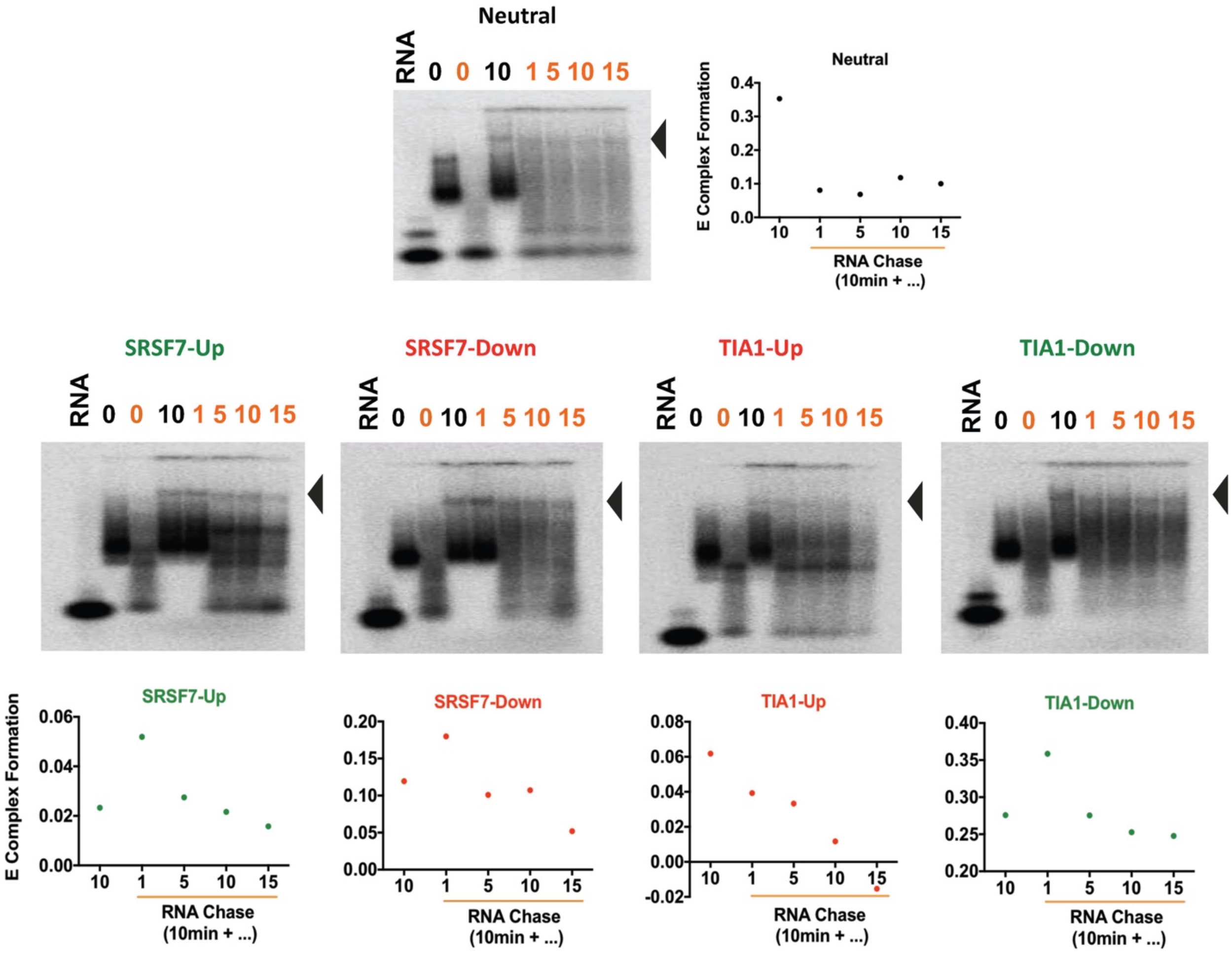


Figure 3.5. Competition Pulse-Chase E Complex Formation on Half-Substrate RNA. Loss of E complex formation during pulse-chase experiments. Native complex formation was performed on half-substrate RNA, pulse-chased with high concentrations of neutral RNA with no binding site for SRSF7 or TIA-1 for 10 minutes plus an additional 0-15 minutes (orange numbers). Non-pulse chase samples were incubated for 0 or 10 minutes (black numbers). Nuclear extract used was ATP depleted. Half-substrate RNA not introduced to nuclear extract was loaded onto gel as a control. E complex denoted with black arrow.

Changes in U1 snRNP complex integrity occur due to the presence of an activator or repressor SR protein/hnRNP, however, there was a concern that the presence or absence of ATP, an essential energy source for the conversion of E to A complex, played a role in U1 snRNP recruitment to the half substrate RNAs. To test the influence of ATP, the pulse-chase pull-down was performed on all half-substrate RNAs in the absence or presence of ATP (Figure 3.6A, 3.6B vs Figure 3.6C). No discernible differences in the levels of U1-70K (Figure 3.6) or U1-C (data not shown as protein intensity was minimal) were detected. However, these pulse-chase experiments were not as reproducible as necessary to finalize strong conclusions (Figure 3.6). In addition, the chase control, RNA 1st' sample, was expected to display reduced levels of U1-70K, as introducing the neutral RNA in excess prior to incubation with NE should have prevented the association of U1 snRNP to the RNA of interest. This, however, was not the case. Many rounds of pull-down and troubleshooting efforts were unsuccessful in confirming the accuracy of this pull-down method. Nevertheless, the data presented thus far is promising as it demonstrates that the composition of U1 snRNP is dynamic and that these U1 snRNP integrity alterations are mediated through repressive or activating SREs.

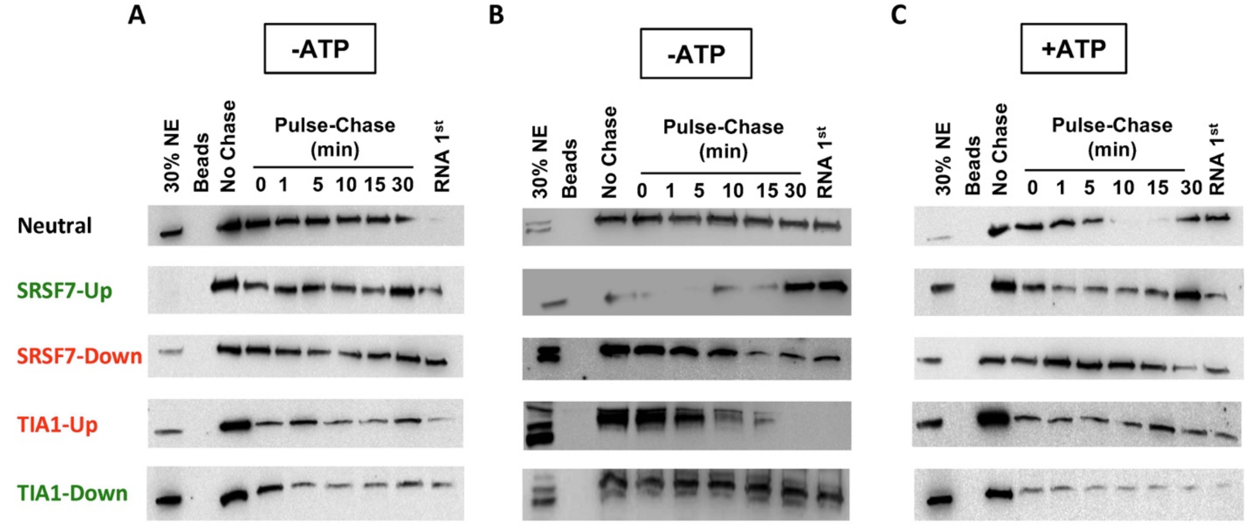
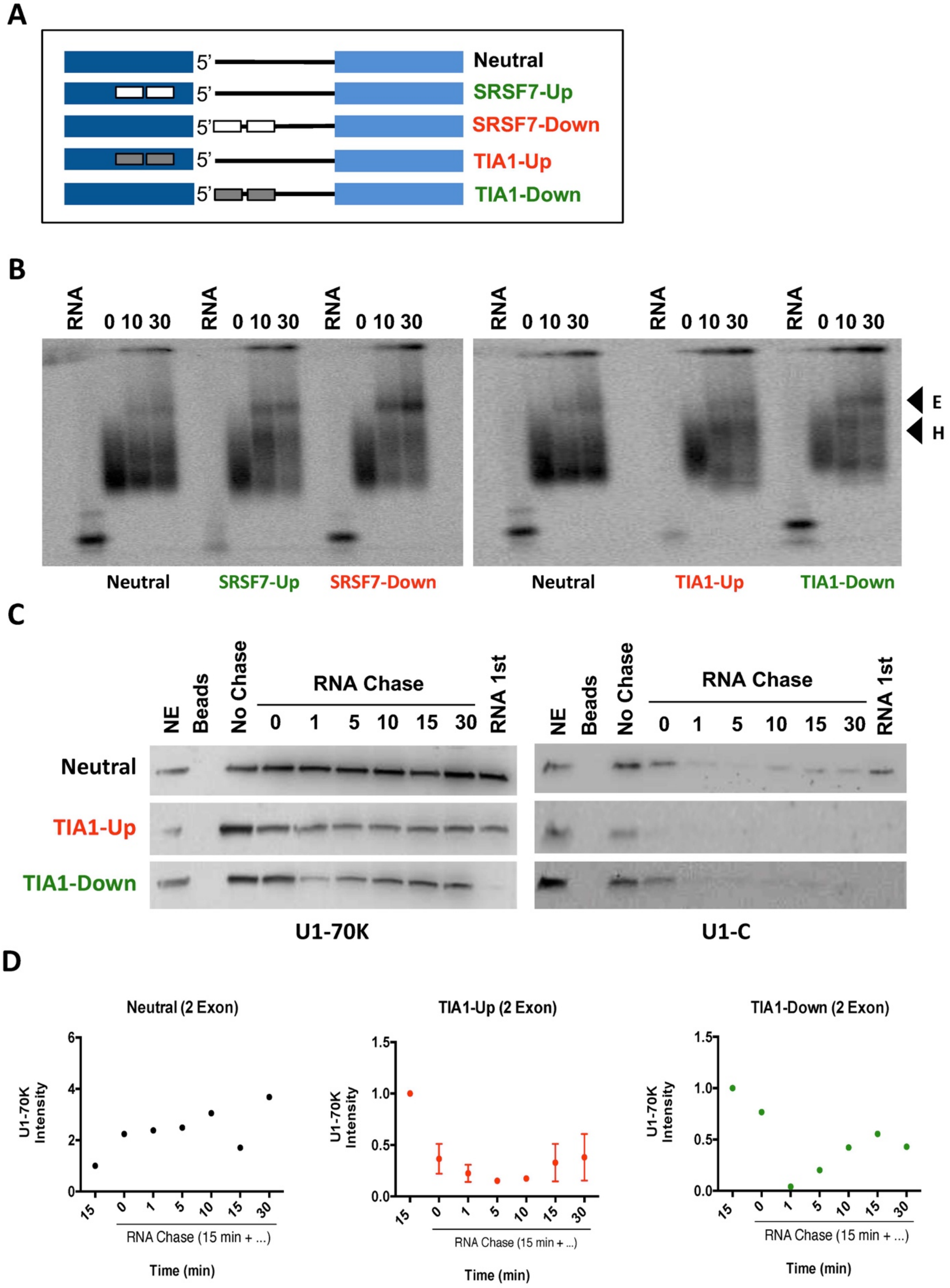


Figure 3.6. Pulse-Chase Pull-Down with or Without ATP on Half-Substrate RNA. Presence or absence of ATP does not change levels of U1-70K. Western blot analysis was performed on half-substrate RNA pulse-chase pulldown samples. Pulse-chase was performed using high concentrations of neutral RNA with no binding site for SRSF7 or TIA-1 for 15 minutes plus an additional 0-30 minutes. A, B) Pulse-chase without ATP, in duplicate C) Pulse-chase with ATP. 30% NE was loaded onto the gel as control, bead sample contains no half-substrate RNA and was used as non-specific binding control, 'No Chase' sample denotes no chase RNA and should mimic pulse-chase time point 0. RNA 1st denotes samples with chase RNA introduced prior to the 15 minute incubation with NE, as an internal assay control.

Two exon mini gene to probe U1 snRNP affinity changes.

It is possible that the half-substrate RNAs used in our assays were not fully reflective of accurate assembly modes, and any changes observed in the association of U1 snRNP to activated- or repressed-splice sites could be due to the lack of proper interactions between protein components at the 5'ss and 3'ss downstream. To test this concern, a two-exon mini gene was designed and tested with the pulse-chase pull-down (Figure 3.7A). The two-exon genes were verified for formation of E complex, and all RNAs demonstrated E complex formation (Figure 3.7B). Pulse-chase pull-down experiments were then performed, however, there was still a lack of reproducibility. The same general trend seemed to be apparent for the neutral RNA, tight association of U1-70K to the neutral RNA, however, not much information could be obtained for the activating and repressive state RNAs (Figure 3.7C, 3.7D).

Figure 3.7. E Complex Formation on Two Exon-Mini Genes.
Two exon mini genes demonstrate **l**oss of E complex formation in general. A) Schematic of two exon RNA constructs used. B) Western blot analysis was performed on two-exon RNA. Pulse-chase was performed using high concentrations of neutral RNA with no binding site for SRSF7 or TIA-1 for 15 minutes plus an additional 0-30 minutes using ATP depleted NE. 30% NE was loaded onto the gel as control, bead sample contains no half-substrate RNA and was used as non-specific binding control, 'No Chase' sample denotes no chase RNA and should mimic pulse-chase time point 0. RNA 1st denotes samples with chase RNA introduced prior to the 15 minute incubation with NE, as an internal assay control. C) Levels of U1-70K intensity as measured from the western blot were plotted over time. D) Two exon mini genes demonstrate **l**oss of H complex and formation of E complex. Native complex formation was performed on low melt agarose. RNA was added to ATP depleted nuclear extract mix for 0, 10 or 30 minutes. RNA not introduced to nuclear extract was loaded onto gel as a control. E complex and H complex denoted with black arrows.


A
5'
Neutral
SRSF7-Up
SRSF7-Down
TIA1-Up
TIA1-Down
B
RNA
0 10 30
E
H
Neutral
SRSF7-Up
SRSF7-Down
Neutral
TIA1-Up
TIA1-Down
C
NE
Beads
No Chase
RNA Chase
0 1 5 10 15 30
RNA 1st
Neutral
TIA1-Up
TIA1-Down
U1-70K
U1-C
D
Neutral (2 Exon)
TIA1-Up (2 Exon)
TIA1-Down (2 Exon)
U1-70K Intensity
15 0 1 5 10 15 30
RNA Chase (15 min + ...)
Time (min)

Some of the major concerns with the affinity pull-down model using both the half-substrate and two exon mini genes involved bead loss and thus sample loss, as well as loss of U1 snRNP components through the many wash steps. The adipic acid dihydrazide beads are sticky, and bead loss occurred through the many pipetting and wash steps. Additionally, there was a concern that introducing multiple washes to remove any non-specifically bound protein or snRNPs could be washing away non-tightly associated U1 snRNP components. Furthermore, with the current method, no internal loading control could be used, so any pipetting errors in sample loading on the gel could not be ruled out, and any major differences between protein intensity from one sample to another could not be confirmed.

To address some of these issues, a different pull-down method was employed using an MS2 affinity tag as a means to pull-down the RNA of interest. The goal of this new method was to improve the pull-down through the use of magnetic amylose beads to minimize bead loss, as well as to reduce the concentration of RNA needed to perform the pull-downs. Three MS2 hairpins were cloned into the half-substrate RNA immediately downstream of the T7 promoter and upstream of the exonic sequence (Figure 3.8A). MS2-MBP (maltose binding protein) fusion proteins were introduced to the MS2-RNAs of interest, subsequently incubated in NE, and pulled-down using magnetic amylose beads. The affinity purified proteins were then eluted and analyzed by western blot. Using this method of affinity purification allowed for (10-fold) far less RNA to be used to perform the pull-down.

Affinity purification using SRSF7-Down demonstrated an increase in U1-70K levels over time during the RNA excess chase (Figure 3.8B). However, upon repeat trials, U1-70K levels oscillated between time points and were not consistent with previous results (Figure 3.8C). Reproducibility of the data was still a key concern. Additionally, U1-C levels, as previously seen, were too low to appear on the western blot. This pull-down was repeated numerous times, and subsequent tests yielded different results and trends. The number of washes, temperature at which the pull-down was performed, as well as order of the experimental time points were all tested, and reproducible data was not attained. Additionally, a new control was introduced to reduce the number of washes needed after incubation with NE, potentially addressing any loss of weakly or transiently bound U1snRNP over the time course of the wash steps. This control involved adding MS2 protein in excess to the amylose beads, prior to their introduction to the MS2-MBP RNA complex. Unfortunately, this additional control did not lower the non-specific wash conditions.

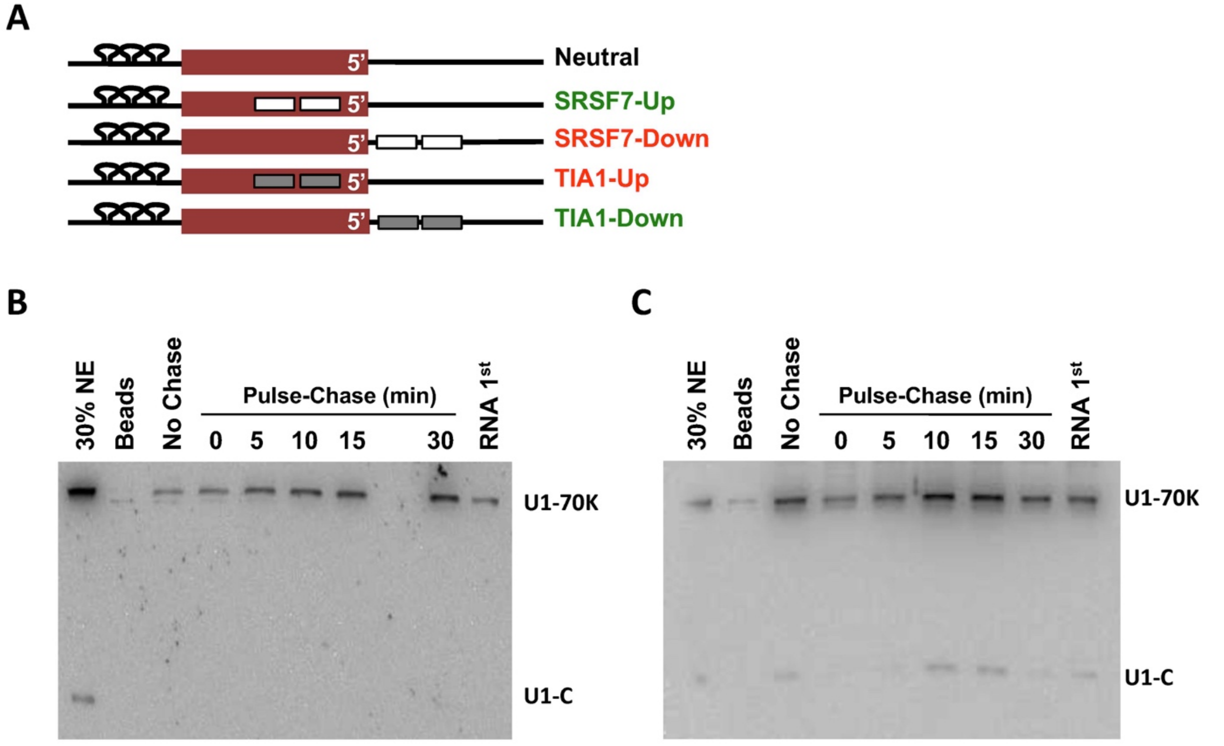


Figure 3.8. 3MS2 Hairpin Half-Substrate RNA Pulse-Chase Pull-Down.
RNA pulldown using 3MS2 hairpin RNA did not improve reproducibility issues. A) Schematic of RNA used. Three hairpins denote MS2 binding sites, red box denotes exon, black line denotes intron, white box denotes SRSF7 binding site. B, C) Western blot analysis was performed on pulse-chase pulldown samples using SRSF7-Down RNA and probed for U1-70K and U1-C levels. Pulse-chase was performed using high concentrations of neutral RNA with no binding site for SRSF7 or TIA-1 for 0-30 minutes. NE denotes 30% nuclear extract loaded onto gel, bead sample contains no RNA, 'No Chase' sample denotes no pulse-chase RNA, and RNA first denotes samples with chase RNA introduced prior to 15 minute incubation, used as control. Two gels depicted are replicates experiments.

The internal RNA 1st control continued to demonstrate high levels of U1-70K, suggesting an inefficient chase. To address this concern, a new chase RNA was designed. The chase RNA that was initially used was the half-substrate neutral sequence that did not contain MS2 hairpins (Figure 3.2A), however, it was seen that low levels of SRSF7 and TIA-1 bound non-specifically to the neutral sequence, when the neutral sequence was introduced at high concentrations (data not shown). Therefore, a new chase sequence was designed, removing any potential SR protein/hnRNP binding sites, while maintaining a strong 5'ss (new neutral chase sequence: 5'-GAGCUCCAGGUGAGUACACAUAU-3'). The new chase RNA did not diminish the RNA 1st signal and did not improve the oscillating levels of U1-70K.

Overall, this new pull-down method, although more manually efficient and requiring less RNA input, did not improve the reproducibility issues. Further work is needed to troubleshoot and optimize this method, in particular, improving the chase RNA and in turn improve the RNA 1st control results.

Discussion

Understanding the mechanism by which splicing regulation occurs allows for greater insights into the complexity of RNA processing. The finding that SR proteins and hnRNPs behave in a similar but opposite fashion, yet still recruit U1 snRNP to the 5'ss, demonstrates the need to understand the mechanism of position-dependent regulation. Based on the findings thus far, the U1 snRNP complex displays differing protein component integrity, when binding to the 5'ss is influenced by a splicing activator or repressor (Figure 3.3, 3.4). A change in the strength of U1 snRNP complex interconnections, as measured by variable U1-70K and U1-C dissociation rates when comparing activating or repressive conditions, suggests that the U1 snRNP complex is highly dynamic. These dynamic changes could be associated with altered protein component affinities, perhaps even altered complex compositions and/or structural conformations that ultimately dictate the assembly of a functional spliceosome. In other words, splicing repression and activation at the 5'ss may be mediated through U1 snRNP structural and complex integrity changes. As suggested by the results highlighted in Figure 3.4, in the presence of a repressor SRSF7 bound downstream of the 5'ss, a faster rate of dissociation for U1-70K was seen. In other words, U1-70K has a lower affinity for U1 snRNP in the presence of a repressor. Overall, the data thus far suggests dynamic binding of U1 components at the 5'ss in the presence of activator/repressor. It is possible that changes to the integrity of protein components associated with U1 snRNP impact the overall affinity of U1 snRNP for the RNA.

It has been shown that there are two differing isoforms of U1-70K that contain RNA binding domains [156, 157]. These two isoforms differ by only 9 amino acids which affect the phosphorylation status of U1-70K. U1-70K isoform one, the larger of both isoforms, demonstrated a more stable interaction with U1-C, whereas isoform two bound more stably to Sm-B [156]. Sm-B has been shown to be an important component for association of additional non-snRNP factors with U1 snRNP [158]. Furthermore, mass spec results showed that U1-C and U1-A were not retained in complexes where U1-70K was lost [156]. This suggests that depending on the isoform of U1-70K that is bound, U1 snRNP integrity could differ, and the stability of U1 snRNP may not be achieved until U1-70K, U1-C and U1-A are all stably associated. This could explain the difficulty in calculating dissociation rates for U1-A and U1-C, when U1-70K is lost due to the presence of the repressor. Moreover, given that U1-70K acts as an anchor for U1-C [159–161], it is plausible that U1 snRNP dynamic changes that are seen, in the presence of a repressor, could be due to changes in association between

U1-70K and U1-C or U1-70K and Sm-B, but also depend on the isoform of U1-70K that is associated. These could lead to varying modes of U1 snRNP that associate with the 5’ss. Furthermore, phosphorylation status could also be altered, affecting the way these protein component associate with U1-70K and other associated factors. Therefore, it is of great importance to validate the isoform of U1-70K that is associated, the phosphorylation status of this protein as well as probing for the other U1 snRNP components, such as the Sm proteins. In the presence of a repressor, loss of U1-70K may modify not only U1-C levels, but also Sm-B levels and these changes could affect the overall stability of U1 snRNP for the 5’ss, leading to the stalled E complex formation that was seen [57].

In addition to altered protein component affinities, altered structural conformations could also play a role in U1 snRNP complex dynamics. The Black lab has demonstrated that stem loop 4 (SL4) of U1 snRNA interacts across the intron with SF3A1, a U2 snRNP component [162]. Likewise, they have shown that polypyrimidine tract binding protein (PTB) binding to a distinct SL4 region of SL4 inhibits splicing [163]. It has also been shown that TIA-1 directly interacts with U1-C when located downstream of the 5’ss [164]. Given that U1-C requires U1-70K for stable incorporation into U1 snRNP [161], it is possible that the U1-C/U1-70K interaction could result in changes to the accessibility of stem loop 1 of U1 snRNA to which U1-70K binds. Changes in accessibility of other stem loops could also alter the overall conformation of U1 snRNA impacting how the U1 snRNP components interact. For instance, it is possible that the accessibility of SL4 of U1 snRNA changes in the presence of a SR protein or hnRNP. This could potentially occur through a conformational change of SL4, thereby allowing U1 snRNP to toggle between permissive and non-permissive structures. In this model the permissive structure would promote the formation of a productive U1 snRNP complex, which is required for cross-intron pairing with U2 snRNP components to assemble functional pre-spliceosomal complexes.

U1 snRNP has been shown to interact with the pre-mRNA at multiple locations, independent of a canonical 5'ss. Consequently, the spliceosome must be able to differentiate between U1 snRNP that binds to the RNA outside of splice sites and those that mark exon/intron junctions or pseudo splice sites and cryptic sites [124, 165]. Based on the data shown thus far, splicing regulatory proteins modulate interactions between U1 snRNP components and these interactions may in turn regulate the ability of U1 snRNP to engage in productive higher order spliceosomal complex formation. Thus, an activated U1 snRNP may act as a gatekeeper of splicing initiation by altering the interaction profile of the U1 snRNP components. Collectively, these results demonstrate the need to understand how regulatory proteins interact with and/or alter U1 snRNA accessibility and/or U1 snRNP integrity. The overall preliminary findings are promising, and further investigation would allow for greater insight into the nature of U1 snRNP as a molecular checkpoint that influences splicing decisions of the pre-mRNA.

Materials & Methods

Half-Substrate Construct Sequences

Half-substrate DNA sequences are as follows. SRSF7 sequence used in constructs was 5'-
AGACAACGATTGATCGACTA-3' and TIA1 sequence used was 5'-TCTTTTTAAGTCGT
ACCTAA-3'.

Neutral (N2 sequence):
5'-TAATACGACTCACATAGGGCCAAACAACCAAACAACCAAACAAGAGCTCCTGGTGAGTA
CCCAAACAACCAAACAACCAAACAACCAAACAACCAAACAACTTAAGCTCTCCGAAGACAGT
GG-3'.

SRSF7-Upstream:
5’-TAATACGACTCACTATAGGGCCTAGGAATTCAGACAACGATTGATCGACTA
AGACAACGATTGATCGACTAGAGCTCCTGGTGAGTACCTTAAGCTCTCCGAAGACAGTGG-3’

SRSF7-Downstream:
5’-TAATACGACTCACTATAGGGCCTAGGAATTCCCAAACAACCAAACAACCAAACAAGAG
CTCCTGGTGAGTACAGACAACGATTGATCGACTAAGACAACGATTGATCGACTACTTAAGCTCTC
CGAAGACAGTGG-3’

TIA1-Upstream:
5’-TAATACGACTCACTATAGGGCCTAGGAATTCTCTTTTTAAGTCGTACCTAATCTTTTT
AAGTCGTACCTAAGAGCTCCTGGTGAGTACCTTAAGCTCTCCGAAGACAGTGG-3’

TIA1-Downstream:
5’-TAATACGACTCACTATAGGGCCTAGGAATTCCCAAACAACCAAACAACCAAACAAGA
GCTCCTGGTGAGTACTCTTTTTAAGTCGTACCTAATCTTTTTAAGTCGTACCTAACTTAAGCTCT
CCGAAGACAGTGG-3’

Cloning of 3MS2 Hairpins in Half-Substrate DNA (SD3MS2 DNA)

Half-substrate DNA (Neutral, SRSF7-Up, SRSF7-Down, TIA1-Up, and TIA1-Down) was cloned into pBluescript II KS+ vector containing three MS2 hairpins downstream of T7 promoter. Xba1 and HindIII were restriction sites used to clone into the vector. MS2 sequence used was 5’-CGTACACCATCAGGGTACG-3’. Final sequence of new SD3MS2 DNA are as follows.

Neutral:
5’-TAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGGCGTACACCATCAGGGTAC
GAGCAAGCCCATTGCGTACACCATCAGGGTACGACTAGTACATTCGTACACCATCAGGGTACGGT
ATTCCATCTAGATATAGGGCCAAACAACCAAACAACCAAACAAGAGCTCCTGGTGAGTACCCAAA
CAACCAAACAACCAAACAACCAAACAACCAAACAACTTAAGCTCTCCGAAAAGCTTACA-3’

SRSF7-Upstream:
5’-TAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGGCGTACACCATCAGGGTACG
AGCAAGCCCATTGCGTACACCATCAGGGTACGACTAGTACATTCGTACACCATCAGGGTACGGTA
TTCACATCTAGATATAGGGCCTAGGAATTCAGACAACGATTGATCGACTAAGACAACGATTGAT
CGACTAGAGCTCCTGGTGAGTACCTTAAGCTCTCCGAAAAGCTTACA-3’

SRSF7-Downstream:
5’-TAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGGCGTACACCATCAGGGTA

CGAGCAAGCCCATTGCGTACACCATCAGGGTACGACTAGTACATTCGTACACCATCAGGGTACGG
TATTCACATCTAGATATAGGGCCTAGGAATTCCCAAACAACCAAACAACCAAACAAGAGCTCCTG
GTGAGTACAGACAACGATTGATCGACTAAGACAACGATTGATCGACTACTTAAGCTCTCCGAAA
AGCTTACA-3’

TIA1-Upstream:
5’-TAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGGCGTACACCATCAGGGTA
CGAGCAAGCCCATTGCGTACACCATCAGGGTACGACTAGTACATTCGTACACCATCAGGGTACGG
TATTCACATCTAGATATAGGGCCTAGGAATTCTCTTTTTAAGTCGTACCTAATCTTTTTAAGTCG
TACCTAAGAGCTCCTGGTGAGTACCTTAAGCTCTCCGAAAAGCTTACA-3’

TIA1-Downstream:
5’-TAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGGCGTACACCATCAGGGTAC
GAGCAAGCCCATTGCGTACACCATCAGGGTACGACTAGTACATTCGTACACCATCAGGGTACGGT
ATTCACATCTAGATATAGGGCCTAGGAATTCCCAAACAACCAAACAACCAAACAAGAGCTCCTGG
TGAGTACTCTTTTTAAGTCGTACCTAATCTTTTTAAGTCGTACCTAACTTAAGCTCTCCGAAAAG
CTTACA-3’

Generation of Two Exon Mini Genes Via PCR Stitching

Using the half-substrate DNAs (Neutral, SRSF7-Up, SRSF7-Down, TIA1-Up, and TIA1-Down), a two-exon mini gene was generated via PCR stitching using 135bp sequence of a β-globin exon. All PCR amplification reaction mixtures used the following: 2mM dNTP’s, 1x Thermopol buffer (New England Biolabs), 10uM forward and reverse primer, and 0.4μL Taq polymerase (APEX).

PCR amplification of a linker sequence. Half substrate RNA was PCR amplified with a modified T7 forward primer with sequence 5’-TTTTTGGAGGTAATACGACTCACTATAGGG-3’ and a β-globin reverse “linker” primer with the sequence 5’-AAGCTCTCCGAAGACAGTGGCAGAGAAGACTCTTGGGTTTC-3’. PCR amplification was performed in a two-step fashion with initial denaturation at 95°C, 3min. Cycle step 1: denaturation at 95°C, 45sec, annealing at 68°C, 45sec and extension at 72°C, 1min for 5 cycles. Cycle step 2: denaturation at 95°C, 45sec, annealing at 54°C, 30sec and extension at 72°C, 1min for 20 cycles. Final extension at 72°C, 5min. Hold at 4°C.

PCR amplification of β-globin region. The β-globin sequence used was as follows: 5’-
CAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACC
CTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCAC
TCCTGATGCTGTTATGGGCAACCCTAAGG-3’. The β-globin region of interest was amplified using a β-globin forward primer with sequence 5’-CAGAGAAGACTCTTGGGTTTCTGA-3’ and a reverse primer with sequence 5’-CCTTAGGGTTGCCCATAACAGC-3’. PCR amplification program used was the following: initial denaturation at 95°C, 3min., followed by denaturation at 95°C, 45sec, annealing at 58°C, 30sec and extension at 72°C, 1min for 25 cycles. Final extension at 72°C, 5 min. Hold at 4°C.

Half-substrate linker DNA and β-globin DNA stitching. A modified T7 forward primer with sequence 5’-TTTTTGGAGGTAATACGACTCACTATAGGG-3’ and a β-globin reverse primer with sequence 5’-CCTTAGGGTTGCCCATAACAGC-3’ were used to amplify the final two-exon DNA sequence of interest. PCR amplification program used was the following: initial denaturation at 95°C, 3min., followed by denaturation at 95°C, 45sec, annealing at 54°C, 30sec and extension at 72°C, 1min for 25 cycles. Final extension at 72°C, 5 min. Hold at 4°C. Splicing efficiency of two-exon constructs was tested following the method as described in Appendix A.

Final sequence of two-exon mini genes are as follows.

Neutral (2 exon):
5’-TTTTTGGAGGTAATACGACTCACTATAGGGCCAAACAACCAAACAACCAAACAAGAGCT
CCTGGTGAGTACCCAAACAACCAAACAACCAAACAACCAAACAACCAAACAACTTAAGCTCTCCG
AAGACAGTGGCAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCTA

TTTTCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGG
ATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGG-3’

SRSF7-Upstream (2 exon):
5’-TTTTTGGAGGTAATACGACTCACTATAGGGCCTAGGAATTCAGACAACGATTGATCGA
CTAAGACAACGATTGATCGACTAGAGCTCCTGGTGAGTACCTTAAGCTCTCCGAAGACAGTGGCA
GAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACCCT
TAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTC
CTGATGCTGTTATGGGCAACCCTAAGG-3’

SRSF7-Downstream (2 exon):
5’-TTTTTGGAGGTAATACGACTCACTATAGGGCCTAGGAATTCCCAAACAACCAAACAAC
CAAACAAGAGCTCCTGGTGAGTACAGACAACGATTGATCGACTAAGACAACGATTGATCGACTA
CTTAAGCTCTCCGAAGACAGTGGCAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCT
GCCTATTGGTCTATTTTCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTT
GAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGG-3’

TIA1-Upstream (2 exon):
5’-TTTTTGGAGGTAATACGACTCACTATAGGGCCTAGGAATTCTCTTTTTAAGTCGTACCTA
ATCTTTTTAAGTCGTACCTAAGAGCTCCTGGTGAGTACCTTAAGCTCTCCGAAGACAGTGGCAGA
GAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACCCTTA
GGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCT
GATGCTGTTATGGGCAACCCTAAGG-3’

TIA1-Downstream (2 exon):
5’-TTTTTGGAGGTAATACGACTCACTATAGGGCCTAGGAATTCCCAAACAACCAAACAAC
CAAACAAGAGCTCCTGGTGAGTACTCTTTTTAAGTCGTACCTAATCTTTTTAAGTCGTACCTAAC
TTAAGCTCTCCGAAGACAGTGGCAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTG
CCTATTGGTCTATTTTCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTG
AGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGG-3’

PCR amplification of DNA Constructs

10-25 ng DNA was PCR amplified using 2mM dNTP’s, 1x Thermopol buffer (New England Biolabs), 10uM forward and reverse primer, and 0.4uL Taq polymerase (APEX). For half-substrate RNA, primers used were forward (T7) sequence 5’-TAATACGACTCACTATAGGG-3’ and reverse sequence 5’-AAGCTCTCCGAAGACAGTGG-3’. For PCR amplification of SD3MS2 RNA, primers used were forward (T7) sequence 5’-TAATACGACTCACTATAGGG-3’ and reverse sequence 5’-CTTAAGCTCTCCGAAAAGCTTACA-

3'. PCR amplification program was the following: initial denaturation at 95°C, 3min., followed by denaturation at 95°C, 45sec, annealing at 50°C or 52°C (Tm=50°C for half-substrate DNA and T_m=52°C for SD3MS2 DNA) 30sec and extension at 72°C, 1min for 30 cycles. Final extension at 72°C, 5 min. Hold at 4°C.

RNA Transcription

RNA was transcribed using the T7 Ribomax Kit (Promega). Briefly, 20μL of T7 5X transcription buffer, 30uL of rNTP's, 40μL of 5-10ug of PCR amplified half-substrate DNA or SD3MS2 DNA constructs (Neutral, SRSF7-Up, SRSF7-Down, TIA1-Up, and TIA1-Down) and 10μL of T7 enzyme mix were incubated at 37°C for 2.5hrs. RNA was cleaned using ZYMO Clean & Concentrator Kit-25 with DNase on column. RNA eluted from column with 60μL nuclease free water. Double elute if needed.

RNA Affinity Pull-Down Protocol

3' end modification of RNA. Dilute 300 pmoles RNA in a 400uL reaction containing 100mM NaOAc (Sodium Acetate), pH 5 and 10mM Na(m)IO4 (Sodium Meta-Periodate) with RNase free water. Incubate 1hr at room temperature in the dark. Note: Na(m)IO4 *is* light sensitive. Precipitate RNA with 1mL ice cold 100% EtOH at -80°C for 5min. Centrifuge RNA solution at room temperature, 16,000 RPM for 10min. Remove all supernatant.

Preparation of adipic acid dihydrazide agarose beads. Determine the number of adipic acid dihydrazide agarose beads (Sigma) needed: (125μL beads) x (# of reactions). Use 125μL beads/reaction and wash beads with 100mM ice cold NaOAc using 2x the bead volume. Do not pipet to mix. Spin beads at 4°C, 500 RPM for 3 min. Aspirate supernatant carefully so as not to disrupt or discard beads. Repeat wash step 4x. After the final wash, remove all residual

100mM NaOAc. Measure remaining bead volume and bring total volume up to 250µL of beads/reaction with 100mM NaOAc.

Preparation of RNA/Bead complex. Gently vortex resuspended beads before adding to each RNA sample to maintain an even suspension. Add 250µL of bead mixture to each RNA sample. Incubate on rotator overnight at 4°C. Pellet beads at room temperature, 1,000 RPM, 3min. Wash beads with 1mL ice cold 2M NaCl, 2x. Spin samples between washes at room temperature, 1,000 RPM for 3 min. Wash beads for an additional 3x with ice cold Buffer D (20 mM HEPES-KOH pH 7.6, 20% v/v glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM DTT). Spin samples between washes at room temperature, 1,000 RPM for 3 min. Remove any residual wash/supernatant and keep samples on ice.

Reaction incubation and elution. Dilute nuclear extract (NE) to final desired concentration with Buffer D (use 30% NE). Add 200uL of diluted NE to beads. Mix by pipetting gently, rotate on shaker at 30°C for 10-15min. Wash reaction 5x with 1mL Buffer D+$MgCL_2$ (20 mM HEPES-KOH [pH 7.6, 20% [vol/vol] glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM DTT, 1mM $MgCl_2$). Spin samples at room temperature, 1,000 RPM for 3min to pellet beads. Resuspend pellet in equal volume of desired solution for downstream use. Elute proteins using 60µL 2X Laemmli's protein sample buffer (4% w/v SDS, 20% v/v glycerol, 120mM Tris-HCl pH 6.8, 200mM DTT, 0.1% w/v bromophenol blue). Incubate beads at 95°C for 10min. Store at -20°C.

Pulse-Chase in conjunction with RNA affinity pull-down. Pulse-chase experiment was performed by incubating pull-down RNA with 30% NE for 15 minutes at 37°C and then chasing with neutral RNA for a further 0, 1, 5, 10, 15 or 30 minutes at 37°C. To confirm NE does not bind to beads, a bead only control was used and incubated with NE for 15 minutes.

A no-chase sample was used as positive control and should mimic the 0 pulse-chase time point, as no time is allowed for the neutral RNA to chase U1 snRNP away. As an internal control for accuracy of the competition assay, neutral RNA was introduced to a sample prior to incubation with NE. This was called 'RNA 1st'. This control was aimed at confirming efficient chasing by the neutral sequence being used.

Protein Analysis and Western Blot Assay

RNA pull-down samples were separated on pre-cast 12% Tris-Glycine SDS-PAGE gels (Bio-Rad) at 125V for 50-60 min at room temperature. Gels were transferred to methanol activated Immun-Blot® PVDF membrane (Bio-Rad) using Tris-Glycine transfer buffer at 60V, 2hrs, 4°C, with an ice pack and stirring of transfer buffer. Membrane was blocked for 1hr, at room temperature with 5% non-fat milk in tris buffered saline with Tween-20 (TBS-T), followed by an overnight incubation with primary antibody at 4°C. Primary antibody was diluted in 2.5% non-fat milk in TBS-T. Primary antibody conditions in this work were the following: U1-70K at 1:100 (Santa Cruz Biotechnology, clone C-18 or Millipore mouse monoclonal, clone 9C4.1), U1-C at 1:200 (Sigma, rat monoclonal, clone 4H12), U1-A at 1:100 (Santa Cruz Biotechnology, mouse monoclonal, clone BJ-7), hnRNP A1 at 1:250 (Santa Cruz Biotechnology, mouse monoclonal, clone 9H10), SRSF7 at 1:1000 (MBL, rabbit polyclonal) and TIA-1 at 1:100 (Santa Cruz Biotechnology, goat polyclonal, clone C-20). Membrane was washed with TBS-T prior to incubation with secondary. Secondary antibodies were diluted in 2.5% non-fat milk in TBS-T and incubated for 1hr at room temperature. Membrane was washed further with TBS-T and then developed for 1min using SuperSignal™ West Pico (or

Femto) Chemiluminescent Substrate (ThermoFisher). Blots were imaged on a Kodak Imager Station. Exposure times varied depending on protein intensity levels.

Native Complex Formation

Native agarose gel formation. Prepare 1.5% low melt agarose (Invitrogen) and allow the gel to completely polymerize overnight at 4°C for optimal results. Add pre-cooled 1X Tris-Glycine running buffer (50mM Tris, 50mM Glycine) to the gel and pre-warm the gel for 20min at 5W. Run all reactions on the gel at 4°C.

Reaction setup. For E complex formation, deplete ATP by incubating NE at room temperature for 20-30 minutes. For higher order complex formation (A, B, C complexes), add 5ul of 4mg/mL heparin to the reaction. Prepare and keep all samples on ice. Prepare reaction master mix with a final concentration of 3.2mM MgOAc (magnesium acetate), 10 units of RNasin® (Promega), 1mM DTT, 42.6mM KOAc (potassium acetate), 12mM HEPES, pH 7.9 and 30% NE in a 25µL reaction. Lastly, add 1µL of freshly prepared P^{32}-radiolabelled RNA to 25µL reaction master mix. P^{32}-radiolabelled transcribed RNA was prepared as described in Appendix A. Incubate samples at 30°C for given time points. Make sure all samples end incubation at the same time and place samples on ice. Add 5uL of nucleic acid sample loading buffer (1X TBE, 20% glycerol, 0.25% bromophenol blue, 0.25% xylene cyanol). Load sample to gel and run samples at 5W, 4-5hrs at 4°C. Do not run the gel for less than 4hrs to maintain optimal separation for E complex formation.

Fix the gel with 10% methanol, 10% glacial acetic acid solution for 30 min. Lay the gel onto pre-cut Whatman paper. Cover the gel with saran wrap and expose to a phosphor screen for minimally 1hr.

MS2-MBP Fusion Protein RNA Affinity Protocol

*Preparation of magnetic amylose beads***.** Vortex magnetic amylose beads (New England Biolabs) prior to use. Aliquot 50µL per reaction of bead suspension into a clean 1.5mL non-stick tube. Add 500µL (per 100µL of beads) of RNA wash buffer (20 mM Hepes-KOH, pH 7.9, 100 mM KCl, 1 mM $MgCl_2$, 1% Triton X-100, 0.5 mM DTT [add fresh before use]). Vortex gently to resuspend. Apply to MagStand for 30sec and decant supernatant 30sec or spin the mixture at 2,000xg for 2 min. Repeat wash 1x. Resuspend beads in 50uL working volume per reaction with RNA wash buffer

*RNA affinity purification and elution***.** Incubate ~70 pmoles of 3MS2-SD RNA with 20 pmoles of MS2-MBP (MS2-MBP was a generous gift from Dr. Yongsheng Shi). Add RNA wash buffer to bring the total volume up to 50µl. Incubate for 30min on ice. Add 50µl pre-washed magnetic amylose beads to each reaction and rotate for 1hr at 4^{o}C. Add 200µL NE (30% final NE concentration) and mix well. Incubate for 15 min at 30^{o}C while rotating. Place samples on MagStand for 30sec or spin the mixture at 2,000xg for 2 min and remove supernatant. Add 1mL RNA wash buffer. Place tubes on MagStand for 30sec-1min or spin the mixture at 2,000xg for 5 min and remove supernatant. Repeat the wash 3x with 1mL of RNA wash buffer.

Elute complex by adding 100µl elution buffer (RNA wash buffer + 12 mM D-maltose monohydrate) to the beads and rotate for 10-20 min at 4^{o}C. Place tubes on MagStand for 30sec-1min or spin down for 2 min at 2,000xg and transfer supernatant to a new 1.5mL tube. Repeat and combine eluates. Add 800µL ice cold acetone and place at -20^{o}C for 20 min to overnight. Spin down proteins at 16,000xg for 20 min at 4°C. Air dry pellet for 10 min at room temperature. Add 50µL of 1x SDS loading buffer (3X loading buffer: 187.5mM Tris-HCl, pH 6.8, 6% w/v SDS, 30% v/v glycerol, 150mM DTT, 0.03% w/v BPB, 2% BME) to dissolve pellet. Heat samples at 95^{o}C for 5 min and continue with western blot analysis.

CHAPTER 4

Coupling Between Alternative Polyadenylation and Alternative Splicing is Limited to Terminal Introns

Summary

Alternative polyadenylation has been implicated as an important regulator of gene expression. In some cases, alternative polyadenylation is known to couple with alternative splicing to influence last intron removal. However, it is unknown whether alternative polyadenylation events influence alternative splicing decisions at upstream exons. Knockdown of the polyadenylation factors CFIm25 or CstF64 in HeLa cells was used as an approach in identifying alternative polyadenylation and alternative splicing events on a genome-wide scale. Although hundreds of alternative splicing events were found to be differentially spliced in the knockdown of CstF64, genes associated with alternative polyadenylation did not exhibit an increased incidence of alternative splicing. These results demonstrate that the coupling between alternative polyadenylation and alternative splicing is usually limited to defining the last exon. The striking influence of CstF64 knockdown on alternative splicing can be explained through its effects on UTR selection of known splicing regulators such as hnRNP A2/B1, thereby indirectly influencing splice site selection. We conclude that changes in the expression of the polyadenylation factor CstF64 influences alternative splicing through indirect effects.

Introduction

Eukaryotic gene expression and proteomic diversity is dependent on the appropriate removal of introns from pre-mRNAs, a process orchestrated by the spliceosome. Multiple mRNA isoforms are generated through alternative splicing (AS), a highly regulated process that has been identified in ~95% of human genes [8, 9]. AS requires the use of different combinations of splice sites resulting in several classes of splicing patterns, such as: alternative 5' splice site selection (Alt5), alternative 3' splice site selection (Alt3), the skipping of complete exons (SE), or the retention of introns (RI).

Many splicing decisions are believed to occur co-transcriptionally [166–170], both through the use of alternative promoters [171–175] as well as in defining the first [128] and last exons. First and last exons only have one flanking splice site and are, therefore, recognized differently than internal exons. Terminal exon definition has been shown to be aided by the polyadenylation machinery [66, 67] through interactions between U2AF and the polyadenylation polymerase (PAP) [119] or cleavage factor 1m (CF1m) [118], or through interactions between splicing factor 3b (SF3b), a component of U2 snRNP, and the cleavage and polyadenylation specificity factor (CPSF) [120]. Furthermore, the U1 snRNP component U1A has been shown to stimulate polyadenylation through interaction with CPSF160 [121], however, U1 snRNP also plays a role in preventing premature cleavage and polyadenylation through binding of U1 snRNA to cryptic poly(A) sites [122] as well as through direct interactions between U1-70K and the PAP [123]. More recent findings have demonstrated that a terminal splice acceptor site [126] and poly(A) tail may be necessary for terminal intron removal with splicing promoted through the coordinated efforts of PAP and nuclear poly(A) binding protein, PABPN1 [127]. Thus, the interactions of splicing and polyadenylation factors at terminal exons play important roles in the regulation and enhancement of both splicing and polyadenylation [176].

Alternative polyadenylation (APA) has gained renewed consideration as an important regulator of gene expression. Similar to AS, the use of alternative poly(A) sites also allows for multiple mRNA isoforms to be generated from a single transcript. Recent analyses has identified APA events in about 70% of human genes [98]. The ability to choose a particular poly(A) site depends on the strength of the poly(A) signal, as well as surrounding *cis*-elements [101, 102]. Genes/transcripts may have alternative poly(A) sites located in internal introns or exons leading to the formation of different protein isoforms. Alternatively, APA events that occur in the 3' untranslated regions (UTR) will result in transcripts with different 3'UTR lengths but will code for the same protein. Changing the length of the 3'UTR is believed to affect the stability, localization, transport, and even translation of the mRNA through altered interactions with microRNAs or other regulatory RNA binding proteins [22, 103–106]. Several polyadenylation factors have been implicated in APA including CF1m [112, 114] and CstF [117, 177]. For example, knockdown of CFIm25 (CFIm25 KD) or CstF64 (CstF64 KD) has been shown to result in the activation of alternative poly(A) sites [114, 177]. Thus, APA events are mediated through factors that are also engaged in interactions with spliceosomal components.

The mechanism of interaction and the extent of coupling between APA and AS are still relatively unexplored. While previous work demonstrated the coupling between splicing and polyadenylation for terminal exons, it is unknown whether there is a mechanistic connection between APA and AS events upstream of the terminal exon. In other words, does an upstream AS event dictate a preference for an APA event or vice versa? Two scenarios can be envisioned. One scenario describes AS and APA events as mechanistically uncoupled. Each process is executed independently from one another, resulting in spliced mRNA isoforms, each of which has a similar distribution of variable 3'UTR lengths as generated from APA (Figure 4.1, left arm). However, if AS and APA events are mechanistically coupled, the resulting mRNA isoforms would be characterized by a selective preference for one type of APA with a particular AS event (Figure 4.1, right arm). As a consequence of this coupled scenario, specific mRNA isoforms could be selectively stabilized or destabilized over other isoforms. Thus, the mechanistic coupling between AS and APA could significantly influence the expression of mRNA isoforms. Using genome-wide approaches, we tested the hypothesis that APA and upstream AS events are functionally linked.

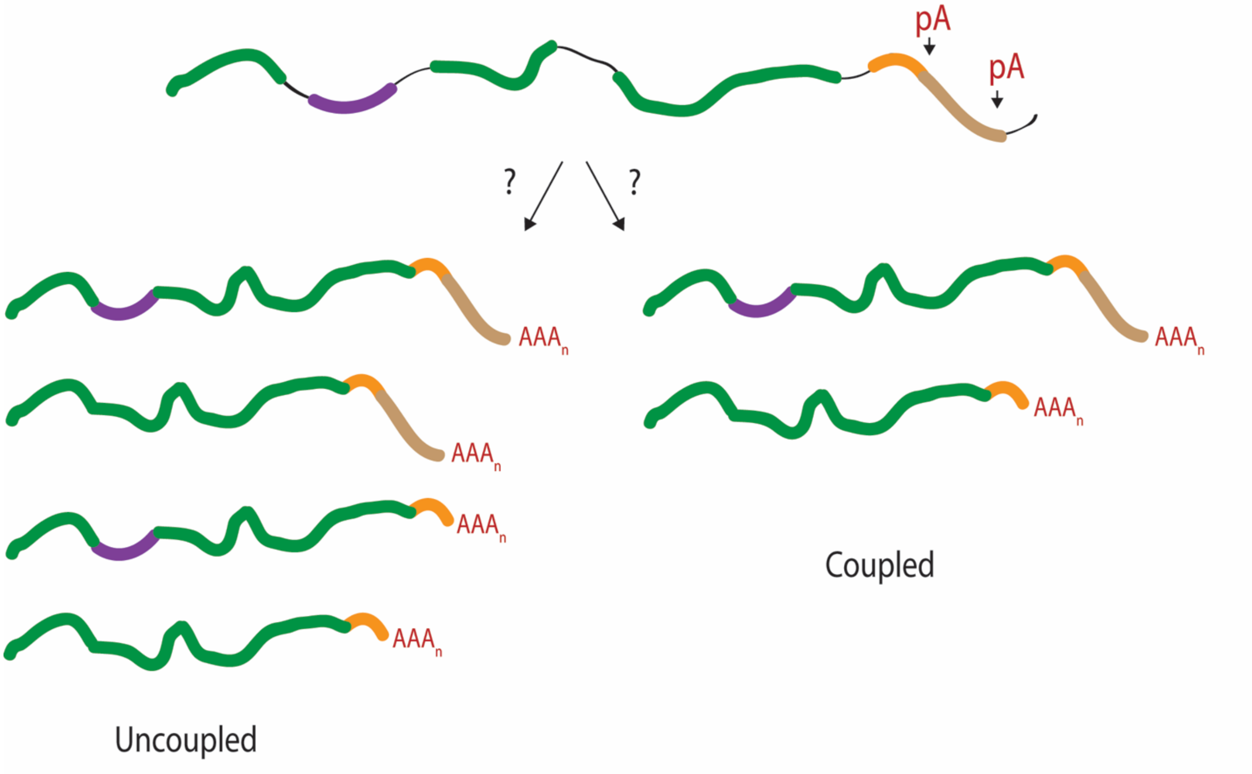


Figure 4.1. Schematic of the Potential Mechanistic Connections Between APA and AS. Consequences of a mechanistic connection between APA and AS events upstream of the terminal exon. It is unknown whether APA and AS events are functionally coupled beyond the definition of the terminal exon. If the RNA processing events are uncoupled (left arm) mRNA isoforms generated through AS would not display a selective preference for a particular APA event. A coupled event (right arm) would preferentially associate one form of APA with a specific AS event. The alternatively spliced exon is depicted in purple, introns shown as black lines, distal poly(A) (pA) site shown in brown, and the proximal pA site is depicted in orange.

Results

Changes in APA correlate with changes in AS at the terminal intron in CFIm25 KD cells.

Previous studies have elucidated several protein interactions between the splicing and polyadenylation machineries, suggesting a coupled process *in vivo*. To test the hypothesis that altering polyadenylation site selection induces AS changes, we induced APA through RNAi mediated knockdown of CFIm25 in HeLa cells, which was shown to induce proximal poly(A) site selection in *TIMP-2*, *Syndecan2*, *ERCC6* and *DHFR* [114]. At efficient CFIm25 KD conditions (60-80%) (Figure 4.2A), reduced CFIm25 levels resulted in alternative *TIMP-2* poly(A) site selection as demonstrated by 3' RACE analysis (Figure 4.2B, 4.2C). *TIMP-2*, an inhibitor of matrix metalloproteinases, contains two functional poly(A) sites, producing mRNAs of 1.2 kb and 3.8 kb. In agreement with previous results [114], PCR-mediated 3' end amplifications demonstrate that CFIm25 KD significantly induced proximal poly(A) site selection (Figure 4.2C).

To determine if changes in APA correlate with changes in AS, RT-PCR analysis was performed along the *TIMP-2* gene using cDNA from CFIm25 KD and control cells. At the terminal exon, depletion of CFIm25 resulted in a 7-fold decrease of an intron retention event (Figure 4.2D, 4.2E). Thus, proximal poly(A) selection increases terminal splicing efficiency in *TIMP-2*.

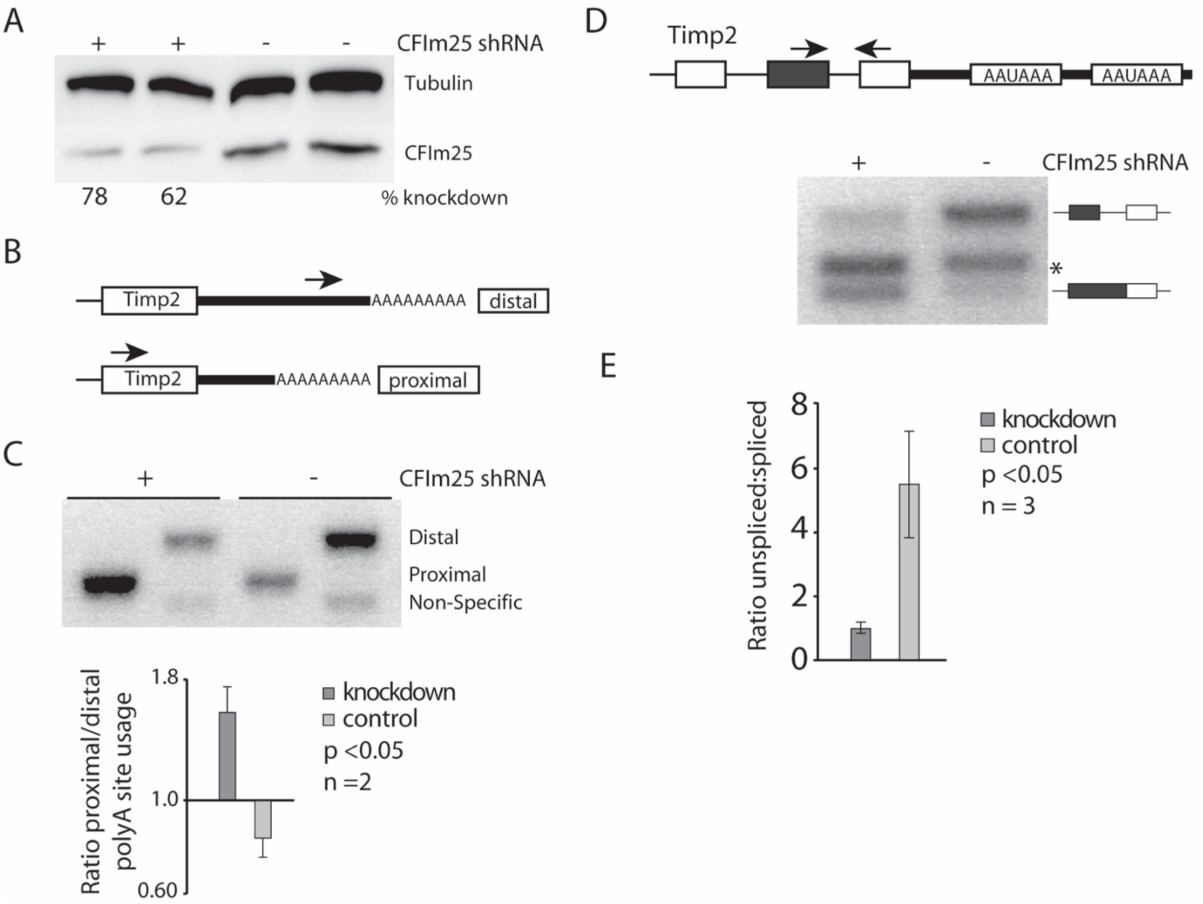


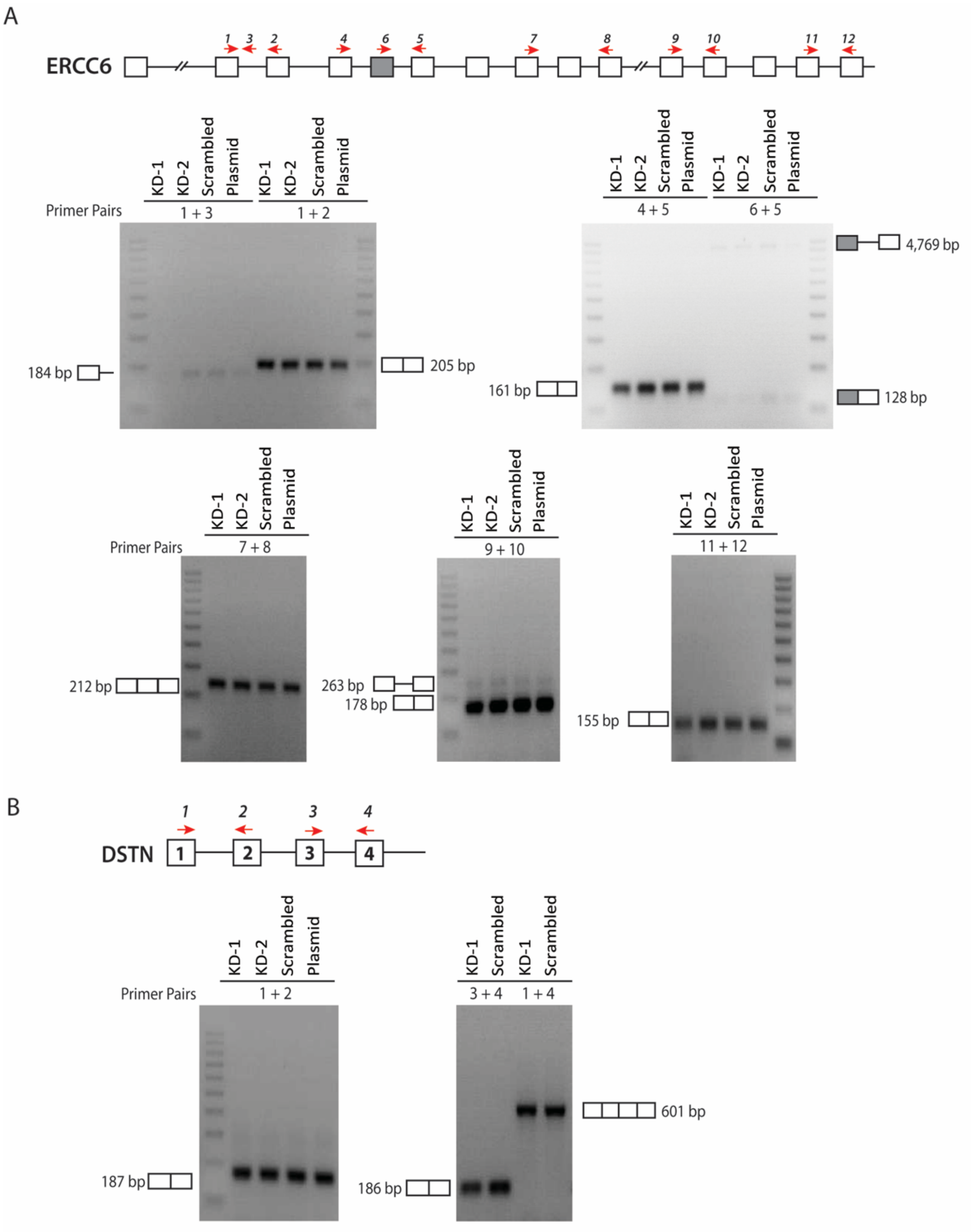
Figure 4.2. Impact of CFIm25 KD on APA.

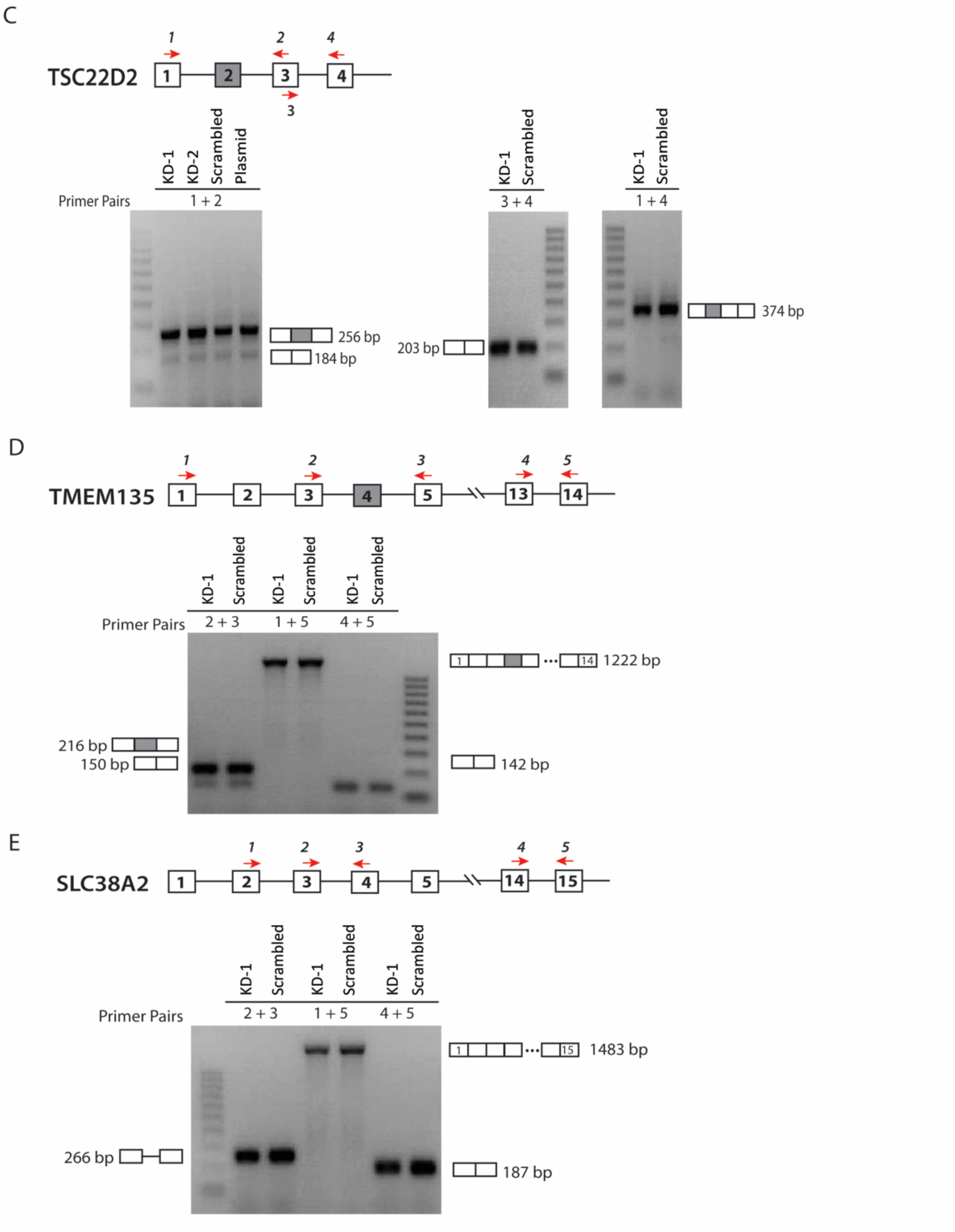
CFIm25 KD induces APA leading to increased last intron retention in *TIMP-2*.

A) Western blot illustrating the efficiency of CFIm25 KD in HeLa cells. B) Diagram of primers used for 3' RACE amplification of *TIMP-2*. The distal primer binds to the 3' UTR sequence of the 3.8 Kb mRNA isoform. The proximal primer binds to the coding sequence of *TIMP-2* but amplifies only from the poly(A) tail of the 1.2 Kb isoform. C) 3' RACE cDNA amplification of *TIMP-2* from cells transfected with CFIm25 shRNA or a negative control vector. The identities of the amplified PCR products are indicated on the side of the representative gel. The results of independent experiments are displayed in the graph below as the ratio of proximal:distal poly(A) site usage. D) Diagram of PCR primers flanking the terminal intron retention event in *TIMP-2*. Asterisk (*) denotes non- specific binding. E) RT-PCR analysis of *TIMP-2* intron retention in cells treated with CFIm25 shRNA or control shRNA. The unspliced and spliced products amplified are indicated on the left of the representative gel. The results of independent experiments are displayed in the graph as the ratio of unspliced:spliced.

These results demonstrate that proximal poly(A) site selection influences the efficiency of terminal intron removal. To address the question of whether splicing events upstream of the terminal exon were affected by CFIm25 KD, the inclusion of known alternative exons within genes that harbor conserved tandem APA events [178] were evaluated using RT-PCR approaches (*TMEM135, DSTN, TSC22D2, SLC38A2,* and *ERCC6*). None of the AS events tested displayed significant differences upon knockdown of CFIm25 (Figure 4.3). These initial results suggest that changes in poly(A) site usage can influence AS near the 3' UTR. However, there does not appear to be a correlation between upstream AS alterations and APA, supporting the notion that mechanistic ties between upstream AS events and APA are limited to terminal intron removal.

Figure 4.3. RT-PCR Analysis of Upstream AS evens in CFIm25 KD Samples.
No change in upstream AS events identified for genes with known changes in APA due to CFIm25 KD. RT-PCR analysis performed on primer pairs for five genes. A) AS events tested for ERCC6: RI (primer pairs 1+3, 1+2, 9+10, 11+12), and SE (primer pairs 4+5, 6+5, 7+8). B) RI events tested for DSTN. C) AS events tested for TSC22D2: SE (primer pair 1+2) and RI (primer pair 3+4). D) AS events tested for TMEM135: RI (primer pair 2+3) and SE (primer pair 4+5). E) AS events tested for SLC38A2: RI (primer pair 2+3, 4+5). Primers depicted as forward and reverse red arrows. KD-1 and KD-2 are replicate samples of CF1m25 KD. Scrambled and plasmid run as controls. Gray boxes depict a skipped exon. Expected product sizes listed on side of gel images. 100 bp ladder used as size markers.


A
ERCC6
1
3
2
4
6
5
7
8
9
10
11
12
Primer Pairs
KD-1
KD-2
Scrambled
Plasmid
1 + 3
1 + 2
184 bp
205 bp
4 + 5
6 + 5
4,769 bp
161 bp
128 bp
7 + 8
212 bp
9 + 10
263 bp
178 bp
11 + 12
155 bp
B
DSTN
1
2
3
4
1 + 2
187 bp
3 + 4
1 + 4
186 bp
601 bp



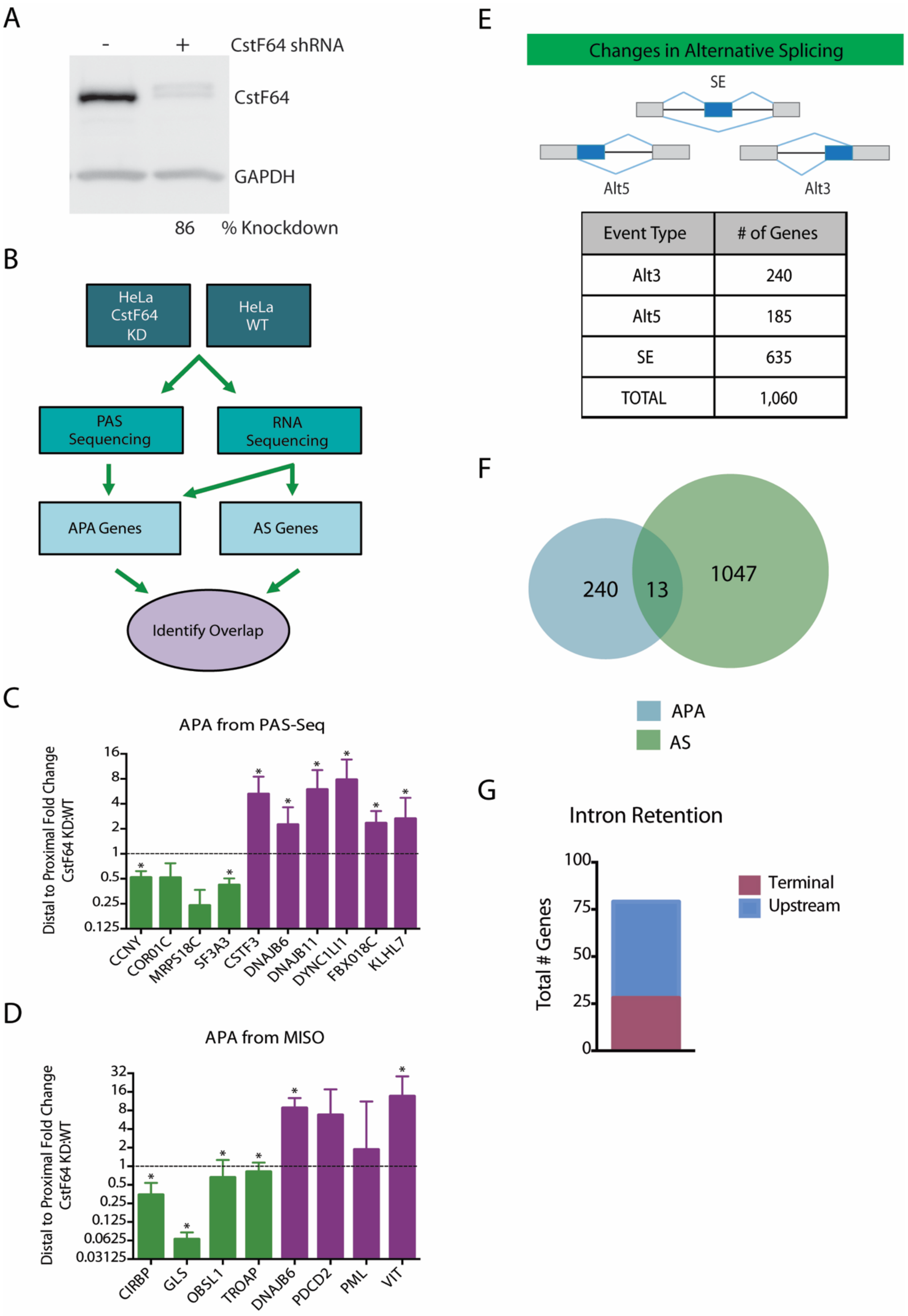
C
TSC22D2
1
2
3
4
1
2
3
4
KD-1
KD-2
Scrambled
Plasmid
Primer Pairs
1 + 2
256 bp
184 bp
KD-1
Scrambled
3 + 4
203 bp
KD-1
Scrambled
1 + 4
374 bp
D
TMEM135
1
2
3
4
5
1
2
3
4
5
13
14
KD-1
Scrambled
KD-1
Scrambled
KD-1
Scrambled
Primer Pairs
2 + 3
1 + 5
4 + 5
1
14
1222 bp
216 bp
150 bp
142 bp
E
SLC38A2
1
2
3
4
5
1
2
3
4
5
14
15
KD-1
Scrambled
KD-1
Scrambled
KD-1
Scrambled
Primer Pairs
2 + 3
1 + 5
4 + 5
1
15
1483 bp
266 bp
187 bp

Genome-wide approach to correlate APA and upstream AS events.

The major limitation of the approach described above is the small number of confirmed APA cases. To increase the power of the correlative analysis, we took advantage of a HeLa CstF64 KD cell line and used genome-wide approaches to identify APA and AS events [110]. After verifying >80% knockdown efficiency in CstF64 protein levels (Figure 4.4A) APA events were identified using poly(A) site-sequencing (PAS-Seq) [99] and standard mRNA-Seq on both CstF64 KD and wild-type HeLa cells (Figure 4.4B). Subsets of these APA events were verified using RT-qPCR analysis (Figure 4.4C-4.4D, Figure 4.5).

Figure 4.4. Genome-Wide Approach for Identification of a Larger Sample Pool for Correlation Studies.

A) Western blot demonstrating CstF64 KD in HeLa cells. B) Workflow for genome-wide analysis of CstF64 KD. To identify examples of APA genes, PAS-Seq and RNA deep sequencing was performed on both CstF64 KD and wild-type HeLa cells. APA and AS genes were then identified and the genes from each group were compared to each other to identify any overlap. C) Fold change of the ratio of distal to proximal poly(A) sites between CstF64 KD and wild-type (WT) HeLa cells depicted for 10 genes identified by PAS-Seq, as determined by RT-qPCR, n=3, *p<0.05. Genes with a shift in APA from distal to proximal (purple bars) and proximal to distal (green bars) are depicted. D) Fold change of the ratio of distal to proximal poly(A) sites between CstF64 KD and WT HeLa cells depicted for 8 genes identified by MISO, as determined by qPCR, n=3, *p<0.05. Genes with a shift in APA from distal to proximal (purple bars) and proximal to distal (green bars) are depicted. E) Alternative splicing analysis identifies CstF64 as a potential regulator of AS. Alternative splicing analysis of RNA-Seq data identified a large number of significant AS changes upon CstF64 KD in HeLa cells (1,060 genes), indicating CstF64 as potential regulator of AS. F) Overlap between AS and APA genes. Out of the 1,060 AS genes that were identified and 251 APA genes, only 13 genes were contained in the list of APA genes. G) Number of CstF64 KD induced genes with intron retention located at the terminal intron (maroon) or upstream of the terminal intron (blue).



Event Type	# of Genes
Alt3	240
Alt5	185
SE	635
TOTAL	1,060

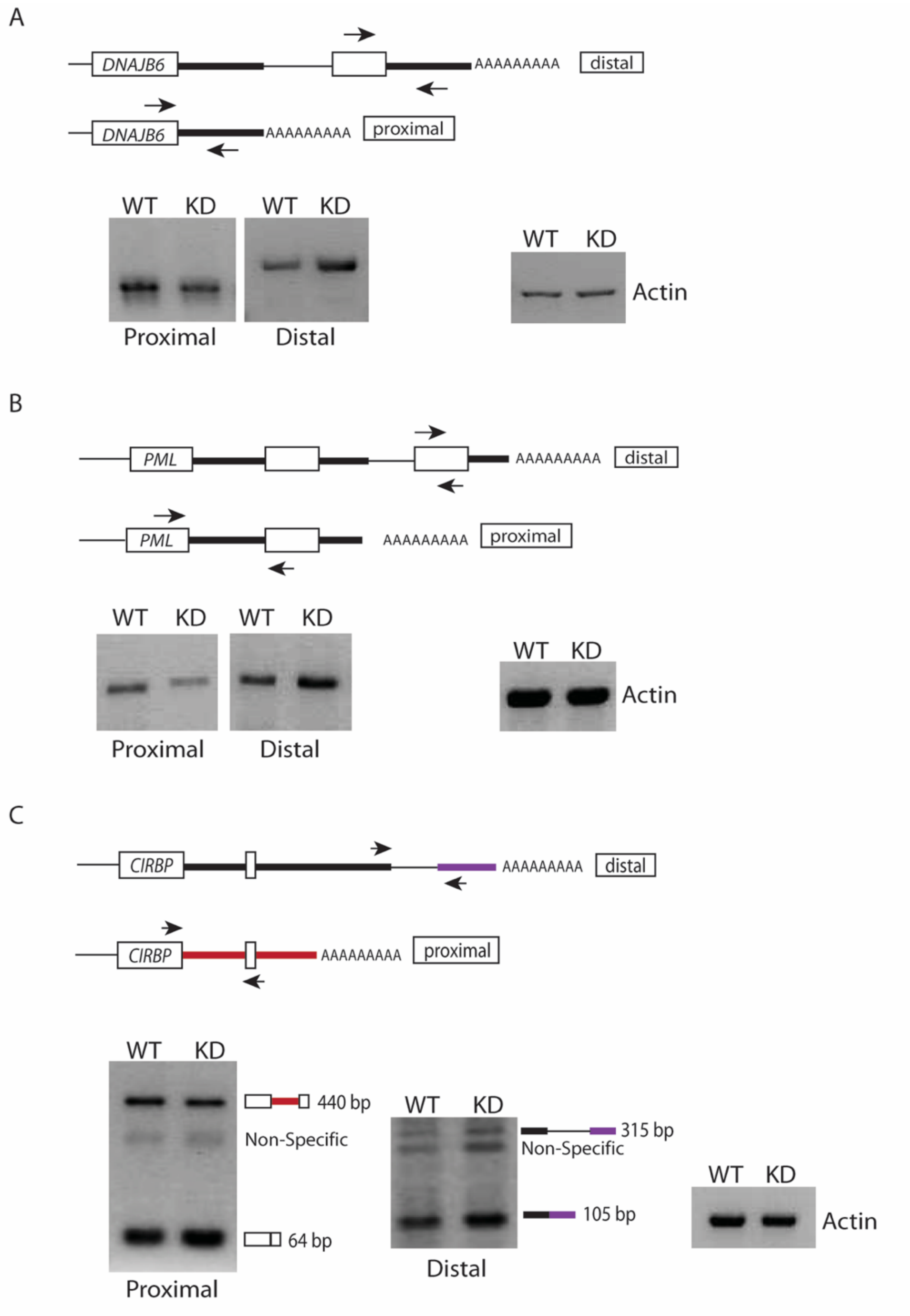


Figure 4.5. Impact of CstF64 KD on APA.
CstF64 KD leads to changes in APA as identified by PAS-Seq and RNA-Seq (MISO). RT-PCR analysis is depicted for three of the genes identified by MISO to contain APA events: DNAJB6 (A), PML (B), CIRBP (C) for wild-type (WT) and CstF64 KD (KD). Forward and reverse arrows depict primer location and direction, white boxes represent exons, thick black or colored lines represent the 3'UTR and thin black lines represent introns.

PAS-Seq and its applied bioinformatics pipeline identified 19 genes displaying statistically significant changes in APA. Differential splicing analysis of RNA-Seq datasets (by MISO [179]) derived from CstF64 KD and wild-type HeLa cells identified 234 additional statistically significant alternative last exons (ALE), which also represent APA events. Furthermore, MISO analysis identified 1,060 genes that display statistically significant Alt3, Alt5 and SE alternative splicing events (Figure 4.4E). These results suggest that in addition to mediating APA, CstF64 also exerts a regulatory role on AS.

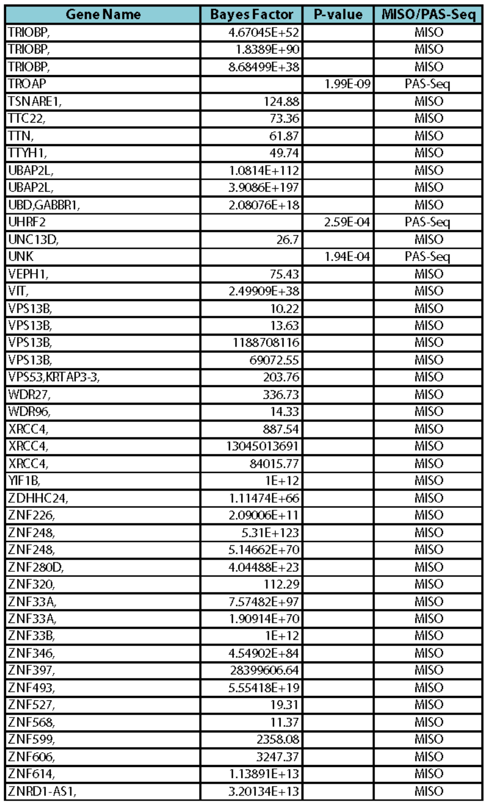
To determine the frequency of 'splicing and polyadenylation linked' events, the overlap between APA and AS events was determined. APA events detected from PAS-Seq (19 genes) and RNA-Seq (234 genes), a total of 253 genes (Table 4.1), were then correlated with all types of AS events. Interestingly, of the total 1,060 AS genes that were identified, only 13 genes were contained in the list of APA genes (Figure 4.4F, Table 4.2). This minimal overlap does not change when the statistical stringency is relaxed (see Materials and Methods). Given the small overlap between genes that display significant changes in APA and AS, we conclude that there is no general mechanistic coupling between APA and upstream AS in the conditions tested here. While this analysis does not exclude the possibility that a mechanistic link between the processing events can exist, for the majority of the alternative pre-mRNA processing events evaluated, APA and upstream AS are carried out independently.

Table 4.1. List of all APA Genes/Events as Derived from MISO and PAS-Seq Combined. Duplicate gene names in column A represent more than one event for that particular gene. Bayes factor for MISO generated events and p-values for PAS-Seq generated events are as listed for their respective genes/events.

Gene Name	Bayes Factor	P-value	MISO/PAS-Seq
ABL2,	46.69		MISO
ADAM33,	2309.5		MISO
ADAM33,	2267406.27		MISO
AGAP3,AX747175,	2.9456E+159		MISO
AHSG,	425.57		MISO
AIM1L,	1545.08		MISO
AK7,	81256519.72		MISO
AKAP6,	12.24		MISO
ALG9,	2.32975E+90		MISO
ANK3,	11.13		MISO
ANKRD11,	3.30448E+61		MISO
ARHGAP25,	962521.51		MISO
ARHGEF1,	1.04551E+37		MISO
ARHGEF1,	6.1045E+18		MISO
ARNTL,	29.68		MISO
ARSD,	36063.19		MISO
ATG16L2,	1.96536E+63		MISO
ATXN7,	1111399.85		MISO
ATXN7,	16.51		MISO
AX748058,AK8,	380.17		MISO
AX748291,ANKRD11,	2.44071E+36		MISO
BBS7,	215.69		MISO
BBS7,	1.33304E+13		MISO
BCAN,	25.63		MISO
BCAP29,	3.04052E+41		MISO
BCCIP,	5.7814E+137		MISO
BCCIP,	4.4355E+213		MISO
BCCIP,	2.0182E+178		MISO
BTAF1,	38.99		MISO
BTBD7,	12.01		MISO
C15orf27,	142563808.8		MISO
C17orf57,	823.13		MISO
C20orf112,	25944.05		MISO
C2orf42		1.50E-04	PAS-Seq
C5orf41,	13363554.58		MISO
C5orf56,	1.68573E+36		MISO
C6orf48,	1E+12		MISO
CARD14,	2100.57		MISO
CARD8		5.21E-05	PAS-Seq
CCDC114,	1.19865E+15		MISO
CCDC13,	501.65		MISO
CCDC157,	10625497.12		MISO
CCDC80,	22704613.48		MISO
CCDC88B,	3.14032E+72		MISO
CCDC88B,	341402.61		MISO
CCDC88B,	4.5139E+130		MISO
CCDC88C,	7.24947E+46		MISO
CCNY		9.00E-05	PAS-Seq
CD4,	79.09		MISO
CDH23,	36.4		MISO
CDH23,	19.95		MISO
CDH23,	607.95		MISO
CDH23,	16.04		MISO
CDH23,	6225.18		MISO
CDKL3,	12472.67		MISO
CEP63,	76.88		MISO
CEP85L,	14.08		MISO
CFLAR,	13.09		MISO
CFLAR,	10.88		MISO
CFLAR,	143.97		MISO
CFLAR,	71.44		MISO
CIRBP,	6.5308E+146		MISO
CLCN6,	21.05		MISO
CLK4,	19.84		MISO
COCH,	278300183		MISO
COL11A2,	40562479.86		MISO
COL11A2,	104934804.7		MISO
COL11A2,	15999.85		MISO
COL11A2,	3635.58		MISO
COL24A1,	129.85		MISO
CORIN,	14.11		MISO
CRB1,	66.63		MISO
CRYZL1,DONSON,	468.78		MISO
CSNK1E,	555903.15		MISO
CSNK1E,	84.06		MISO
CTCFL,	513.72		MISO
DENND4C,	4.34619E+50		MISO
DENND4C,	8.11649E+45		MISO
DIP2A,	16.38		MISO
DKFZp686B07190,ANKRD10,	6.5159E+186		MISO
DKFZp686B07190,ANKRD10,	3.3877E+228		MISO
DNAH6,	169.95		MISO
DNAJB11		6.18E-06	PAS-Seq
DNAJB6,		1.03E-21	PAS-Seq
DNAJB6,	1E+12		MISO
DNHD1,	5.10858E+11		MISO
DNHD1,	1.17291E+25		MISO
DNHD1,	6.95347E+15		MISO
DNHD1,	8.75496E+11		MISO
DNHD1,	6.20808E+18		MISO
DNHD1,	2.50026E+12		MISO
DOCK8,	125.75		MISO
DPP8,	20.29		MISO
DYNC1LI1		8.79E-07	PAS-Seq
EFNB1,	44893.21		MISO
EIF4E2,	1.4443E+126		MISO
ENDOV,	10.01		MISO
EPGN,	988.41		MISO
ERICH1,	95104.08		MISO
ESR1,	36241.43		MISO
EXD3,	236.34		MISO
EXD3,	12080.71		MISO
FAH,	4.95645E+21		MISO
FAM129C,	7401537.64		MISO
FAM135A,	10.04		MISO
FANCD2,	37.98		MISO
FEZ1,	43.41		MISO
FHAD1,	1121.1		MISO
FHAD1,	1.34188E+33		MISO
GANC,CAPN3,	2.19307E+13		MISO
GLS,	1E+12		MISO
GLS,	1E+12		MISO
GON4L,YY1AP1,	182.92		MISO
GPI		6.60E-13	PAS-Seq
GPNMB,	1712.18		MISO
GRIN1,	24936313.16		MISO
GRIN1,	7018.18		MISO
GSTM4,GSTM2,	8.2068E+178		MISO
GTF2H2,	3323673.85		MISO
GTF2H2,	585.48		MISO
GTF2H2B,SMA4,GTF2H2,AHRR,	5.66947E+13		MISO
GTF2H2B,SMA4,GTF2H2,AHRR,	52.03		MISO
GTF2H2C,GTF2H2D,	37.8		MISO
GTF2H2C,GTF2H2D,AHRR,	50504.35		MISO
GTF2H2C,GTF2H2D,AHRR,	1640491902		MISO
GTF2IRD2,	16.34		MISO
HDAC9,	4.37808E+49		MISO
HDAC9,	39309016227		MISO
HEATR7A,	8.61895E+19		MISO
HELQ,	11.03		MISO
HERC4,	7.50863E+35		MISO
HES2,	1.07133E+29		MISO
HIRA,C22orf39,	7.4505E+271		MISO
HM13		1.91E-15	PAS-Seq
HNF1A,	54.93		MISO
HSD3B2,	219.66		MISO
IDS,	32.01		MISO
IQCH,	92.25		MISO
IQCH,	2822.33		MISO
IRX5,	105371.97		MISO
JAK3,	1.28039E+21		MISO
JMJD5,	12506.71		MISO
KALRN,	38.33		MISO
KALRN,	101819.58		MISO
KCNH2,	18266.98		MISO
KDM5C		7.85E-05	PAS-Seq
KIAA0513,	9.95149E+11		MISO
KIAA1841,	1054556.78		MISO
KLHL7,	1358616378		MISO
KLK4,	142.02		MISO

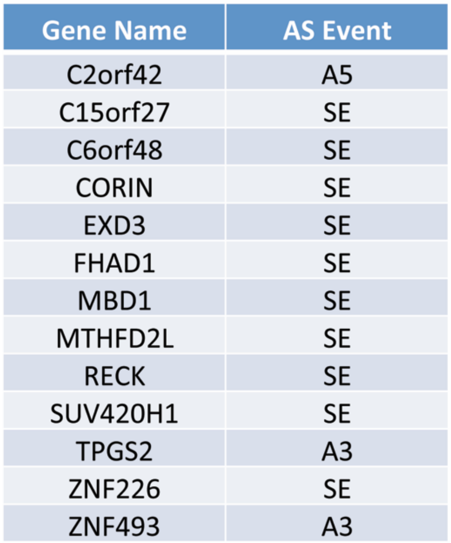
Gene Name	Bayes Factor	P-value	MISO/PAS-Seq
KRTAP3-2,	32.52		MISO
KTN1,	106771095.8		MISO
KTN1,	28.59		MISO
LARS		2.42E-06	PAS-Seq
LDB3,	336.76		MISO
LEPR,	61.31		MISO
LILRA6,LILRB3,	84.95		MISO
LTA,	14		MISO
MAN2C1		1.55E-04	PAS-Seq
MAPK12,	5.22971E+11		MISO
MARK4,	2052648.48		MISO
MAST1,	9.33035E+31		MISO
MBD1,	14.96		MISO
MEF2BNB,MEF2BNB-MEF2B,MEF2B,	4.54233E+20		MISO
MEF2BNB,MEF2BNB-MEF2B,MEF2B,	8.89524E+38		MISO
MEIS3,	13790.46		MISO
MEIS3,	1.30175E+13		MISO
METTL21A,	1.06762E+56		MISO
METTL6,	1217.86		MISO
MORN4,	2260.69		MISO
MORN4,	269.35		MISO
MRPL2		2.44E-04	PAS-Seq
MST1,	24.37		MISO
MTHFD2L,	3.65923E+11		MISO
MTL5,	6.46447E+11		MISO
MTL5,	2.849E+102		MISO
MTMR7,	6806.13		MISO
MXD3,	1E+12		MISO
MYH11,	2.01873E+12		MISO
MYH11,	6.68857E+11		MISO
MYH11,	16764.79		MISO
MYH7B,	147201.25		MISO
MYO3B,	89.28		MISO
MYRIP,	95.72		MISO
N4BP2L2,	4.22812E+20		MISO
N4BP2L2,	5.5132E+106		MISO
NBPF9,NOTCH2NL,NBPF14,	1015646851		MISO
NFASC,	775468.17		MISO
NFATC1,	164.37		MISO
NFATC4,	21277.44		MISO
NOTCH4,	156.45		MISO
NOTCH4,	57.85		MISO
NPHP1,	495.74		MISO
NRG1,	890351935.8		MISO
NRG1,	13.41		MISO
NRG1,	3217853463		MISO
NRG1,	5.43208E+30		MISO
OAS1,	22.5		MISO
OBSCN,	39.63		MISO
OBSCN,	11.53		MISO
OBSL1,	1.0847E+187		MISO
OBSL1,	1424791587		MISO
OTUD7A,	128259.32		MISO
P2RX5,	8.55288E+34		MISO
P2RX5,	1E+12		MISO
PATZ1,	2.70593E+44		MISO
PCDH11X,	17072.65		MISO
PCDH11X,	7252710.6		MISO
PCDHA2,PCDHA1,	44.85		MISO
PDCD2,		1.28E-04	PAS-Seq
PDCD2,	29.9		MISO
PDDC1,BC048998,	31.64		MISO
PEX6		2.08E-06	PAS-Seq
PKD1L2,	24.83		MISO
PLD1,	25.69		MISO
PML,	23097.72		MISO
PML,	148.65		MISO
PML,	9.6944E+132		MISO
PML,	208.06		MISO
PML,	2.86718E+14		MISO
PML,	1.20024E+11		MISO
PML,	193955.21		MISO
PML,	2.8E+275		MISO
PML,	1.02907E+23		MISO
PML,	1.3477E+245		MISO

Gene Name	Bayes Factor	P-value	MISO/PAS-Seq
PML,	1E+12		MISO
PML,	2.1065E+99		MISO
PPAPDC1B,	217.95		MISO
PPFIA3,	409.26		MISO
PPP2R1B,	9.45043E+37		MISO
PRDM2,	1.31292E+44		MISO
PRDM2,	3.60125E+49		MISO
PWWP2A,	3.5365E+270		MISO
RANBP17,	196.42		MISO
RAPGEF3,	2148.81		MISO
RAPGEF3,	31.98		MISO
RAPGEF3,	10.04		MISO
RBPMS,	841092.05		MISO
RBPMS,	52461041.37		MISO
RECK,	4.1122E+141		MISO
RECK,	9.0009E+102		MISO
RFC3,	6.2153E+195		MISO
RGL3,	7.80153E+22		MISO
RGNEF,	13.04		MISO
RGS12,	22.87		MISO
RGS3,	30678462239		MISO
RIBC1,	19.06		MISO
RPGRIP1L,	11.27		MISO
RPS6KA5,	18.91		MISO
RTKN2,	15.42		MISO
S100PBP,	168.45		MISO
SCN9A,	28.88		MISO
SF3A3		2.36E-07	PAS-Seq
SIN3B,	6.73174E+23		MISO
SLC19A1		2.59E-04	PAS-Seq
SLC26A1,	1.27531E+18		MISO
SLC30A5,	6.24994E+24		MISO
SLC4A5,	164.93		MISO
SLC9B2,	65.85		MISO
SP100,	3931.8		MISO
SP100,	2.3754E+144		MISO
SP110,	967.08		MISO
SPAG16,	10.01		MISO
SPEF2,	6.1303E+15		MISO
SPHK2,	113140.98		MISO
SPTB,	1.274E+30		MISO
SPTBN4,	62.39		MISO
SPTLC3,	149800.66		MISO
SRCIN1,	3.209E+189		MISO
SRCIN1,	6.91288E+32		MISO
SRCIN1,	7.91518E+83		MISO
SRPK3,PLXNB3,	30.51		MISO
SRR,	2.74897E+12		MISO
SSPO,	22.43		MISO
STAG1,	13127.05		MISO
SUSD4,	26.25		MISO
SUV420H1,	3.2888E+204		MISO
SUV420H1,	6.46764E+14		MISO
SVEP1,	18.65		MISO
SYNE1,	926.85		MISO
SYNPO2,	3835393469		MISO
TAP2,	386272.97		MISO
TAP2,	791025.3		MISO
TAP2,	2.22741E+60		MISO
TAP2,	5.77132E+46		MISO
TAP2,	6.33415E+43		MISO
TAP2,	1.44779E+42		MISO
TAP2,	1.36449E+34		MISO
TAP2,	4.33649E+26		MISO
TCFL5,	2.63855E+38		MISO
TCP11L2,	3.69504E+18		MISO
THAP3,	9.2834E+123		MISO
THAP6,	49549655.77		MISO
TM4SF19,	8007.13		MISO
TMPRSS3,	8.60648E+28		MISO
TMTC4,	9.25835E+37		MISO
TNFRSF19,	5267064451		MISO
TPGS2,	9.1108E+260		MISO
TRIM4,	2.33192E+46		MISO
TRIOBP,	5.37175E+57		MISO



Gene Name	Bayes Factor	P-value	MISO/PAS-Seq
TRIOBP,	4.67045E+52		MISO
TRIOBP,	1.8389E+90		MISO
TRIOBP,	8.68499E+38		MISO
TROAP		1.99E-09	PAS-Seq
TSNARE1,	124.88		MISO
TTC22,	73.36		MISO
TTN,	61.87		MISO
TTYH1,	49.74		MISO
UBAP2L,	1.0814E+112		MISO
UBAP2L,	3.9086E+197		MISO
UBD,GABBR1,	2.08076E+18		MISO
UHRF2		2.59E-04	PAS-Seq
UNC13D,	26.7		MISO
UNK		1.94E-04	PAS-Seq
VEPH1,	75.43		MISO
VIT,	2.49909E+38		MISO
VPS13B,	10.22		MISO
VPS13B,	13.63		MISO
VPS13B,	1188708116		MISO
VPS13B,	69072.55		MISO
VPS53,KRTAP3-3,	203.76		MISO
WDR27,	336.73		MISO
WDR96,	14.33		MISO
XRCC4,	887.54		MISO
XRCC4,	13045013691		MISO
XRCC4,	84015.77		MISO
YIF1B,	1E+12		MISO
ZDHHC24,	1.11474E+66		MISO
ZNF226,	2.09006E+11		MISO
ZNF248,	5.31E+123		MISO
ZNF248,	5.14662E+70		MISO
ZNF280D,	4.04488E+23		MISO
ZNF320,	112.29		MISO
ZNF33A,	7.57482E+97		MISO
ZNF33A,	1.90914E+70		MISO
ZNF33B,	1E+12		MISO
ZNF346,	4.54902E+84		MISO
ZNF397,	28399606.64		MISO
ZNF493,	5.55418E+19		MISO
ZNF527,	19.31		MISO
ZNF568,	11.37		MISO
ZNF599,	2358.08		MISO
ZNF606,	3247.37		MISO
ZNF614,	1.13891E+13		MISO
ZNRD1-AS1,	3.20134E+13		MISO

Table 4.2. List of Genes with APA and Upstream AS Events.
13 genes contain an APA event (from PAS-Seq and RNA-Seq) and upstream AS event.



Gene Name	AS Event
C2orf42	A5
C15orf27	SE
C6orf48	SE
CORIN	SE
EXD3	SE
FHAD1	SE
MBD1	SE
MTHFD2L	SE
RECK	SE
SUV420H1	SE
TPGS2	A3
ZNF226	SE
ZNF493	A3

The results evaluating the impact of induced APA in *Timp2* (Figure 4.2D, 4.2E) suggested that terminal intron removal may be influenced by the knockdown of polyadenylation factors. To test whether the knockdown of CstF64 induced intron retention events at the terminal end, we determined the extent of terminally located RI. MISO identified 77 genes displaying significant differences in the efficiency of intron removal. Interestingly, 35% of these RI events were terminal in nature (Figure 4.4G, Figure 4.5C). These observations suggest that CstF64 KD preferentially impacts the removal efficiency of terminal introns, presumably through modulating interactions with the splicing and polyadenylation machinery at the 3' end of the pre-mRNA.

Alternative splicing analysis identifies CstF64 as a potential regulator of AS.

Interestingly, the large number of AS events that are detected upon CstF64 KD suggests that this polyadenylation factor may be a potential regulator of AS, either directly or through indirect effects. To test the hypothesis that CstF64 binding to the pre-mRNA directly influences AS, we mapped CstF64 iCLIP-Seq reads [179] to the human genome and evaluated whether unique binding signatures could be identified around alternatively spliced exons induced by CstF64 KD. Compared to control groups no significant differences in CstF64 binding densities were observed in upstream intronic binding surrounding Alt3 (Figure 4.6A) or downstream intronic binding of Alt5 (Figure 4.6B). However, the upstream intronic region around alternatively included exons displayed increased CstF64 binding coverage, as demonstrated by the horizontal peak shift (Figure 4.6C). No measurable differences were detected for introns downstream of alternatively included exons (Figure 4.6D).

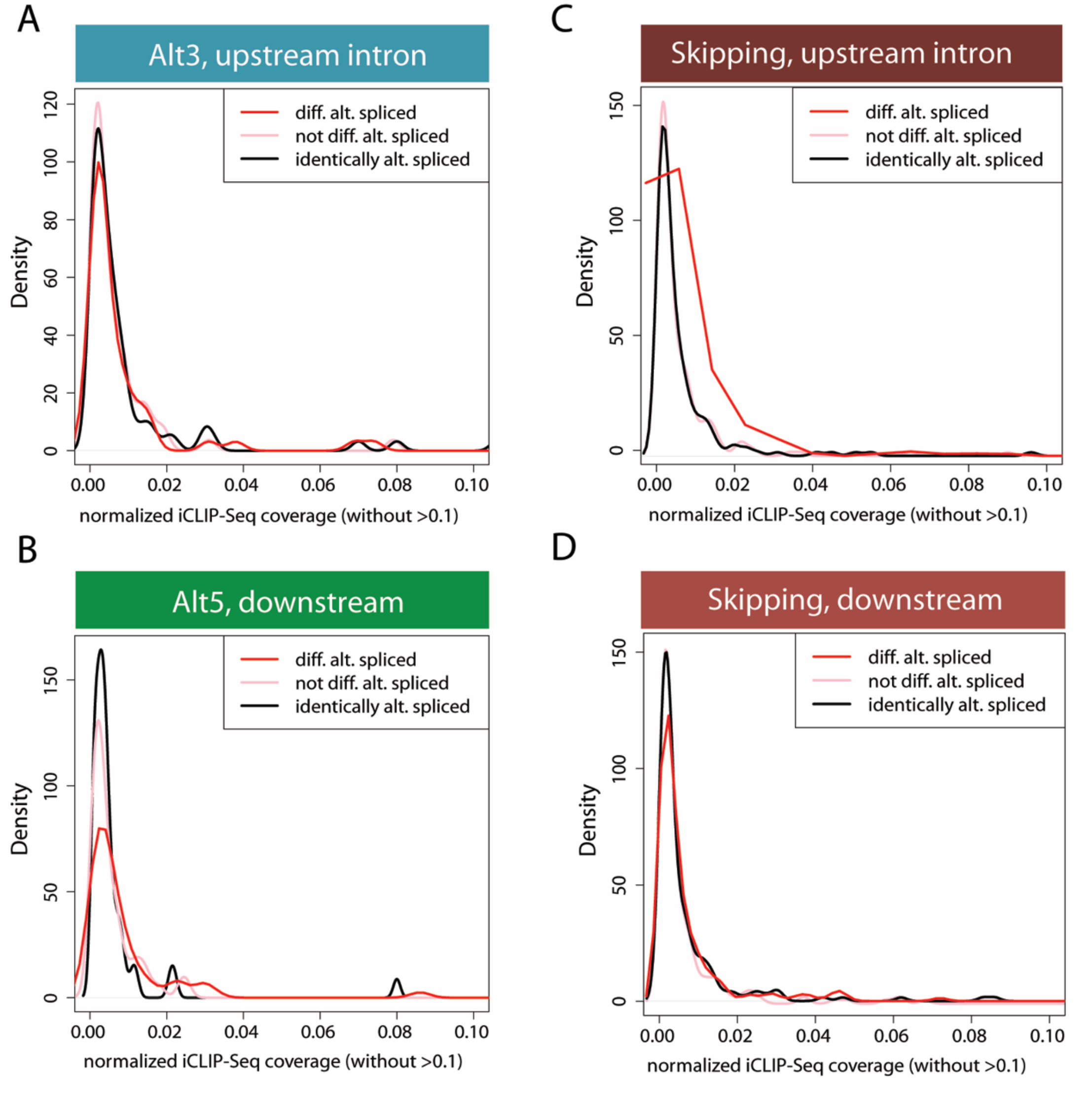


Figure 4.6. Density Distribution of CstF64 Binding in the Vicinity of AS Exons.
To distinguish the location of intronic binding of CstF64, iCLIP-Seq coverage was analyzed. The density of the distribution of normalized iCLIP-Seq coverage within introns surrounding alternative exons was determined. A) Intronic CstF64 binding distribution upstream of an alternatively spliced 3' splice site (Alt3). B) Intronic CstF64 binding distribution downstream of an alternatively spliced 5' splice site (Alt5). C) Intronic CstF64 binding distribution upstream of an alternatively skipped exon. D) Intronic CstF64 binding distribution downstream of an alternatively skipped exon. "not differentially alternatively spliced" and "identically alternatively spliced" are control categories. The first control group includes exons that are not in the group of "differentially alternatively spliced" exons. The second control group ("identically alternatively spliced") includes exons specifically filtered for no change in their splicing behavior. The plots show the fraction of events in those exon sets/groups that overlap with iCLIP-Seq clusters.

Further analysis of the increase in CstF64 binding coverage around upstream introns of alternatively included exons identified the gene hnRNP A2/B1, which displayed strong CstF64 binding signals at APA sites (Figure 4.7A). CstF64 KD leads to a decrease in hnRNP A2/B1 protein levels (Figure 4.7B), which correlates with splicing changes within its 3'UTR (Figure 4.7C). There are three potential APA sites within the 3'UTR of hnRNP A2/B1 and selection of the proximal APA site leads to the formation of hnRNP A2 with no discernable difference between WT and CstF64 KD RNA levels. However, selection of the most distal APA site generates multiple splicing isoforms, one of which is the hnRNP B1 mRNA isoform, which is targeted by NMD [180]. Thus, CstF64 KD leads to an increase in the less stable hnRNP B1 mRNA isoform, providing a molecular basis for the reduced protein levels observed (Figure 4.7B). These observations suggest that AS changes within the 3'UTR of hnRNP A2/B1 lead to the activation of different APA sites, resulting in the generation of different terminal exons (Figure 4.7D). As hnRNP A2/B1 has also been implicated in the regulation of AS [181], it is highly possible that many of the observed AS changes activated upon CstF64 KD were caused by the altered expression of this or other splicing regulatory proteins. Indeed, our analysis of CstF64 KD cells also identified hnRNP AB, hnRNP C, hnRNP H3, SRSF5 and SRSF6 as AS target genes (GEO submission GSE79157).

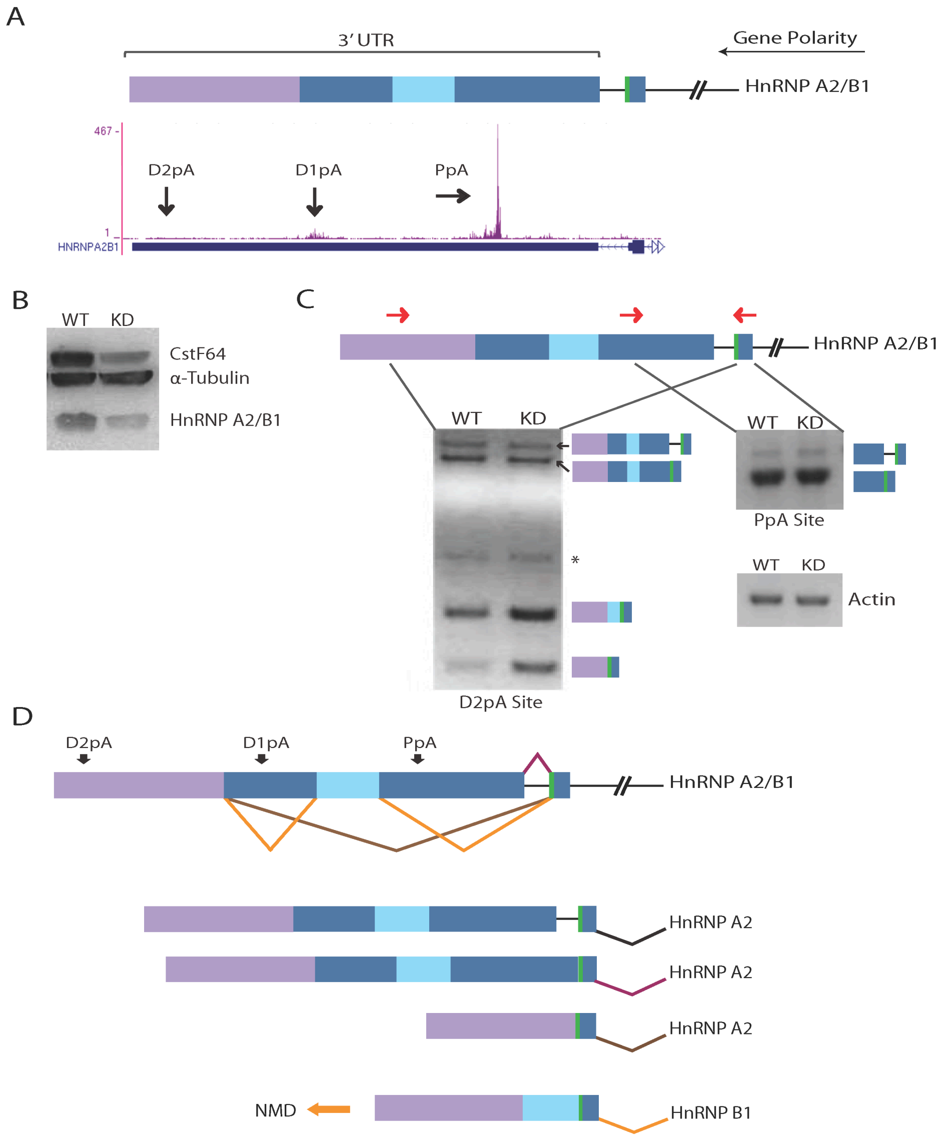


Figure 4.7. CstF64 Binding Within the 3'UTR of HnRNP A2/B1 Gene Triggers AS.
A) Cartoon depicts alternatively spliced regions (blue, light blue, purple boxes) located in the 3'UTR of the hnRNP A2/B1 gene. The polarity of the gene is from right (5' end) to left (3' end) as depicted by the long black arrow. The green bar depicts a stop codon. Snapshot from the UCSC genome browser depicts iCLIP-Seq track information for the binding of CstF64 and black arrows show the location of the proximal poly(A) site (PpA) and two distal poly(A) sites (D1pA and D2pA). B) Western blot analysis demonstrating loss in hnRNP A2/B1 protein levels due to CstF64 KD. C) RT-PCR analysis of the hnRNP A2/B1 3'UTR. Primer locations are depicted as forward and reverse red arrows. The identities of the PCR products are indicated on the side of the representative gels. Asterisk (*) denotes a non-specific amplicon. D) Alternative splicing pattern in the 3'UTR of hnRNP A2/B1 depicting the generation of hnRNP B1 and hnRNP A2 mRNA isoforms. The orange arrow indicates the isoform (hnRNP B1) that undergoes nonsense-mediated decay (NMD) [180].

To test the hypothesis that the reduced expression of hnRNP A2/B1 at CstF64 KD conditions accounts for many of the observed AS changes, we determined the overlap between CstF64 knockdown-mediated AS differences and hnRNP A2/B1 knockdown-mediated AS differences [46]. Interestingly, about one third of all genes that are affected by CstF64 KD are also affected by hnRNP A2/B1 KD (Figure 4.8), thus supporting the proposal that CstF64's influence on the processing of hnRNP A2/B1 indirectly mediates changes in AS. In summary, our analysis demonstrates that CstF64 pre-mRNA binding does not correlate with induced AS. Rather, the selection of alternative poly(A) sites or the alternative splicing of several splicing regulators is altered upon CstF64 KD, suggesting that most AS differences observed are likely to be indirect effects of CstF64 KD.

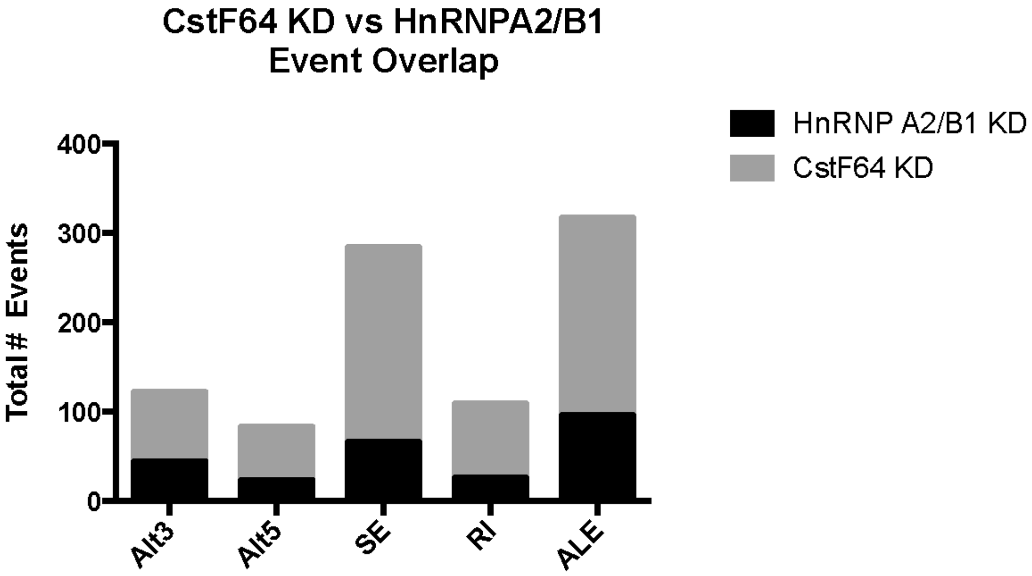


Figure 4.8. Overlap Between CstF64 and HnRNP A2/B1 AS Events.
Overlap in MISO derived AS events identified between CstF64 KD and hnRNP A2/B1 KD. One third of genes listed as having an AS event in CstF64 KD are contained in the list of genes affected by hnRNP A2/B1 KD.

Discussion

We tested the hypothesis that APA and upstream AS events are mechanistically coupled. APA was induced by the knockdown of the polyadenylation factors CFIm25 or CstF64. We then tested whether genes with APA also undergo AS upstream of the terminal exon. Using global approaches to increase the number of cases investigated, it was demonstrated that upstream AS and APA events occur independently from each other and that the major mechanistic coupling between APA and pre-mRNA splicing was limited to terminal exon processing. Direct interactions between the splicing and polyadenylation machineries have been documented previously [182], as demonstrated by U1 snRNP preventing premature cleavage and polyadenylation [122] or the inhibition of terminal exon splicing and polyadenylation through mutations of the AAUAAA consensus poly(A) signal or the 3' splice site [66, 183]. In combination, these observations strongly suggest that the mechanistic coupling between APA and pre-mRNA splicing is generally limited to terminal exon definition.

The lack of a 5' splice site at the 3' end of the pre-mRNA dictates the need for alternative mechanisms to aid in the definition of the terminal exon. However, in the case of AS events upstream of the terminal exon, polyadenylation does not appear to exert a major influence on decision-making processes. This may be because each internal exon is efficiently recognized through the combinatorial contributions of 3' and 5' splice site interactions with the splicing machinery. Furthermore, kinetic barriers could limit interactions between the splicing and polyadenylation machinery to terminal exons as upstream exons are synthesized prior to polyadenylation sites. Thus, the "first come, first served" concept [184–186] may have a significant influence on allowing productive interactions between splicing factors recruited to upstream exons and the polyadenylation machinery assembled at the 3’ end. Moreover, it should also be noted that although transcription occurs in the 5’ to 3’ direction, not all splicing occurs in that order as demonstrated by the adenine phosphoribosyltransferase (APRT) gene where the first of four introns is the last to be removed [187].

Given the significance of AS and APA in contributing to genome diversity, it is of considerable interest to determine whether these pre-mRNA processing events are functionally linked beyond the previously demonstrated definition of terminal exons. It is known that splicing defects play a role in many human diseases [35, 132], including numerous types of cancer [133–136]. Similarly, APA has a demonstrated regulatory role in human disease [18–22]. Thus, coordinated APA and upstream AS could generate alternatively spliced mRNA isoforms with unique 3’UTRs that dictate mRNA half-lives and protein product functions (Figure 4.1, right arm). Our genome-wide analysis demonstrates that the mechanistic coupling between APA and pre-mRNA splicing is limited to terminal exon definition (Figure 4.1, left arm). Thus, the regulation of mRNA isoform stability and translatability is mainly driven by the generation of different 3’UTRs.

Our results also identified an intriguing role for CstF64 in mediating AS. As shown, reduced levels of CstF64, led to a considerable change in AS events. However, neither enrichment nor depletion of CstF64 binding to alternatively spliced exons was observed when compared to control groups, arguing against a direct role. Instead, hnRNP A2/B1 was identified to be one of several splicing regulatory proteins that alter APA or AS patterns upon CstF64 KD. Given hnRNP A2/B1’s important role in regulating AS [181] and mRNA stability [180, 188], these observations strongly suggest that CstF64 influences AS decisions through indirect mechanisms. Investigating the mechanisms that lead to 3’ UTR diversification and alternative terminal exon definition will therefore be an important future research endeavor.

Materials and Methods

Cell Culture and Transfections

CFIm25 KD was performed using shRNA constructs as previously described.[114] The siRNA-408 sequence (5’-GCAAUCGUCAAUGACCCAGUCUUGC-3’) was used as a target sequence in the design of an shRNA insert oligo and was cloned into the pSUPERIOR.puro vector (Oligoengine, VEC-IND-0006). To create the shRNA insert, the sense oligonucleotide 5’-GATCCCCGCAAGACTGGGTCATTGACGATTGCTTCAAGAGAGCAATCGTCAATGACCCAGTCTTGCTTTTTA-3’ was annealed with the antisense oligonucleotide 5’-AGCTTAAAAAGCAAGACTGGGTCATTGACGATTGCTCTCTTGAAGCAATCGTCAATGACCCAGTCTTGCGGG-3’ and ligated to linearized vector using T4 DNA ligase (Promega). HeLa cells were cultured in MEM and transfected with 1 μg of either CFIm25 KD plasmid or a control plasmid using Lipofectamine-2000 (Invitrogen). Media was replaced with MEM supplemented with 10% FBS and 3 mg/mL puromycin post transfection. RNA extraction was performed via Trizol and proteins were isolated with RIPA buffer.

CstF64-RNAi HeLa cell lines were obtained [110] and grown in high glucose DMEM media plus 10% FBS, 1% Na-Pyruvate and 1.5 μg/mL of puromycin for selection of stably transfected cells. RNA extraction was performed via Trizol and proteins were isolated with RIPA buffer.

Western Blot Analysis

CFIm25 and CstF64 proteins were separated on 10% SDS-PAGE and subjected to standard western blotting procedures. Primary antibodies used were as follows: anti-rabbit NUDT21 (Proteintech Group, 10322-1-AP), anti-rabbit CstF64 (Bethyl Labs, A301-092A), anti-mouse hnRNP A2/B1 (AbCam, DP3B3) and anti-mouse α-Tubulin (Calbiochem, DM1A). Typically, 75ug of total protein were loaded per lane.

3' RACE and RT-PCR

3' RACE was performed to analyze 3' ends, as previously described [189]. Total RNA was reverse transcribed using the primer 5'-CCAGTGAGCAGAGTGACGAGGACTCGA GCTCAAGCTTTTTTTTTTTTTTTTT-3'. *TIMP-2* gene specific primers used were: forward primer 1 5'-CGCAACAGGCGTTTTGCAAT-3' for proximal poly(A) tail product amplification, *TIMP-2* forward primer 2 5'-CTGTTCGCTTCCTGTATGGT-3' for distal poly(A) tail product amplification [114], and reverse primer 5'-CCAGTGAGCAGAGTGACG-3'.

For RT-PCR analysis, total RNA was treated with DNase I and reverse transcribed using iScript (BioRad). For the *TIMP-2* gene, terminal intron retention was evaluated using forward primer 5-AGGGAAGCACACCTGCAGTA-3' with reverse primer 5-GTGCCCGTTGATGTTCTTCT-3'. In addition to *TIMP*-2, five additional genes with documented cases of AS events were analyzed using RT-PCR. Three AS events were evaluated for each of the following genes: *DSTN, TSC22D2, TMEM135, SLC38A2,* and seven AS events for *ERCC6*. Primer sequences are available upon request. PCR reactions were performed using Taq polymerase under standard conditions and resolved on a 0.8% agarose gel stained with ethidium bromide.

PAS-Seq APA events were verified by RT-qPCR as described previously [110]. RNA-Seq MISO designated AS events in the 3'UTR for *hnRNP A2/B1* were evaluated in addition to APA events for eight genes: *CIRBP, DNAJB6, GLS, OBSL1, PDCD2, PML, TROAP, VIT*. Primer sequences are available upon request. Total RNA from CstF64 KD cells were treated with DNaseI and reverse transcribed using Oligo(dT) primers with SuperScript II (Invitrogen). RT-PCR analysis was performed with Taq polymerase under standard conditions and resolved on a 1.5-2% agarose gel stained with ethidium bromide. Quantitative real-time PCR (qPCR) was performed with the same primer pairs and iTaq Universal SYBR Green supermix (BioRad) using a three step qPCR protocol with an annealing temp (Tm)=55°C and 1 minute extension at 72°C.

RNA Deep Sequencing

Cell lines stably transfected with shRNA plasmids targeting CstF64 were obtained from the Shi lab [110]. Cell lines were maintained by culturing in the presence of 1 mg/ml puromycin. Total RNA was isolated from CstF64 KD cells and wild-type control cells using Trizol according to manufacturer's protocols. CstF64 protein levels were measured via Western Blot analysis using rabbit anti-CstF64 antibody (Bethyl, A301-092A). Total RNA isolated from two biological replicates was used to generate independent libraries with the Illumina TruSeq RNA sample preparation kit according to manufacturer's protocols. Each library was diluted to approximately 10 pM prior to loading and sequenced using the Illumina GA IIx instrument generating 50 bp single-end reads. This produced ~50 million reads for each wild-type and CstF64 KD biological replicate.

Computational Analysis

Identification of AS and APA events. RNA-Seq results were trimmed by 10 bases at the 5' end. Trimmed 40 bp long reads were then mapped to the human genome (GRCh37/hg19) using TopHat (version 2.0.4) [190] with default parameters. To identify AS events, the datasets were compared using MISO [179]. To do so, replicate samples were combined, and only uniquely mapped reads were kept. For every known AS event expressed in both cell lines (SE, Alt3, Alt5, RI, ALE/APA) only events showing an absolute change in inclusion level or splice site usage ≥15% with a bayes factor of ≥10 were filtered for. Applying these filters resulted in the identification of 733 SE (636 genes), 267 Alt3 (240 genes), 200 Alt5 (185 genes), 109 RI (77 genes) and 328 APA (234 genes) events. To identify additional APA events, a PAS-Seq dataset was utilized comparing polyadenylation between CstF64 KD and wild-type HeLa cells [110]. PAS-Seq and its applied bioinformatics pipeline [110] identified 19 genes displaying statistically significant changes in APA, using a cutoff of ≥2 reads with FDR ≤0.05 and a change in APA ≥15% (a change in the ratio of proximal to distal shift in polyadenylation between CstF64 KD and wild-type HeLa cells).

Correlation between APA and AS events. To determine the frequency of ‘splicing and polyadenylation linked’ events, the overlap between APA and AS events was determined. APA events detected from PAS-Seq (19 genes) and RNA-Seq (234 genes), a total of 253 genes, were correlated with all types of AS events. Overlapping APA and AS genes were identified by determining those that only had AS events occurring upstream of both proximal and distal poly(A) sites. Of the 19 PAS-Seq derived list of APA genes, only one gene was located upstream of both proximal and distal poly(A) sites. Of the RNA-Seq derived list of APA genes, only 12 genes with AS events upstream of both poly(A) sites were identified.

Correlation between AS events and CstF64 binding. AS events were identified using the RNA-Seq datasets described above and MISO [179] filters of ≥20% change in the absolute level or splice site usage (for Alt3 or Alt5) with a bayes factor of ≥10 filtered for. Of these, events supporting only one of the two possible isoforms with coverage of ≥10 reads were kept for subsequent analysis. These remaining exons were considered to be "differentially alternatively spliced". A first list included cassette exons, which either demonstrated an increase or decrease in inclusion of the knockdown sample. A second and third list consisted of exons with an Alt3 or Alt5, respectively. For each of these three lists, two different control groups were created. The first control group for each set (denoted "not differentially alternatively spliced") included exons that did not pass the filter, but showed the same type of AS. The second control group (denoted "identically alternatively spliced") was filtered in an opposite fashion than the test sets: events showing a change in inclusion level or splice site usage of <0.05 and a bayes factor of <0.1, which meant a change in splicing behavior was 10 times less likely than having no change.

Data from triplicate CstF64-iCLIP-Seq in HeLa cells was obtained from the Gene Expression Omnibus database (accession no. GSE40859) [110] and was used to identify whether AS is correlated with CstF64 binding. The frequency of CstF64 binding sites was examined within the set of differentially alternatively spliced exons and their surrounding introns, as well as within the exons (and their neighboring introns) of the two control groups (not differentially alternatively spliced as well as identically alternatively spliced exons).

To separate real CstF64 binding sites from background peaks, clusters of iCLIP-Seq reads in close proximity to the alternatively spliced exon were searched for. A cluster was defined as a window of at least 10 nucleotides displaying at least two iCLIP-Seq tags after averaging all three iCLIP-Seq data sets. The clusters of CstF64 binding sites were overlapped with alternatively spliced exons and the neighboring introns in both the test and control sets to compute their total iCLIP-Seq coverage. Those values were then normalized by the length of the introns. The density of the distribution of the normalized iCLIP-Seq coverage in the different sets was plotted (excluding introns with a coverage of 0). All the density functions are only plotted for normalized coverage of ≤0.1.

Correlation between CstF64 KD and hnRNP A2/B1 KD. Data from duplicate HnRNP A2/B1 KD and five control HeLa RNA-Seq samples were obtained from the Gene Expression Omnibus database (accession no. GSE34992) [46]. To identify AS events, the datasets were compared using MISO [179]. Replicate control and hnRNP A2/B1 KD samples were combined and only uniquely mapped reads were kept. For every known AS event expressed in both cell types (SE, Alt3, Alt5, RI, ALE/APA) only events showing an absolute change in inclusion level or splice site usage ≥15% with a bayes factor of ≥10 were filtered for. Overlap between CstF64 KD and hnRNP A2/B1 KD was determined for all types of AS events (SE, Alt3, Alt5, RI, ALE/APA) by comparing which exact AS events were contained in both groups.

Data Accession Code. All sequencing and MISO data has been submitted to the National Center for Biotechnology Information Gene Expression Omnibus database (accession no. GSE79157).

CHAPTER 5

PERSPECTIVES

Evolutionary Nature of Exon Conservation

The mechanisms leading to nucleotide variation within species and how conservation has shaped vital processes needed for survival and adaptation are of great importance. The discovery of split genes has led scientists to a better understanding of gene evolution and to an increased appreciation of the vast complexity of gene expression. At its core, splicing is a fundamental process, and its study has led to the discovery of key elements important for maintaining genomic stability. Mutations within the genome have supported diversity and variations within species, however, these changes have also led to a number of human diseases. The potential to predict splicing-associated diseases based on sequence analysis is critical, not only for understanding splicing regulation mechanistically, but also for developing disease-appropriate therapeutic approaches.

The ultimate goal of generating an exon conservation database was to improve the prediction of splicing patterns of disease-associated SNPs. To this end, the first step was piecing together genomic connections between human and other vertebrate species. The exon conservation database allowed for an initial analysis of physical parameters that are conserved, representing key features that are necessary for gene expression. Of particular interest was the conserved nature of length and sequence among exons. Most *internal exons* are between 50-250nts long, an exon length shown to be optimal for splice site recognition. Exons that were longer in length (>250nts) were mainly represented by *last exons*, which are size conserved between human and primates only. Most *last exons* contain untranslated regions, which will significantly contribute to sequence and size variation between species. Thus, *last exons* are expected to undergo increased evolutionary drift, reducing overall size conservation.

Evaluation of the *internal exon*s that are longer in length (>250nts) that also display a population of low length-conserved exons could be explained as a newly emerged and an evolutionary young population. Given the non-optimal nature of these longer length exons, it is possible that they are still undergoing optimization for efficient recognition and splicing. It has been shown that new exons have a higher frequency of AS [191, 192]. Although, currently under investigation, it is possible that this population of longer *internal exons* also undergoes a higher degree of AS and may be composed of a large percentage of Alu elements. Alu elements are transposable genomic sequences that are created through the process of exonization and are found only in primates (and prosimians) [85]. This suggests that Alu elements have played an important role in the evolution of exon creation in primates. Moreover, it has been shown that Alu elements are very rare within constitutive exons, however, alternatively spliced exons show a high frequency of these fragments [82]. Therefore, it is probable that the population of primate only, size-conserved *internal exons* that are longer in length may have emerged through the process of the exonization of Alu fragments.

S*ingle exons* (or intronless genes) also revealed an interesting finding. The majority of *single exons* had low length conservation (<10). Interestingly, 30% of *single exons* are olfactory genes, which displayed moderate length conservation (10-40) but very low sequence conservation. Previous studies on single exons have suggested that these intronless genes have emerged relatively recently [193] and their emergence could be due to gene duplication or retroposition of mRNA [194]. Therefore, the low length-conserved nature of the majority of these intronless genes could be explained by the burst of retroposition during the sudden emergence of primate evolution [194]. The population of low length-conserved single exons was largely composed of genes involved in signal transduction pathways and metabolic processes, representing their importance for species survival. Furthermore, their lack of introns could represent the need for rapid processing of these genes to achieve quick expression [195]. Certain intronless genes have been shown to accumulate in the cytoplasm [196], however, their inability to splice would mean their export would not rely on splicing-dependent mRNA export. However, recent studies have shown that these intronless genes may rely on a different mechanism such as the use of the transcription/export (TREX) pathway for export to the cytoplasm [196]. For the 30% of intronless genes that were represented by olfactory genes, the moderate length-conservation of the genes contributes to their likely evolutionary significance, but, as primates evolved, the need for these genes was diminished and it is plausible that these genes have now been selected against and have started on their journey out.

The exon length and sequence conservation analysis revealed a tight relationship. High sequence conservation typically demonstrated higher length conservation, however, there were exceptions to this rule, as discussed in Chapter 2. These findings demonstrate an evolutionary advantage in maintaining both the size of an exon and, therefore, important distances between splicing elements, but also maintain important sequence elements, presumably important for protein coding and splicing regulation. Although sequence variation can play an important role in species to species genetic variation, certain sequences have been optimized and changes to these sequences can cause defects in splicing patterns leading to various genetic diseases [197, 198].

Nucleotide conservation has been previously demonstrated to be an important framework for increasing the predictive power of the splicing code [73, 199]. However, due to the co-evolving nature of the splicing and genetic code, it is difficult to separate sequence elements that overlap between these two. By limiting such an analysis to synonymous positions, it is possible to enrich for sequences that are already negatively selected for, in turn validating important evolutionary conservation patterns. Nevertheless, understanding whether certain sequence variations and mutations lead to changes in splicing requires access to a large dataset of disease-causing mutations with recorded splicing outcomes. Through the use of a publicly available library of disease-associated SNPs [145], it was possible to use the exon conservation database as a means to predict splicing patterns of size-conserved exons. Although the SNP data was mainly representative of non-synonymous nucleotide changes, presumably leading to amino acid changes, this dataset was a great starting point for testing the model of using exon size-filtered sequence alignments as a method for splicing prediction. It is widely accepted that variations at splice sites influence splicing, something that was also evident within this dataset. However, understanding whether this phenomenon applies to non-junction sites was of particular importance. Can sequence variations due to SNPs lead to predictive control of an exon's splicing potential? Through defined SNP category designations and analysis of the mutations across codon positions the greatest variability within SNP categories was observed at the wobble position, as expected. As discussed in Chapter 2, using exon size-filtered sequence alignments at the wobble position therefore allowed for a certain level of prediction on how and if a SNP influenced splicing.

There were several limitations within the dataset, including a large proportion of protein coding disease-associated SNPs and an underrepresentation of SNPs that cause changes in splicing at the wobble position. This could represent the detrimental impact of a SNP within the wobble position or the lack of importance of a SNP at that position. This also demonstrates the necessity for larger datasets focused on disease-associated SNPs at the third codon position to increase the predictive capability of this type of analysis to further define the role of SNPs at the wobble position.

Through the utilization of exon size conservation patterns and the important sequence requirements of trans-acting factors, greater insight into important sequence patterns can be obtained. The ability to determine the impact that a sequence of RNA will have on how an exon is spliced could aid in predicting the functional implication of a mutational variant that is introduced. This could enhance disease prediction and lead to the potential development of therapeutic capabilities. Predicting splicing defective mutational sites within an exon could aid in repair and replacement of a defective sequence within the pre-mRNA. For instance, spliceosome-mediated RNA trans-splicing (SMaRT) is a method used to replace an exon with a defective sequence [200]. In short, through the use of trans-splicing, an artificially engineered pre-mRNA trans-splicing molecule (PTM) with the correct sequence is introduced, and through spliceosome-mediated splicing, replacement and removal of the defective RNA sequence can be achieved. The potential implications of such a therapeutic method are vast. Thus, addressing allowable variations from defective variations within a codon through their conservation pattern could guide the predictive nature of splicing. The work described here is thus a stepping stone for further investigations into methods of using conservation patterns and synonymous mutations as a means of defining the roles that disease-associated SNPs have on splicing within the human genome.

Position-Dependent Mechanism of Regulatory Proteins

Many trans-acting factors aid in the recognition of an exon. Evaluating nucleotide conservation is one avenue of determining important splicing elements. However, understanding the nature of splicing regulation requires the assessment of SREs and their binding partners. SR proteins and hnRNPs are two classes of proteins that directly bind and regulate the pre-mRNA through SREs. In the past, these regulatory proteins were characterized by defined roles, SR proteins as splicing enhancers and hnRNPs as splicing silencers. However, their unilateral role was questioned by more recent evidence demonstrating a dual functionality, whereby both classes of proteins could enhance or repress splicing, contingent on their binding position within the pre-mRNA [44–49, 55]. The Hertel lab demonstrated position-dependent regulation of SR proteins and hnRNPs relative to the 5’ss [57]. This work also revealed a potential mechanism of repression whereby spliceosomal complex formation was stalled at E complex. SRSF7 and TIA-1 at repressive positions demonstrated E complex formation, however, subsequent formation of A complex was hindered. Based on these findings, it was of great interest to determine the mechanistic connection between spliceosomal complex formation and splicing efficiency. The stalling of E complex by SRSF7 and TIA-1 when bound in their repressive positions could be dependent on the kinetic changes between protein components of U1 snRNP. U1 snRNP initiates splicing through its base pairing interaction with the 5’ss to form E complex. Competition pulse chase experiments, through titration of U1 snRNP, demonstrated altered thermodynamics of the U1 snRNP components UI-70K and U1-C. In the presence of a repressor, U1-70K was observed to be less tightly associated with U1 snRNP. These results demonstrate that U1 snRNP is a dynamic complex, one that can adjust the integrity of binding interactions with its core proteins, U1-70K, U1-C, and U1-A. This dynamic U1 snRNP behavior could represent a checkpoint in regulating the splicing of the pre-mRNA it associates with. U1 snRNP is one of the most abundant spliceosomal snRNPs [201]. Not only does it initiate splicing, but it also participates in non-canonical splicing and polyadenylation activities, through interaction with pseudo splice sites and poly(A) sites [124, 165]. Given these multiple roles in RNA metabolism, there must be pathways to allow U1 snRNP to function in these different roles. It is possible that the U1 snRNP that initiates splicing is structurally or compositionally distinct from those that bind a pseudo-splice site or one that inhibits premature polyadenylation. These differences could lie within the complex integrity of U1 snRNP. Depending on the binding location of surrounding splicing regulators, U1 snRNP's compositional stability may be altered, in turn defining the role that it plays on the RNA it interacts with. For example, in the presence of an activator, U1 snRNP may be fully formed and tightly associated with U1-70K, U1-C, and U1-A. However, when a repressor is nearby, the associations between U1 snRNP components could be weakened, altering the integrity, stability and overall structure of U1 snRNP at the 5'ss.

Probing for other protein components of U1 snRNP, namely Sm proteins, would give a more accurate picture of the dynamic interactions that are occurring. U1 snRNP is a complex molecule and focusing on single protein components does not allow for a complete view of changes that occur. However, given that U1-70K association is vital for the stability and binding of other U1 snRNP proteins [202], this protein may be a key regulator of U1 snRNP integrity and changes in its affinity may be critical for the alterations that are observed.

The pulse-chase experiments that were performed focused on using a neutral chase RNA with a strong 5'ss to outcompete U1 snRNP binding for the RNAs of interest. However, other chase RNAs could be used that would allow for titrating SRSF7 or TIA-1 during the chase. Such additional experiments could lead to further insights into how relieving an activator or repressor impacts U1 snRNP recruitment and protein component affinities. The nature of the chase RNA used is critical for reliable measurement and interpretations of changes that are being observed. In the case of the neutral RNA sequence, experimental reproducibility was an issue. It is possible that the RNA chase sequence was not ideal or adding the chase RNA in excess was over-saturating the reaction leading to inconsistent results. It is also possible that the many wash steps necessary to reduce non-specific binding uncontrollably altered the final signal of the evaluated U1 components. Constant dilution through washing could also have aided in the loss of SRSF7 or TIA-1, alleviating the repressive/activating state of the RNA, thus altering how U1 snRNP components were interacting with each other.

At high concentrations of nuclear extract, which was important for visualization of U1-70K on western blots, there was binding of TIA-1 to the RNA containing binding sites for SRSF7 only and vice versa. Although the sequences were specific for their respective proteins, this demonstration of non-specific binding could have led to some of the ambiguous results that were being seen at times. It is possible that the fluctuating levels of U1-70K were due to the antagonistic behavior of SRSF7 and TIA-1 on the same RNAs. SRSF7 bound within the designated position with TIA-1 binding elsewhere could counter SRSF7's activity, preventing stable association of U1 snRNP components. To resolve these issues, S100 cytoplasmic extracts rather than nuclear extract could be used. S100 does not contain SR proteins and contains much lower levels of TIA-1 [203]. This would allow for addition of SRSF7 or TIA-1 as needed for each RNA being tested, without the worry for cross-interactions.

U1 snRNPs dynamic molecular composition may place this splicing initiator at the forefront of splicing initiation and regulation. The work done thus far demonstrates the importance of understanding how regulatory proteins interact with spliceosomal factors and how their directional biases guide the stability of key players of splicing initiation, namely U1 snRNP. The preliminary work is encouraging and supports further investigation into the role that U1 snRNP plays as a molecular checkpoint influencing splicing decisions of the pre-mRNA.

Correlation Between APA and AS

Splicing is one of several RNA processing events that collectively regulates eukaryotic gene expression. The synergy between splicing and polyadenylation increases the genome's coding potential and regulates the outcome of a gene's expression. APA, much like AS, has shown extensive contribution towards proteomic diversity and complexity. Therefore, understanding the functional connections between these two critical mRNA processing events is of great importance.

Coupling between polyadenylation and splicing has been shown for terminal exon recognition [66, 67, 118–124]. However, whether this phenomenon is also seen between APA and upstream AS events had not yet been elucidated. One scenario would describe AS and APA as two mechanistically uncoupled events carried out independently of each other.

This would result in the generation of spliced mRNA isoforms with a comparable distribution of variable 3'UTR lengths. However, if AS and APA are mechanistically coupled events, the mRNA isoforms that are generated would display a selective preference for one type of APA event with a specific AS event. The results of such a coupling could lead to a synergistic influence over the expression of one particular mRNA isoform over another. Therefore, the hypothesis that upstream AS events go hand in hand with a particular APA event was of particular interest.

Initial work focused on identifying a correlation between APA and AS through the knockdown of a polyadenylation factor known to influence APA, CFIm25 [111, 112, 114]. Through analysis of four genes via RT-PCR analysis, no mechanistic connection was seen between APA and upstream AS events. However, a drawback of this approach was the insufficient number of altered APA cases. To increase the power of analysis, RNA-Seq and PAS-Seq analysis was performed on CstF64 knockdown (KD) cells. Through this genome-wide analysis, mechanistic coupling between APA and pre-mRNA splicing was shown to be limited to terminal exon definition. However, CstF64 was demonstrated to be an important indirect regulator of AS. For example, CstF64 KD led to the formation of a less stable mRNA isoform of the gene hnRNP A2/B1, specifically the B1 isoform. HnRNP B1 is generated through the selection of a more distal APA site within its 3' UTR, which when used leads to nonsense-mediated decay. However, hnRNP B1 was shown to increase in expression through CstF64 KD. Therefore, it is highly possible that many of the AS changes that were observed upon CstF64 KD were caused by changes in the expression of hnRNP A2/B1 or other splicing regulatory proteins. These results demonstrated additional indirect regulatory connections between polyadenylation and alternative splicing factors.

One of the limitations of this study was the limited list of AS events, a consequence of the MISO database used. In addition the AS analysis software MISO was not able to take into account biological variations. Since the publication of our study, new AS analysis algorithms and models have been developed that improve the accuracy and sensitivity of previous methods such as, replicate MATS (rMATS) [204] and modeling alternative junction inclusion quantification (MAJIQ) [205]. These new tools may aid in detection of additional alternatively spliced events that were previously unidentified. Thus, it might be worthwhile to reanalyze the datasets using the most modern AS algorithms.

Through knockdown of CstF64, statistically significant APA events were identified, although at relatively low numbers. Thus, it may be possible that in other biological contexts, mechanistic links between APA and upstream AS do exist. Certain cells and tissues have been shown to favor certain APA isoforms [206, 207], and there are several other mRNA processing factors that have been shown to play important roles in APA regulation, including a paralogue of CstF64, CstF64τ [110], the poly(A) binding protein nuclear 1 protein (PABPN1) [108], polyadenylate-binding protein 1 (PABP1) [109], and Fip1, a component of the CPSF complex, identified as an important regulator of APA in embryonic stem cells [21]. A deeper analysis of the impact tissue specificity or any of these additional polyadenylation factors have on the correlation between APA and AS events would be beneficial. Knockdown of these key APA regulatory factors would most likely significantly increase the number of APA events being analyzed and may enhance the impact of correlative or non-correlative observations. However, the striking lack of overlap between APA and AS observed thus far provides strong circumstantial evidence in support of the notion that APA and upstream AS are not mechanistically linked.

REFERENCES

1. Black DL (2003) Mechanisms of alternative pre-messenger RNA Splicing. Annual Review of Biochemistry 72:291–336 . doi: 10.1146/annurev.biochem.72.121801.161720

2. Hegele A, Kamburov A, Grossmann A, et al (2012) Dynamic protein-protein interaction wiring of the human spliceosome. Mol Cell 45:567–580 . doi: 10.1016/j.molcel.2011.12.034

3. Jurica MS, Moore MJ (2003) Pre-mRNA splicing: awash in a sea of proteins. Mol Cell 12:5–14

4. Wahl MC, Will CL, Lührmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 136:701–718 . doi: 10.1016/j.cell.2009.02.009

5. Reed R (1996) Initial splice-site recognition and pairing during pre-mRNA splicing. Current Opinion in Genetics & Development 6:215–220 . doi: 10.1016/S0959-437X(96)80053-0

6. Moore MJ, Query CC, Sharp PA (1993) Splicing of precursors to mRNA by the spliceosome. In: The RNA World. pp 303–357

7. Pertea M, Shumate A, Pertea G, et al (2018) Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise. bioRxiv 332825 . doi: 10.1101/332825

8. Pan Q, Shai O, Lee LJ, et al (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40:1413–1415 . doi: 10.1038/ng.259

9. Wang ET, Sandberg R, Luo S, et al (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–476 . doi: 10.1038/nature07509

10. Dvinge H, Bradley RK (2015) Widespread intron retention diversifies most cancer transcriptomes. Genome Med 7:45 . doi: 10.1186/s13073-015-0168-9

11. Roy B, Haupt LM, Griffiths LR (2013) Review: alternative splicing (AS) of genes as an approach for generating protein complexity. Curr Genomics 14:182–194 . doi: 10.2174/1389202911314030004

12. Baek D, Green P (2005) Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. Proc Natl Acad Sci USA 102:12813–12818 . doi: 10.1073/pnas.0506139102

13. Garg K, Green P (2007) Differing patterns of selection in alternative and constitutive splice sites. Genome Res 17:1015–1022 . doi: 10.1101/gr.6347907

14. Lavigueur A, La Branche H, Kornblihtt AR, Chabot B (1993) A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. Genes Dev 7:2405–2417

15. Zheng CL, Fu X-D, Gribskov M (2005) Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. RNA 11:1777–1787 . doi: 10.1261/rna.2660805

16. Dhir A, Buratti E (2010) Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. FEBS J 277:841–855 . doi: 10.1111/j.1742-4658.2009.07520.x

17. Dhir A, Buratti E, van Santen MA, et al (2010) The intronic splicing code: multiple factors involved in ATM pseudoexon definition. EMBO J 29:749–760 . doi: 10.1038/emboj.2009.397

18. Chen J-M, Férec C, Cooper DN (2006) A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3' UTR variants. Hum Genet 120:301–333 . doi: 10.1007/s00439-006-0218-x

19. Conne B, Stutz A, Vassalli J-D (2000) The 3′ untranslated region of messenger RNA: A molecular 'hotspot' for pathology? Nat Med 6:637–641 . doi: 10.1038/76211

20. Di Giammartino DC, Nishida K, Manley JL (2011) Mechanisms and consequences of alternative polyadenylation. Molecular Cell 43:853–866 . doi: 10.1016/j.molcel.2011.08.017

21. Lackford B, Yao C, Charles GM, et al (2014) Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. EMBO J 33:878–889 . doi: 10.1002/embj.201386537

22. Mayr C, Bartel DP (2009) Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138:673–684 . doi: 10.1016/j.cell.2009.06.016

23. Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463:457–463 . doi: 10.1038/nature08909

24. Hertel KJ (2008) Combinatorial control of exon recognition. J Biol Chem 283:1211–1215 . doi: 10.1074/jbc.R700035200

25. Roca X, Sachidanandam R, Krainer AR (2005) Determinants of the inherent strength of human 5' splice sites. RNA 11:683–698 . doi: 10.1261/rna.2040605

26. Yeo G, Burge CB (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 11:377–394 . doi: 10.1089/1066527041410418

27. Zhang XH-F, Chasin LA (2004) Computational definition of sequence motifs governing constitutive exon splicing. Genes Dev 18:1241–1250 . doi: 10.1101/gad.1195304

28. Moore MJ (2000) Intron recognition comes of AGe. Nat Struct Biol 7:14–16 . doi: 10.1038/71207

29. Reed R (1989) The organization of 3' splice-site sequences in mammalian introns. Genes Dev 3:2113–2123

30. Voithenberg LV von, Sánchez-Rico C, Kang H-S, et al (2016) Recognition of the 3′ splice site RNA by the U2AF heterodimer involves a dynamic population shift. PNAS 113:E7169–E7175 . doi: 10.1073/pnas.1605873113

31. 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 . doi: 10.1128/MCB.01071-09

32. Shepard PJ, Choi E-A, Busch A, Hertel KJ (2011) Efficient internal exon recognition depends on near equal contributions from the 3' and 5' splice sites. Nucleic Acids Res 39:8928–8937 . doi: 10.1093/nar/gkr481

33. Shen H, Kan JLC, Green MR (2004) Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell 13:367–376

34. Wu JY, Maniatis T (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell 75:1061–1070

35. Cartegni L, Chew SL, Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285–298 . doi: 10.1038/nrg775

36. Hertel KJ, Lynch KW, Maniatis T (1997) Common themes in the function of transcription and splicing enhancers. Curr Opin Cell Biol 9:350–357

37. Lam BJ, Hertel KJ (2002) A general role for splicing enhancers in exon definition. RNA 8:1233–1241

38. Tacke R, Manley JL (1999) Determinants of SR protein specificity. Curr Opin Cell Biol 11:358–362 . doi: 10.1016/S0955-0674(99)80050-7

39. Schaal TD, Maniatis T (1999) Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol Cell Biol 19:261–273

40. Staknis D, Reed R (1994) SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol Cell Biol 14:7670–7682

41. Tian M, Maniatis T (1994) A splicing enhancer exhibits both constitutive and regulated activities. Genes Dev 8:1703–1712

42. Chiara MD, Gozani O, Bennett M, et al (1996) Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly. Mol Cell Biol 16:3317–3326

43. Geuens T, Bouhy D, Timmerman V (2016) The hnRNP family: insights into their role in health and disease. Hum Genet 135:851–867 . doi: 10.1007/s00439-016-1683-5

44. Cho S, Moon H, Loh TJ, et al (2014) HnRNP M facilitates exon 7 inclusion of SMN2 pre-mRNA in spinal muscular atrophy by targeting an enhancer on exon 7. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1839:306–315 . doi: 10.1016/j.bbagrm.2014.02.006

45. Dredge BK, Stefani G, Engelhard CC, Darnell RB (2005) Nova autoregulation reveals dual functions in neuronal splicing. EMBO J 24:1608–1620 . doi: 10.1038/sj.emboj.7600630

46. Huelga SC, Vu AQ, Arnold JD, et al (2012) Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep 1:167–178 . doi: 10.1016/j.celrep.2012.02.001

47. Hung L-H, Heiner M, Hui J, et al (2008) Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. RNA 14:284–296 . doi: 10.1261/rna.725208

48. Schaub MC, Lopez SR, Caputi M (2007) Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. J Biol Chem 282:13617–13626 . doi: 10.1074/jbc.M700774200

49. Wang E, Mueller WF, Hertel KJ, Cambi F (2011) G Run-mediated recognition of proteolipid protein and DM20 5’ splice sites by U1 small nuclear RNA is regulated by context and proximity to the splice site. J Biol Chem 286:4059–4071 . doi: 10.1074/jbc.M110.199927

50. Kanopka A, Mühlemann O, Akusjärvi G (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. Nature 381:535–538 . doi: 10.1038/381535a0

51. Simard MJ, Chabot B (2002) SRp30c is a repressor of 3' splice site utilization. Mol Cell Biol 22:4001–4010

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

53. Garneau D, Revil T, Fisette J-F, Chabot B (2005) Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. J Biol Chem 280:22641–22650 . doi: 10.1074/jbc.M501070200

54. Hastings ML, Wilson CM, Munroe SH (2001) A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA. RNA 7:859–874

55. Ibrahim EC, Schaal TD, Hertel KJ, et al (2005) Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. Proc Natl Acad Sci USA 102:5002–5007 . doi: 10.1073/pnas.0500543102

56. McCullough AJ, Berget SM (1997) G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol Cell Biol 17:4562–4571

57. Erkelenz S, Mueller WF, Evans MS, et al (2013) Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. RNA 19:96–102 . doi: 10.1261/rna.037044.112

58. Berget SM (1995) Exon recognition in vertebrate splicing. J Biol Chem 270:2411–2414

59. Deutsch M, Long M (1999) Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res 27:3219–3228

60. Lander ES, Linton LM, Birren B, et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921 . doi: 10.1038/35057062

61. De Conti L, Baralle M, Buratti E (2013) Exon and intron definition in pre-mRNA splicing. Wiley Interdisciplinary Reviews: RNA 4:49–60 . doi: 10.1002/wrna.1140

62. Fox-Walsh KL, Dou Y, Lam BJ, et al (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci USA 102:16176–16181 . doi: 10.1073/pnas.0508489102

63. 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

64. Sakharkar MK, Perumal BS, Sakharkar KR, Kangueane P (2005) An analysis on gene architecture in human and mouse genomes. In Silico Biol (Gedrukt) 5:347–365

65. Izaurralde E, Lewis J, McGuigan C, et al (1994) A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell 78:657–668

66. Cooke C, Hans H, Alwine JC (1999) Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. Mol Cell Biol 19:4971–4979

67. Niwa M, Rose SD, Berget SM (1990) In vitro polyadenylation is stimulated by the presence of an upstream intron. Genes Dev 4:1552–1559 . doi: 10.1101/gad.4.9.1552

68. Clark F, Thanaraj TA (2002) Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. Hum Mol Genet 11:451–464

69. Itoh H, Washio T, Tomita M (2004) Computational comparative analyses of alternative splicing regulation using full-length cDNA of various eukaryotes. RNA 10:1005–1018 . doi: 10.1261/rna.5221604

70. Magen A, Ast G (2005) The importance of being divisible by three in alternative splicing. Nucleic Acids Res 33:5574–5582 . doi: 10.1093/nar/gki858

71. Sorek R, Ast G (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res 13:1631–1637 . doi: 10.1101/gr.1208803

72. Stamm S, Zhu J, Nakai K, et al (2000) An alternative-exon database and its statistical analysis. DNA Cell Biol 19:739–756 . doi: 10.1089/104454900750058107

73. Busch A, Hertel KJ (2015) Splicing predictions reliably classify different types of alternative splicing. RNA 21:813–823 . doi: 10.1261/rna.048769.114

74. Busch A, Hertel KJ (2012) Evolution of SR protein and hnRNP splicing regulatory factors. Wiley Interdisciplinary Reviews: RNA 3:1–12 . doi: 10.1002/wrna.100

75. Keren H, Lev-Maor G, Ast G (2010) Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 11:345–355 . doi: 10.1038/nrg2776

76. Kim E, Magen A, Ast G (2007) Different levels of alternative splicing among eukaryotes. Nucleic Acids Res 35:125–131 . doi: 10.1093/nar/gkl924

77. França GS, Cancherini DV, de Souza SJ (2012) Evolutionary history of exon shuffling. Genetica 140:249–257 . doi: 10.1007/s10709-012-9676-3

78. Sorek R (2007) The birth of new exons: mechanisms and evolutionary consequences. RNA 13:1603–1608 . doi: 10.1261/rna.682507

79. Deininger P (2011) Alu elements: know the SINEs. Genome Biology 12:236 . doi: 10.1186/gb-2011-12-12-236

80. Lev-Maor G, Sorek R, Shomron N, Ast G (2003) The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. Science 300:1288–1291 . doi: 10.1126/science.1082588

81. Nekrutenko A, Li WH (2001) Transposable elements are found in a large number of human protein-coding genes. Trends Genet 17:619–621

82. Sorek R, Ast G, Graur D (2002) Alu-containing exons are alternatively spliced. Genome Res 12:1060–1067 . doi: 10.1101/gr.229302

83. Kim S, Cho C-S, Han K, Lee J (2016) Structural variation of Alu element and human disease. Genomics Inform 14:70–77 . doi: 10.5808/GI.2016.14.3.70

84. Lin L, Shen S, Tye A, et al (2008) Diverse splicing patterns of exonized Alu elements in human tissues. PLOS Genetics 4:e1000225 . doi: 10.1371/journal.pgen.1000225

85. Schmid CW (1996) Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. Prog Nucleic Acid Res Mol Biol 53:283–319

86. Fairbrother WG, Yeh R-F, Sharp PA, Burge CB (2002) Predictive identification of exonic splicing enhancers in human genes. Science 297:1007–1013 . doi: 10.1126/science.1073774

87. Zhang C, Li W-H, Krainer AR, Zhang MQ (2008) RNA landscape of evolution for optimal exon and intron discrimination. Proc Natl Acad Sci USA 105:5797–5802 . doi: 10.1073/pnas.0801692105

88. Faa' V, Coiana A, Incani F, et al (2010) A synonymous mutation in the CFTR gene causes aberrant splicing in an italian patient affected by a mild form of cystic fibrosis. J Mol Diagn 12:380–383 . doi: 10.2353/jmoldx.2010.090126

89. Sauna ZE, Kimchi-Sarfaty C (2011) Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet 12:683–691 . doi: 10.1038/nrg3051

90. Supek F, Miñana B, Valcárcel J, et al (2014) Synonymous mutations frequently act as driver mutations in human cancers. Cell 156:1324–1335 . doi: 10.1016/j.cell.2014.01.051

91. Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134:341–352 . doi: 10.1016/j.cell.2008.05.042

92. Gingold H, Pilpel Y (2011) Determinants of translation efficiency and accuracy. Mol Syst Biol 7:481 . doi: 10.1038/msb.2011.14

93. Sonneborn TM (1965) Degeneracy of the genetic code: extent, nature, and genetic implications. In: Bryson V, Vogel HJ (eds) Evolving Genes and Proteins. Academic Press, pp 377–397

94. Mueller WF, Larsen LSZ, Garibaldi A, et al (2015) The silent sway of splicing by synonymous substitutions. J Biol Chem 290:27700–27711 . doi: 10.1074/jbc.M115.684035

95. Chan SL, Huppertz I, Yao C, et al (2014) CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. Genes Dev gad.250993.114 . doi: 10.1101/gad.250993.114

96. Sun Y, Zhang Y, Hamilton K, et al (2018) Molecular basis for the recognition of the human AAUAAA polyadenylation signal. Proc Natl Acad Sci USA 115:E1419–E1428 . doi: 10.1073/pnas.1718723115

97. Tian B, Manley JL (2017) Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol 18:18–30 . doi: 10.1038/nrm.2016.116

98. Derti A, Garrett-Engele P, Macisaac KD, et al (2012) A quantitative atlas of polyadenylation in five mammals. Genome Res 22:1173–1183 . doi: 10.1101/gr.132563.111

99. Shepard PJ, Choi E-A, Lu J, et al (2011) Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA 17:761–772 . doi: 10.1261/rna.2581711

100. Tian B, Hu J, Zhang H, Lutz CS (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res 33:201–212 . doi: 10.1093/nar/gki158

101. Fu X-D, Ares M (2014) Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet 15:689–701 . doi: 10.1038/nrg3778

102. Millevoi S, Vagner S (2010) Molecular mechanisms of eukaryotic pre-mRNA 3’ end processing regulation. Nucleic Acids Res 38:2757–2774 . doi: 10.1093/nar/gkp1176

103. Szostak E, Gebauer F (2013) Translational control by 3′-UTR-binding proteins. Brief Funct Genomics 12:58–65 . doi: 10.1093/bfgp/els056

104. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136:215–233 . doi: 10.1016/j.cell.2009.01.002

105. Leung AKL (2015) The whereabouts of microRNA actions: cytoplasm and beyond. Trends in Cell Biology 25:601–610 . doi: 10.1016/j.tcb.2015.07.005

106. Reczko M, Maragkakis M, Alexiou P, et al (2012) Functional microRNA targets in protein coding sequences. Bioinformatics 28:771–776 . doi: 10.1093/bioinformatics/bts043

107. Martin G, Gruber AR, Keller W, Zavolan M (2012) Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. Cell Rep 1:753–763 . doi: 10.1016/j.celrep.2012.05.003

108. Jenal M, Elkon R, Loayza-Puch F, et al (2012) The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell 149:538–553 . doi: 10.1016/j.cell.2012.03.022

109. Li W, You B, Hoque M, et al (2015) Systematic profiling of poly(A)+ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. PLoS Genet 11:e1005166 . doi: 10.1371/journal.pgen.1005166

110. Yao C, Biesinger J, Wan J, et al (2012) Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. Proc Natl Acad Sci USA 109:18773–18778 . doi: 10.1073/pnas.1211101109

111. Zhu Y, Wang X, Forouzmand E, et al (2018) Molecular mechanisms for CFIm-mediated regulation of mRNA alternative polyadenylation. Mol Cell 69:62-74.e4 . doi: 10.1016/j.molcel.2017.11.031

112. Brown KM, Gilmartin GM (2003) A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. Mol Cell 12:1467–1476

113. Zhou Z, Licklider LJ, Gygi SP, Reed R (2002) Comprehensive proteomic analysis of the human spliceosome. Nature 419:182–185 . doi: 10.1038/nature01031

114. Kubo T, Wada T, Yamaguchi Y, et al (2006) Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3'-UTRs. Nucleic Acids Res 34:6264–6271 . doi: 10.1093/nar/gkl794

115. Hatton LS, Eloranta JJ, Figueiredo LM, et al (2000) The Drosophila homologue of the 64 kDa subunit of cleavage stimulation factor interacts with the 77 kDa subunit encoded by the suppressor of forked gene. Nucleic Acids Res 28:520–526

116. Takagaki Y, Manley JL (2000) Complex protein interactions within the human polyadenylation machinery identify a novel component. Mol Cell Biol 20:1515–1525

117. Takagaki Y, Seipelt RL, Peterson ML, Manley JL (1996) The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell 87:941–952

118. Millevoi S, Loulergue C, Dettwiler S, et al (2006) An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. EMBO J 25:4854–4864 . doi: 10.1038/sj.emboj.7601331

119. Vagner S, Vagner C, Mattaj IW (2000) The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. Genes Dev 14:403–413

120. Kyburz A, Friedlein A, Langen H, Keller W (2006) Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. Mol Cell 23:195–205 . doi: 10.1016/j.molcel.2006.05.037

121. Lutz CS, Murthy KG, Schek N, et al (1996) Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. Genes Dev 10:325–337

122. Kaida D, Berg MG, Younis I, et al (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature 468:664–668 . doi: 10.1038/nature09479

123. Gunderson SI, Polycarpou-Schwarz M, Mattaj IW (1998) U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. Mol Cell 1:255–264

124. Spraggon L, Cartegni L (2013) U1 snRNP-dependent suppression of polyadenylation: physiological role and therapeutic opportunities in cancer. International Journal of Cell Biology 2013: . doi: 10.1155/2013/846510

125. Fortes P, Cuevas Y, Guan F, et al (2003) Inhibiting expression of specific genes in mammalian cells with 5' end-mutated U1 small nuclear RNAs targeted to terminal exons of pre-mRNA. Proc Natl Acad Sci USA 100:8264–8269 . doi: 10.1073/pnas.1332669100

126. Davidson L, West S (2013) Splicing-coupled 3′ end formation requires a terminal splice acceptor site, but not intron excision. Nucl Acids Res gkt446 . doi: 10.1093/nar/gkt446

127. Muniz L, Davidson L, West S (2015) Poly(A) polymerase and the nuclear poly(A) binding protein, PABPN1, coordinate the splicing and degradation of a subset of human pre-mRNAs. Mol Cell Biol 35:2218–2230 . doi: 10.1128/MCB.00123-15

128. David CJ, Boyne AR, Millhouse SR, Manley JL (2011) The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. Genes Dev 25:972–983 . doi: 10.1101/gad.2038011

129. McCracken S, Fong N, Yankulov K, et al (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature 385:357–361 . doi: 10.1038/385357a0

130. Qu X, Perez-Canadillas J-M, Agrawal S, et al (2007) The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for

mRNA 3′-end processing. J Biol Chem 282:2101–2115 . doi: 10.1074/jbc.M609981200

131. Zeng C, Berget SM (2000) Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. Mol Cell Biol 20:8290–8301

132. Faustino NA, Cooper TA (2003) Pre-mRNA splicing and human disease. Genes Dev 17:419–437 . doi: 10.1101/gad.1048803

133. Brinkman BMN (2004) Splice variants as cancer biomarkers. Clin Biochem 37:584–594 . doi: 10.1016/j.clinbiochem.2004.05.015

134. Carstens RP, Eaton JV, Krigman HR, et al (1997) Alternative splicing of fibroblast growth factor receptor 2 (FGF-R2) in human prostate cancer. Oncogene 15:3059–3065 . doi: 10.1038/sj.onc.1201498

135. Wang Z, Lo HS, Yang H, et al (2003) Computational analysis and experimental validation of tumor-associated alternative RNA splicing in human cancer. Cancer Res 63:655–657

136. Xu Q, Lee C (2003) Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. Nucleic Acids Res 31:5635–5643

137. Lin S, Fu X-D (2007) SR proteins and related factors in alternative splicing. Adv Exp Med Biol 623:107–122

138. Venables JP (2007) Downstream intronic splicing enhancers. FEBS Letters 581:4127–4131 . doi: 10.1016/j.febslet.2007.08.012

139. Duan J, Wainwright MS, Comeron JM, et al (2003) Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. Hum Mol Genet 12:205–216

140. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A (2010) Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 20:110–121 . doi: 10.1101/gr.097857.109

141. Sakharkar MK, Chow VTK, Kangueane P (2004) Distributions of exons and introns in the human genome. In Silico Biol (Gedrukt) 4:387–393

142. Sterner DA, Carlo T, Berget SM (1996) Architectural limits on split genes. Proc Natl Acad Sci USA 93:15081–15085

143. Siepel A (2009) Phylogenomics of primates and their ancestral populations. Genome Research 19:1929–1941 . doi: 10.1101/gr.084228.108

144. Lim KH, Ferraris L, Filloux ME, et al (2011) Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. Proc Natl Acad Sci USA 108:11093–11098 . doi: 10.1073/pnas.1101135108

145. Soemedi R, Cygan KJ, Rhine CL, et al (2017) Pathogenic variants that alter protein code often disrupt splicing. Nat Genet 49:848–855 . doi: 10.1038/ng.3837

146. Britten RJ (2002) Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. Proc Natl Acad Sci USA 99:13633–13635 . doi: 10.1073/pnas.172510699

147. Shastry BS (2002) SNP alleles in human disease and evolution. J Hum Genet 47:561–566 . doi: 10.1007/s100380200086

148. Yang JO, Kim W-Y, Bhak J (2009) ssSNPTarget: genome-wide splice-site Single Nucleotide Polymorphism database. Hum Mutat 30:E1010-1020 . doi: 10.1002/humu.21128

149. Liu HX, Cartegni L, Zhang MQ, Krainer AR (2001) A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. Nat Genet 27:55–58 . doi: 10.1038/83762

150. Blanchette M, Kent WJ, Riemer C, et al (2004) Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res 14:708–715 . doi: 10.1101/gr.1933104

151. Burge CB, Tuschl T, Sharp PA (1999) Splicing of precursors to mRNAs by the spliceosomes. In: The RNA World, 2nd Ed: The Nature of Modern RNA Suggests a Prebiotic RNA World, 2nd ed. pp 525–560

152. Sun H, Chasin LA (2000) Multiple splicing defects in an intronic false exon. Mol Cell Biol 20:6414–6425

153. Fu XD, Maniatis T (1992) The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. Proc Natl Acad Sci USA 89:1725–1729

154. Chen Y, Carlini DB, Baines JF, et al (1999) RNA secondary structure and compensatory evolution. Genes Genet Syst 74:271–286

155. Chiou N-T, Shankarling G, Lynch KW (2013) HnRNP L and HnRNP A1 induce extended U1 snRNA interactions with an exon to repress spliceosome assembly. Molecular Cell 49:972–982 . doi: 10.1016/j.molcel.2012.12.025

156. Hernández H, Makarova OV, Makarov EM, et al (2009) Isoforms of U1-70k Control Subunit Dynamics in the Human Spliceosomal U1 snRNP. PLoS One 4: . doi: 10.1371/journal.pone.0007202

157. Query CC, Bentley RC, Keene JD (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. Cell 57:89–101

158. Mouaikel J, Narayanan U, Verheggen C, et al (2003) Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. EMBO Rep 4:616–622 . doi: 10.1038/sj.embor.embor863

159. Kondo Y, Oubridge C, van Roon A-MM, Nagai K (2015) Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. eLife 4: . doi: 10.7554/eLife.04986

160. Pomeranz Krummel DA, Oubridge C, Leung AKW, et al (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature 458:475–480 . doi: 10.1038/nature07851

161. Rösel-Hillgärtner TD, Hung L-H, Khrameeva E, et al (2013) A novel intra-U1 snRNP cross-regulation mechanism: alternative splicing switch links U1C and U1-70K expression. PLoS Genet 9: . doi: 10.1371/journal.pgen.1003856

162. Sharma S, Wongpalee SP, Vashisht A, et al (2014) Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. Genes Dev 28:2518–2531 . doi: 10.1101/gad.248625.114

163. Sharma S, Maris C, Allain FH-T, Black DL (2011) U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression. Molecular Cell 41:579–588 . doi: 10.1016/j.molcel.2011.02.012

164. Förch P, Puig O, Martínez C, et al (2002) The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5′ splice sites. EMBO J 21:6882–6892 . doi: 10.1093/emboj/cdf668

165. Roca X, Sachidanandam R, Krainer AR (2003) Intrinsic differences between authentic and cryptic 5’ splice sites. Nucleic Acids Res 31:6321–6333

166. Listerman I, Sapra AK, Neugebauer KM (2006) Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nat Struct Mol Biol 13:815–822 . doi: 10.1038/nsmb1135

167. Pandya-Jones A, Black DL (2009) Co-transcriptional splicing of constitutive and alternative exons. RNA 15:1896–1908 . doi: 10.1261/rna.1714509

168. Moore MJ, Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. Cell 136:688–700 . doi: 10.1016/j.cell.2009.02.001

169. Kornblihtt AR (2006) Chromatin, transcript elongation and alternative splicing. Nat Struct Mol Biol 13:5–7 . doi: 10.1038/nsmb0106-5

170. Brugiolo M, Herzel L, Neugebauer KM (2013) Counting on co-transcriptional splicing. F1000Prime Rep 5:9 . doi: 10.12703/P5-9

171. Landry J-R, Mager DL, Wilhelm BT (2003) Complex controls: the role of alternative promoters in mammalian genomes. Trends Genet 19:640–648 . doi: 10.1016/j.tig.2003.09.014

172. Logette E, Wotawa A, Solier S, et al (2003) The human caspase-2 gene: alternative promoters, pre-mRNA splicing and AUG usage direct isoform-specific expression. Oncogene 22:935–946 . doi: 10.1038/sj.onc.1206172

173. Pecci A, Viegas LR, Baranao JL, Beato M (2001) Promoter choice influences alternative splicing and determines the balance of isoforms expressed from the mouse bcl-X gene. J Biol Chem 276:21062–21069 . doi: 10.1074/jbc.M008665200

174. Wang Y, Newton DC, Robb GB, et al (1999) RNA diversity has profound effects on the translation of neuronal nitric oxide synthase. PNAS 96:12150–12155 . doi: 10.1073/pnas.96.21.12150

175. Xin D, Hu L, Kong X (2008) Alternative promoters influence alternative splicing at the genomic level. PLoS ONE 3:e2377 . doi: 10.1371/journal.pone.0002377

176. Rigo F, Martinson HG (2008) Functional coupling of last-intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage. Mol Cell Biol 28:849–862 . doi: 10.1128/MCB.01410-07

177. Yao C, Choi E-A, Weng L, et al (2013) Overlapping and distinct functions of CstF64 and CstF64τ in mammalian mRNA 3' processing. RNA 19:1781–1790 . doi: 10.1261/rna.042317.113

178. Ara T, Lopez F, Ritchie W, et al (2006) Conservation of alternative polyadenylation patterns in mammalian genes. BMC Genomics 7:189 . doi: 10.1186/1471-2164-7-189

179. Katz Y, Wang ET, Airoldi EM, Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods 7:1009–1015 . doi: 10.1038/nmeth.1528

180. McGlincy NJ, Tan L-Y, Paul N, et al (2010) Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay. BMC Genomics 11:565 . doi: 10.1186/1471-2164-11-565

181. Martinez-Contreras R, Fisette J-F, Nasim FH, et al (2006) Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. PLoS Biol 4:e21 . doi: 10.1371/journal.pbio.0040021

182. Shi Y, Di Giammartino DC, Taylor D, et al (2009) Molecular architecture of the human pre-mRNA 3’ processing complex. Mol Cell 33:365–376 . doi: 10.1016/j.molcel.2008.12.028

183. Niwa M, Berget SM (1991) Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. Genes Dev 5:2086–2095 . doi: 10.1101/gad.5.11.2086

184. de la Mata M, Alonso CR, Kadener S, et al (2003) A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell 12:525–532

185. de la Mata M, Lafaille C, Kornblihtt AR (2010) First come, first served revisited: Factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. RNA 16:904–912 . doi: 10.1261/rna.1993510

186. de la Mata M, Muñoz MJ, Alló M, et al (2011) RNA polymerase II elongation at the crossroads of transcription and alternative splicing. Genet Res Int 2011:309865 . doi: 10.4061/2011/309865

187. Kessler O, Jiang Y, Chasin LA (1993) Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. Mol Cell Biol 13:6211–6222

188. Fähling M, Mrowka R, Steege A, et al (2006) Heterogeneous nuclear ribonucleoprotein-A2/B1 modulate collagen prolyl 4-hydroxylase, alpha (I) mRNA stability. J Biol Chem 281:9279–9286 . doi: 10.1074/jbc.M510925200

189. Scotto-Lavino E, Du G, Frohman MA (2006) 3’ end cDNA amplification using classic RACE. Nat Protoc 1:2742–2745 . doi: 10.1038/nprot.2006.481

190. Kim D, Pertea G, Trapnell C, et al (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14:R36 . doi: 10.1186/gb-2013-14-4-r36

191. Alekseyenko AV, Kim N, Lee CJ (2007) Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. RNA 13:661–670 . doi: 10.1261/rna.325107

192. Corvelo A, Eyras E (2008) Exon creation and establishment in human genes. Genome Biol 9:R141 . doi: 10.1186/gb-2008-9-9-r141

193. Shabalina SA, Ogurtsov AY, Spiridonov AN, et al (2010) Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes. Mol Biol Evol 27:1745–1749 . doi: 10.1093/molbev/msq086

194. Marques AC, Dupanloup I, Vinckenbosch N, et al (2005) Emergence of young human genes after a burst of retroposition in primates. PLoS Biol 3:e357 . doi: 10.1371/journal.pbio.0030357

195. Chen C, Gentles AJ, Jurka J, Karlin S (2002) Genes, pseudogenes, and Alu sequence organization across human chromosomes 21 and 22. Proc Natl Acad Sci U S A 99:2930–2935 . doi: 10.1073/pnas.052692099

196. Lei H, Dias AP, Reed R (2011) Export and stability of naturally intronless mRNAs require specific coding region sequences and the TREX mRNA export complex. Proc Natl Acad Sci U S A 108:17985–17990 . doi: 10.1073/pnas.1113076108

197. Caminsky N, Mucaki EJ, Rogan PK (2014) Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. F1000Res 3: . doi: 10.12688/f1000research.5654.1

198. Krawczak M, Reiss J, Cooper DN (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 90:41–54

199. Barash Y, Calarco JA, Gao W, et al (2010) Deciphering the splicing code. Nature 465:53–59 . doi: 10.1038/nature09000

200. Yang Y, Walsh CE (2005) Spliceosome-mediated RNA trans-splicing. Molecular Therapy 12:1006–1012 . doi: 10.1016/j.ymthe.2005.09.006

201. Guiro J, O'Reilly D (2015) Insights into the U1 small nuclear ribonucleoprotein complex superfamily. WIREs RNA 6:79–92 . doi: 10.1002/wrna.1257

202. So BR, Wan L, Zhang Z, et al (2016) A U1 snRNP-specific assembly pathway reveals the SMN complex as a versatile RNP exchange. Nat Struct Mol Biol 23:225–230 . doi: 10.1038/nsmb.3167

203. Del Gatto-Konczak F, Bourgeois CF, Le Guiner C, et al (2000) The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5' splice site. Mol Cell Biol 20:6287–6299

204. Shen S, Park JW, Lu Z, et al (2014) rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. PNAS 111:E5593–E5601 . doi: 10.1073/pnas.1419161111

205. Vaquero-Garcia J, Barrera A, Gazzara MR, et al (2016) A new view of transcriptome complexity and regulation through the lens of local splicing variations. Elife 5:e11752 . doi: 10.7554/eLife.11752

206. Lianoglou S, Garg V, Yang JL, et al (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev 27:2380–2396 . doi: 10.1101/gad.229328.113

207. Zhang H, Lee JY, Tian B (2005) Biased alternative polyadenylation in human tissues. Genome Biol 6:R100 . doi: 10.1186/gb-2005-6-12-r100

APPENDIX A

In Vitro Assay of Pre-mRNA Splicing in Mammalian Nuclear Extract

Chapter 11

In Vitro Assay of Pre-mRNA Splicing in Mammalian Nuclear Extract

Maliheh Movassat, William F. Mueller, and Klemens J. Hertel

Abstract

The in vitro splicing assay is a valuable technique that can be used to study the mechanism and machinery involved in the splicing process. The ability to investigate various aspects of splicing and alternative splicing appears to be endless due to the flexibility of this assay. Here, we describe the tools and techniques necessary to carry out an in vitro splicing assay. Through the use of radiolabeled pre-mRNA and crude nuclear extract, spliced mRNAs can be purified and visualized by autoradiography for downstream analysis.

Key words In vitro splicing, Alternative splicing, Splicing analysis, Pre-mRNA substrate, HeLa cell nuclear extract, In vitro transcription, RNA extraction and purification

1 Introduction

The ability to study biochemical changes associated with pre-mRNA splicing in a cell-free-based assay, also referred to as the in vitro splicing assay, has vastly improved our understanding of this complex, key process of gene expression. Not only has it improved our knowledge of the mechanisms and necessary components involved in splicing, but it has also allowed insights into the regulation of alternative splicing as it is mediated by *cis*-acting elements and *trans*-acting factors.

The ease of use, flexibility, and rapid results provided by an in vitro splicing system allows for tailored investigations into various aspects of the splicing reaction. The major benefit, however, lies with the ability to biochemically manipulate the splicing reaction through utilizing two key components: (1) minigene constructs and (2) mammalian crude nuclear extracts. The use of minigene constructs is a common in vitro technique that employs genomic segments from a gene (introns and exons) that include alternatively spliced regions within flanking genomic regions that are cloned

Klemens J. Hertel (ed.), *Spliceosomal Pre-mRNA Splicing: Methods and Protocols*, Methods in Molecular Biology, vol. 1126, DOI 10.1007/978-1-62703-980-2_11, © Springer Science+Business Media, LLC 2014

downstream of efficient promoters. These minigene constructs allow for identification of specific features that control intron and exon usage as well as the characterization of *cis*-acting elements and *trans*-acting factors that interact and modulate regulatory elements necessary for splicing regulation [1]. Crude nuclear extract is another important component of the in vitro splicing reaction that is usually generated from HeLa cells. Importantly, these nuclear extracts contain the necessary proteins and snRNAs for an efficient splicing reaction (*see* Chapter 8). The advantage associated with in vitro biochemical manipulation allows for insights into various factors and processes. These include, but are not limited to, protein regulatory elements and composition, splice site recognition and selection, the influence of RNA elements and their *trans*-acting factors, the characterization of enhancer and silencer elements, and kinetic insights into the splicing pathway. As with all in vitro-based systems, the assay does come with limitations. The rate of intron removal in vitro is slower than rates determined in vivo [2]. The efficiency of in vitro transcription of pre-mRNA, its purification, and subsequent splicing is restricted by the size of the RNA to be used; RNA should be less than 2,000bp [3]. Because of this, the in vitro splicing assay relies heavily on the use of shorter minigenes that are only a subset of a larger gene. The assay also does not take into account the effects of other events associated with splicing, such as transcription, capping, and polyadenylation.

Methods for in vitro splicing reactions have previously been described [4–7]. In general, these protocols employ the use of radiolabeled pre-mRNAs that are incubated for several hours in nuclear extract supplemented with necessary salts and cofactors. The mRNA is then extracted and purified from the nuclear extract, subjected to denaturation on a polyacrylamide gel, and subsequently dried for visualization by autoradiography via film or phosphor imaging. The pre-mRNA, mRNA, and other intermediates are then identified as bands on the autoradiograph.

2 Materials

All reagents should be high quality, molecular biology grade, and RNase-free. Stock solutions should be stored at 4 °C (unless otherwise indicated). Certain reagents can be substituted for their equivalents from other manufacturers or as otherwise stated. The concentrations of chemicals/reagents listed in the materials are stock concentrations, not final concentrations. Since all steps require working with radioactive isotopes, all necessary precautions must be taken. Carefully follow all hazardous and radioactive waste disposal regulations when disposing of waste materials.

2.1 Splicing Reaction Components

1. Radiolabeled pre-mRNA: generated from an in vitro transcription reaction (*see* **Note 1**).
2. Splicing competent nuclear extract (NE) (*see* Chapter 8).
3. 1 mM adenosine triphosphate (ATP). Store at −20 °C.
4. 0.5 M creatine phosphate (CP). Store at −20 °C.
5. 80 mM magnesium acetate ($Mg(OAc)_2$) (*see* **Note 2**).
6. RNase inhibitor (40 U/μl). Store at −20 °C.
7. 100 mM dithiothreitol (DTT). Store at −20 °C.
8. 1 M potassium acetate (KOAc) (*see* **Note 3**).
9. 0.5 M HEPES buffer, pH 7.9 (*see* **Note 3**).
10. 13 % polyvinyl alcohol (PVA): optional (*see* **Note 4**).
11. Wet ice and dry ice (finely ground or small chunks).
12. Water bath.

2.2 6 % Splicing Gel Components

1. Tris-Borate-EDTA (TBE) buffer: 89 mM Tris Base, 89 mM boric acid, 2 mM EDTA.
2. 7 M urea.
3. 40 % (19:1) acrylamide:bis-acrylamide solution: acrylamide is dissolved in 1× TBE/7 M urea.
4. *N,N,N′,N′*-Tetramethylethylenediamine (TEMED).
5. 10 % Ammonium persulfate (APS).
6. Formamide/EDTA stop dye: formamide with 0.1 % bromophenol blue and 0.1 % xylene cyanol and 2 mM EDTA.
7. Radiolabeled RNA ladder/molecular marker.
8. Electrophoresis glass plates: 8″ × 8″ (two): one glass plate should notch to allow for the addition of a comb.
9. 0.4 mm gel plate spacers (three).
10. 0.4 mm comb (same thickness as the spacers).
11. 1¼″ binder clips (four).
12. Aluminum plate: 8″ × 8″ or longer and precooled (*see* **Note 5**).
13. Silicon Gel Slick® Solution (Lonza Rockland) or equivalent.
14. 70 % ethanol.
15. 30–50 ml syringe, with and without a needle (two).
16. Flat gel loading tips.
17. Putty knife/gel spatula.
18. Vertical gel electrophoresis system.
19. Whatman paper, cut into an 8″ × 8″ square.
20. Plastic wrap (such as Saran™ Wrap).
21. Power pack for an electrophoresis system with a temperature probe.

22. Bio-Rad Gel Dryer or equivalent.
23. Bio-Rad Personal Molecular PhosphorImager System or similar. Film may also be used.

2.3 Splicing Digest and RNA Purification

1. Proteinase K 10 mg/ml.
2. 2× Proteinase K buffer: 20 mM Tris Base, 2 % SDS, 200 mM NaCl, 2 mM EDTA, pH 7.5.
3. 100 % ethanol.
4. Glycogen.
5. Phenol, chloroform, isoamyl alcohol solution (25:24:1 pH 8.0).

3 Methods

Carry out all steps of the reaction on ice unless otherwise stated.

3.1 Splicing Reaction

1. Thaw NE on ice.
2. Thaw ATP, CP, $Mg(OAc)_2$, DTT, HEPES, KOAc, and radiolabeled RNA at room temperature. Once thawed, place them immediately on ice. RNase inhibitor should be kept on ice.
3. Determine the Master Mix reaction volume and reaction size (*see* **Note 6**):
   (a) (# of reactions) + 1 = Master Mix reaction size.
   (b) Reaction volume: 12.5 μl or 25 μl reaction volume total.
4. Mix reagents to a final concentration of 1 mM ATP, 20 mM CP, 3.2 mM $Mg(OAc)_2$, 10 U RNase inhibitor, 1 mM DTT, 10–50 % NE (should be optimized for each extract and substrate used), 72.5 mM KOAc, 12 mM HEPES (*see* **Note 3**), and 3 % PVA (optional). Use sterile water to bring up the Master Mix volume if needed.
5. For each experimental reaction condition: add the appropriate Master Mix volume, 0.01–0.1 nM RNA (~1,000 cpm) (*see* **Note 7**), experimental variant (i.e., protein), and/or sterile water to bring up the volume. Add NE last and pipet carefully to mix (*see* **Note 8**). Keep all reaction tubes on ice. Prepare a time 0 tube as control and immediately place on dry ice after addition of NE (*see* **Note 9**).
6. Incubate all reactions, except the time 0 reaction, at 30 °C water bath for 90 min (*see* **Note 10**).
7. While the splicing reactions are running, prepare the 6 % acrylamide gel.
8. Once the incubation time is complete, immediately place the tubes on dry ice to stop the reactions (*see* **Note 11**).

3.2 Splicing Gel Preparation

1. Prepare a 20 % acrylamide:bis solution: dilute 40 % (19:1) acrylamide:bis-acrylamide solution in 1× TBE with 7 M urea.

2. Prepare a 6 % polyacrylamide mixture from the 20 % acrylamide solution: in a 50 ml conical tube, dilute the 20 % acrylamide solution with the desired amount of 1× TBE/7 M urea buffer to obtain a mixture at the required percentage.
3. Carefully clean the inside face of a siliconized plate (*see* **Note 12**) with 70 % ethanol. Wipe dry with lint-free paper towels.
4. Carefully clean the non-siliconized plate using water and 70 % ethanol. Make sure the gel plates are completely clean, with no small pieces of debris present (*see* **Note 13**). Wipe dry with lint-free paper towels.
5. Place the spacers around the outside edge (bottoms and sides) of a non-siliconized plate. Lay the siliconized notched plate on top and clip the glass plates together using binder clips.
6. Once the gel cassette is ready, add the appropriate amount of 10 % APS and TEMED (*see* **Note 14**) to 20 ml of 6 % acrylamide and mix gently.
7. Using a syringe (without needle), aspirate the acrylamide and gently dispense the mixture between the plates. Once the cassette is filled, lay it flat, place a gel comb with an appropriate well size into the top of the gel, and allow the gel to set at room temperature for approximately 30 min (or until polymerized) (*see* **Note 15**).
8. Pre-run the gel before adding your samples (*see* **Note 16**): clamp the gel cassette onto the vertical gel electrophoresis apparatus, fill the chambers with 1× TBE (*see* **Note 17**), and run the gel at 30 W (100 V), 45 °C, for 15 min.

3.3 Digest

1. Once the last in vitro splicing tube has been placed on dry ice, prepare the Proteinase K digest mix:
   (a) Determine the desired final volume of the Proteinase K Master Mix: reaction volume (μl) × (# of reactions + 1) = Master Mix volume (μl).
   (b) Proteinase K Master Mix: final concentration of 1× Proteinase K buffer, 0.25 mg/ml glycogen, 0.25 mg/ml Proteinase K, and sterile water, for a final volume of 180 μl per reaction.
2. Add 175 μl of Proteinase K Master Mix (*see* **Note 18**) to each reaction tube and incubate at 37 °C for 10–15 min.

3.4 RNA Purification and Precipitation

1. Once the Proteinase K digest has completed, purify the RNA by adding 200 μl of phenol/chloroform, vortex for 30 s, and spin at 16,500 × *g* for 5 min to separate the aqueous and organic layers.
2. To precipitate the RNA: remove the aqueous (top) phase and place into a separate tube (~200 μl). Add 2.5 times the volume of 100 % ice-cold ethanol (for 200 μl of top phase, add 500 μl of ethanol). Incubate at −20 °C or −80 °C for 10–15 min.

3. Centrifuge the tubes at 16,500 × *g* for 10 min at room temperature to pellet.
4. Remove the ethanol supernatant and allow the pellet to air-dry for no more than 5 min (*see* **Note 19**). Resuspend the pellet in a small amount of stop dye within 5–10 min (10 μl or less). Pipet up and down and vortex for 30 s to mix.

3.5 Visualization of Splicing Reaction

1. Load RNA samples onto the pre-run gel. Clamp an aluminum plate to the front glass plate (*see* **Note 20**). Run the gel at 30 W (100 V), 45 °C, for 90 min or until the dye runs off the gel.
2. Remove the gel cassette from the apparatus and dispose of the buffers in appropriate waste containers. Split the plates apart with a putty knife/spatula. The gel should remain attached to the non-siliconized plate.
3. Center the pre-cut Whatman paper on top of the gel and press gently to allow the gel to adhere evenly to the paper. Carefully peel the Whatman paper upward at an angle to allow for the gel to be peeled away from the glass. Cover the gel with plastic wrap, minimizing the presence of any creases (*see* **Note 21**).
4. Dry the gel for 15–20 min using a Bio-Rad Gel Dryer at 80 °C with suction.
5. Expose the gel to film or preferably a phosphor imaging screen (*see* **Note 22**) or similar equipment for the recommended length of time (generally at least 3 h to overnight; *see* **Note 23**).
6. Once the gel has been exposed and imaged, the appearance of spliced product can be used to determine the amount of RNA spliced (% spliced) in each lane (Fig. 1), which can in turn be used to calculate the efficiency of product appearance (*see* Chapter 12 and **Note 24**). Use a suitable computer program to analyze the digital quantitation file (*see* **Note 25**).

4 Notes

1. Generally, for in vitro splicing reactions, DNA is transcribed using T7 polymerase in a reaction containing radiolabeled nucleotides, phosphorus-32 (^{32}P) α-UTP. This reagent is usually in the 0.3–3 nM range, with an incorporation of around 100,000 cpm/μl.
2. Magnesium chloride can also be used; however, chloride has in some cases been shown to inhibit in vitro splicing reactions [8].
3. Potassium chloride or potassium glutamate may also be used although chloride has been shown to inhibit in vitro splicing reactions [8]. KOAc is used in this reaction because the nuclear extract has been prepared in KOAc (*see* Chapter 8). The final volume of KOAc to add to the Master Mix will depend on

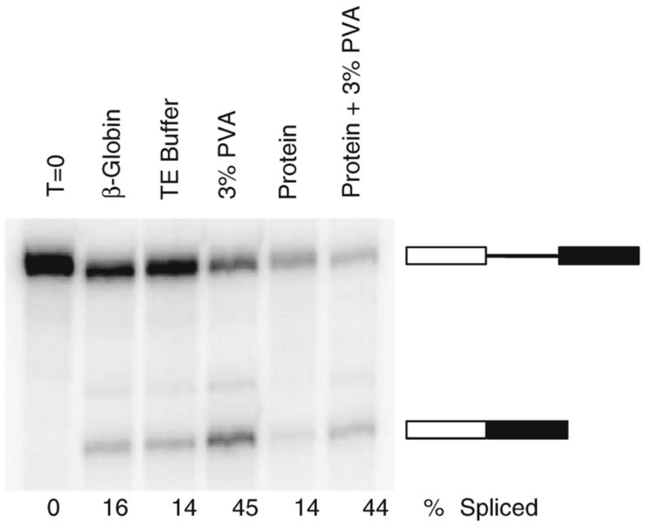


Fig. 1 Autoradiogram of radiolabeled β-globin minigene construct (from *left* to *right*) at time 0, alone, with TE buffer, with 3 % PVA, with 1 μM protein X, with 1 μM protein X with 3 % PVA, run on a 6 % polyacrylamide gel. Analysis of % spliced is performed using Bio-Rad Quantity One (*see* **Note 25**)

how many ions are present in the nuclear extract to begin with. The final volume of HEPES to add will also depend on how many ions are present in the nuclear extract.

4. Addition of PVA is optional but has been shown to potentially increase splicing efficiency in certain reactions [9].
5. Monitor the temperature of the gel using a temperature probe connected to the power pack. It is highly recommended to place a precooled aluminum plate aluminum plate on the front surface the front surface of the gel cassette to keep the cassette cool and prevent it from shattering as well as evenly distribute heat (*see* **Note 19** as well).
6. The extra reaction is to account for pipetting errors. When determining the reaction volume (12.5 μl or 25 μl reaction), consider how many reactions are needed, how much radiolabeled pre-RNA is present, and how radioactive the radiolabeled pre-mRNA is. If the radiolabeled pre-RNA is less than 4,000 cpm/μl, a 12.5 μl reaction may be appropriate with the addition of more pre-mRNA.
7. It is possible to add radiolabeled pre-mRNA to the Master Mix, rather than adding it separately.
8. When adding NE, make sure to prevent any air bubbles from forming. Mix gently by pipetting up and down, and *do not vortex*. Excessive bubbles may reduce splicing efficiency.

9. A time 0 tube should be prepared as a control. Once NE is added to the reaction tube, immediately place the tube on dry ice to prevent the splicing reaction from starting. This time 0 control treatment will be used to adjust for background intensity associated with un-spliced product for all the reactions.

10. The optimal temperature for cleavage at the 5′ splice site is 30 °C [10].

11. The splicing complexes formed on the pre-mRNA will not survive a dry ice freeze/thaw cycle. Therefore, only place the reaction on dry ice if the reaction will not be used to visualize native gel complex formation or for other downstream analyses. In the case of this protocol, only the spliced radiolabeled mRNA products are to be visualized. Therefore, destroying the spliceosomal complexes is not an issue.

12. Coating one of the plates with silicon is not required but highly recommended. A siliconized gel plate allows for easier separation when separating the glass plates. The gel will almost always stick to the uncoated plate, instead of partially sticking to both. Preferably, the notched plate should be siliconized.

13. Both the siliconized and non-siliconized plates should be free of any sort of particles and debris. Make sure to wipe away any debris, as they will form tiny air pockets between the glass plates that will cause leakage when pouring the gel.

14. Altering the amount of APS and TEMED can have different effects on gel polymerization and on how the samples run on the gel [11–13]. Generally a 1:150 dilution of 10 % APS and 1:1,000 dilution of TEMED are used.

15. Avoid the formation of bubbles while making the gel. Hold the clipped gel cassette (with notched plate facing upward) in one hand at a 45° angle, tilted on its corner. Slowly dispense the acrylamide solution. If an air bubble is present, adjust the angle of the cassette to allow the solution to force the air bubble outward. Add the comb immediately before the gel solution has time to harden.

16. Pre-running the gel before adding samples can remove all traces of (APS) and will apply a constant temperature to the gel before use [14].

17. Immediately before loading samples, make sure to flush out the wells with buffer to remove any urea that has leached and deposited into the wells.

18. The spliced RNA solution will be frozen when adding the Proteinase K Master Mix. Pipet the Master Mix up and down slowly in the reaction tube to thaw the spliced RNA. *do not vortex.*

19. Keep track of the orientation of the tubes while centrifuging; the pellet will be very hard to see and sometimes invisible.

If a pellet is not visible, continue to add the stop dye and load samples onto the gel (it is most likely there as the dye will stick to the pellet).

20. As mentioned previously, to prevent the glass plates from cracking and to ensure even conduction of heat, clamp a precooled aluminum plate to the front glass plate using the same binder clips used to hold the gel cassette in place. Make sure the aluminum plate is positioned so that it does not touch any buffer in the lower chamber. Run the gel for an appropriate amount of time; this will differ depending on the splicing products of your reaction and the percent/mix of the gel poured.

21. Make sure there are no creases in the plastic wrap. Remove any extra overhanging plastic wrap using a razor, being careful not to slice the gel. Any extra plastic wrap will bulge and may prevent the gel from being flush with the phosphor imaging screen or film.

22. The PhosphorImager screen is a form of autoradiography that is used to visualize and detect radioactive emission from radiolabeled RNA. Phosphor imaging screens contain BaFBR:Eu^{2+} crystals. When these crystals are exposed to ionizing radiation from radiolabeled RNA, electrons from Eu^{2+} become excited resulting in subsequent oxidation. During screening, the oxidized electrons revert back releasing a photon that can then be detected at certain wavelengths via a photomultiplier system producing a quantitative image [15]. There are many advantages to this method over other methods such as film. These advantages include increased sensitivity over a linear detection range of 5 orders of magnitude, while exposure to film is limited to only 1.5 orders of magnitude, increased exposure time from 10 to 250 times faster than film, easier and faster quantitation of images, and reuse of the phosphor screens indefinitely [16]. Other molecular detection systems similar to the Bio-Rad Molecular Imager are also available.

23. If the radiolabeled pre-mRNA used for the splicing reaction is around 8,000 cpm/μl, 1 h exposure to the PhosphorImager screen or film is sufficient to observe most splicing; however, longer exposures are often needed to see all splicing products or intermediates.

24. Due to the differential rates of decay among some splicing products, not all bands may be suitable for quantification. Depending on the in vitro reaction, lariat formation may be more stable than certain products and can be used as a substitute for calculating % spliced [17]. In addition, certain products may form which will not necessarily be stable in the cell (such as single exons). These RNAs will be degraded in the cell but may persist in an in vitro reaction.

25. % spliced product is obtained by calculating the volume intensity from the digital image of the splicing gel for each band in each lane. Briefly, calculate the sum of the adjusted intensity (taking into account the background of the gel and time 0 reaction) for the spliced and un-spliced band, divide the signal for spliced product by the total signal in the lane, and take the percent:

$$\%\ \text{Spliced} = \frac{\text{Signal from final spliced product}}{\text{Total signal in lane}} \times 100$$

$$(\text{Total signal in lane} = \text{spliced product} + \text{un-spliced product})$$

Acknowledgment

Research in the Hertel laboratory is supported by NIH (RO1 GM62287 and R21 CA149548).

References

1. Cooper TA (2005) Use of minigene systems to dissect alternative splicing elements. Methods 37:331–340
2. Beyer AL, Osheim YN (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev 2:754–765
3. Hicks MJ, Lam BJ, Hertel KJ (2005) Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays. Methods 37:306–313
4. Hernandez N, Keller W (1983) Splicing of in vitro synthesized messenger RNA precursors in HeLa cell extracts. Cell 35:89–99
5. Padgett RA, Hardy SF, Sharp PA (1983) Splicing of adenovirus RNA in a cell-free transcription system. Proc Natl Acad Sci USA 80:5230–5234
6. Hardy SF, Grabowski PJ, Padgett RA et al (1984) Cofactor requirements of splicing of purified messenger RNA precursors. Nature 308:375–377
7. Lee KA, Bindereif A, Green MR (1988) A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. Gene Anal Tech 5:22–31
8. Reichert V, Moore MJ (2000) Better conditions for mammalian in vitro splicing provided by acetate and glutamate as potassium counterions. Nucleic Acids Res 28:416–423
9. Krainer AR, Maniatis T, Ruskin B, Green MR (1984) Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell 36:993–1005
10. Furdon PJ, Kole R (1986) Inhibition of splicing but not cleavage at the 5′ splice site by truncating human beta-globin pre-mRNA. Proc Natl Acad Sci USA 83:927–931
11. Dirksen ML, Chrambach A (1972) Studies on the redox state in polyacrylamide gels. Separ Sci 7:747–772
12. Gelfi C, Righetti PG (1981) Polymerization kinetics of polyacrylamide gels. I. Effect of different cross-linkers. Electrophoresis 2: 213–219
13. Righetti PG, Gelfi C, Bosisio AB (1981) Polymerization kinetics of polyacrylamide gels. III. Effect of catalysts. Electrophoresis 2:291–295
14. Rio DC, Ares M, Hannon GJ, Nilsen TW (2010) Polyacrylamide gel electrophoresis of RNA. Cold Spring Harb Protoc 2010:1–6
15. Voytas D, Ke N (1999) Current protocols in molecular biology – detection and quantitation of radiolabeled proteins and DNA in gels and blots. Curr Protoc Mol Biol 48:A.3A.1–A.3A.10
16. Johnston RF, Pickett SC, Barker DL (1990) Autoradiography using storage phosphor technology. Electrophoresis 11:355–360
17. Kotlajich MV, Crabb TL, Hertel KJ (2009) Spliceosome assembly pathways for different types of alternative splicing converge during commitment to splice site pairing in the A complex. Mol Cell Biol 29:1072–1082

APPENDIX B

Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing

Molecular Cell
Article

Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing

Brandon E. Aubol,[1] Guowei Wu,[2] Malik M. Keshwani,[1] Maliheh Movassat,[3] Laurent Fattet,[1] Klemens J. Hertel,[3] Xiang-Dong Fu,[2] and Joseph A. Adams[1,*]
[1]Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093, USA
[2]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[3]Department of Microbiology and Molecular Genetics, University of California, Irvine, Irvine, CA 92697, USA
*Correspondence: j2adams@ucsd.edu
http://dx.doi.org/10.1016/j.molcel.2016.05.034

SUMMARY

Phosphorylation has been generally thought to activate the SR family of splicing factors for efficient splice-site recognition, but this idea is incompatible with an early observation that overexpression of an SR protein kinase, such as the CDC2-like kinase 1 (CLK1), weakens splice-site selection. Here, we report that CLK1 binds SR proteins but lacks the mechanism to release phosphorylated SR proteins, thus functionally inactivating the splicing factors. Interestingly, CLK1 overcomes this dilemma through a symbiotic relationship with the serine-arginine protein kinase 1 (SRPK1). We show that SRPK1 interacts with an RS-like domain in the N terminus of CLK1 to facilitate the release of phosphorylated SR proteins, which then promotes efficient splice-site recognition and subsequent spliceosome assembly. These findings reveal an unprecedented signaling mechanism by which two protein kinases fulfill separate catalytic features that are normally encoded in single kinases to institute phosphorylation control of pre-mRNA splicing in the nucleus.

INTRODUCTION

Alternative mRNA splicing, prevalent in higher eukaryotic genomes, has the potential to singularly magnify the proteome size from a rather limited set of genes in development and disease (Nilsen and Graveley, 2010; Wang and Cooper, 2007). Splicing is tightly coupled with transcription, nuclear export, and mRNA stability, thereby profoundly influencing the expression of gene products (Maniatis and Reed, 2002; Moore and Proudfoot, 2009; Pandit et al., 2008). Despite significant progress in the past two decades, we still have very limited knowledge of the molecular mechanisms underlying regulated splicing (Shin and Manley, 2004). Because up to 60% of disease-causing mutations are linked to defects in alternative splicing (Wang and Cooper, 2007), an understanding of splicing mechanisms and their regulation continues to represent a major challenge in the post-genome era.

Recent studies have elucidated the role of cellular signaling in the regulation of alternative splicing (Fu and Ares, 2014; Zhou et al., 2012). Most important among different regulatory strategies are the protein kinases and phosphatases that control the activities of various splicing factors through reversible phosphorylation (Stamm, 2008). Protein kinases are key regulators of various cellular processes and have uniquely evolved mechanisms to carry out selective phosphorylation of their substrates. Their substrate selectivity has been attributed, in part, to the catalytic domain that recognizes preferred amino acids (Endicott et al., 2012; Taylor and Kornev, 2011). Protein kinases also encode modular domains in their gene structure that impart secondary binding surfaces for substrate selection as well as serve as signals for their sub-cellular localization and guidance toward physiological substrates (Parsons and Parsons, 2004; Taylor et al., 2012). These modular domains can bind to scaffold proteins that serve as platforms for both the protein kinases and their cognate substrates (Parsons and Parsons, 2004). Additionally, protein kinases have also adopted dimerization strategies for activation and substrate phosphorylation. Some protein kinases form homodimers (e.g., receptor tyrosine kinases), whereas a few form heterodimers (Lavoie et al., 2014; Lemmon and Schlessinger, 2010). The heterodimer assembly between inactivated BRAF and active CRAF activates the latter during MEK-ERK signaling (Poulikakos et al., 2010). RAF heterodimerization suggests a critical disease mechanism, because such elegant arrangement was found in melanoma cells that develop resistance in the background of the most common BRAF (V600E) mutation (Lavoie et al., 2014; Poulikakos et al., 2010).

Alternative splicing is facilitated by the reversible phosphorylation of the SR protein family of splicing regulators (Pandit et al., 2008). Two families of protein kinases specifically phosphorylate these essential splicing factors. The serine-arginine protein kinases (SRPK1–3) strictly phosphorylate Arg-Ser dipeptides, whereas the CDC2-like kinases (CLK1–4) can phosphorylate both Arg-Ser and Ser-Pro dipeptides, common in all SR proteins (Ghosh and Adams, 2011) (Figure 1A). Although SRPKs use a docking groove in the C-terminal lobe of the kinase domain for substrate recognition and phosphorylation, CLKs lack such a

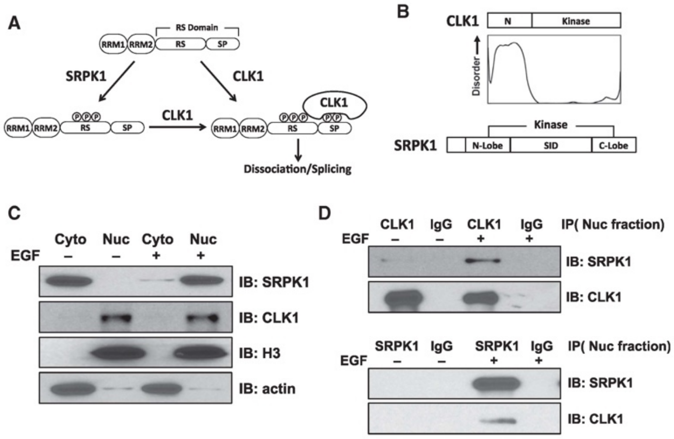


Figure 1. SRPK1 and CLK1 Form a Nuclear Complex

(A) SRPK1 phosphorylates Arg-Ser repeats (RS) and CLK1 phosphorylates these and Ser-Pro dipeptides (SP) in the SRSF1 RS domain.

(B) SRPK1 and CLK1 domain structures. CLK1 N terminus is predicted to be disordered using DISOPRED.

(C) Cytoplasmic and nuclear fractionations of SRPK1 and CLK1 in HeLa cells with and without EGF stimulation.

(D) Co-immunoprecipitation of SRPK1 and CLK1 in nuclear lysates of HeLa cells.

groove and instead use a disordered N-terminal domain to bind SR proteins with very high affinity (Figure 1B).

Although SRPKs are located in both the cytoplasm and nucleus, their function in the cytoplasm is best understood. SRPKs support phosphorylation-dependent transport of SR proteins from the cytoplasm to the nucleus via an SR-specific transportin protein, TRN-SR2 (Kataoka et al., 1999; Lai et al., 2001; Yun et al., 2003). Unlike SRPKs, CLKs possess a nuclear localization signal in their N termini and are thus localized strictly to the nucleus, where they play a vital role in mobilizing SR proteins from speckles to sites of active gene splicing via Ser-Pro phosphorylation (Keshwani et al., 2015a). This has led to a simple model in which SRPKs generate basal SR protein phosphorylation levels in the cytoplasm and CLKs enhance phosphorylation in the nucleus for splicing regulation (Ding et al., 2006; Keshwani et al., 2015a). However, this simplistic relay mechanism fails to explain two long-standing issues regarding the role of phosphorylation in alternative splicing. First, while phosphorylation is thought to enhance the activity of SR proteins in splice-site recognition, high levels of CLK paradoxically inhibit this function (Prasad et al., 1999). Second, as both SRPKs and CLKs are also present in the nucleus, this sets up a potential competitive rather than a cooperative relationship, because we recently found that although SRPK has built-in sequences that assist in the release of SR proteins after phosphorylation (Aubol et al., 2014; Koizumi et al., 1999), CLK does not release fully phosphorylated SR proteins (Aubol et al., 2014). These findings also suggest that CLKs may require a release factor to generate functional SR proteins for splicing function.

Epidermal growth factor (EGF) signaling increases nuclear levels of SRPKs and enhances bulk SR protein phosphorylation (Zhou et al., 2012). This mechanism involves the Akt-dependent release of several chaperones (Hsp70/90) from the spacer insert domain (SID) in SRPK1 that pins the kinase in the cytoplasm (Zhong et al., 2009) (Figure 1B). Such signaling affects the alternative splicing of numerous genes and hints to a role for SRPKs in the nucleus. Furthermore, osmotic stress increases nuclear SRPK levels by also shedding several chaperones, suggesting that multiple signals converge on this cytoplasmic kinase complex (Zhong et al., 2009; Zhou et al., 2012). Given the presence of SRPKs in the same compartment as the CLKs, we wish to further delineate the unique function of SRPKs in the nucleus. We here report that endogenous SRPK1 and CLK1 interact in the nucleus, forming a complex. Surprisingly, rather than acting in a competitive manner, the dual kinase complex acts symbiotically to associate with and phosphorylate their substrates with SRPK1, functioning as a release factor for CLK1, unleashing phosphorylated SR proteins for the promotion of spliceosome assembly. These findings solve some major puzzles from previous studies and, more importantly, reveal a signaling paradigm where two protein kinases are required to catalyze a full phosphorylation cycle (from binding to release) of their substrates.

RESULTS

SRPK1 Forms a Complex with CLK1 in Cells

SRPKs and CLKs have separate functions in the cell with regard to SR proteins. Whereas cytoplasmic SRPKs facilitate phosphorylation-dependent transport of SR proteins to the nucleus, CLKs perform additional phosphorylation steps that mobilize SR proteins for splicing activity in the nucleus (Keshwani et al., 2015a). What has gone unappreciated is that SRPKs are present not only in the cytoplasm but also in the nucleus (Colwill et al., 1996; Nayler et al., 1997; Wang et al., 1998). We showed in a previous study that, upon EGF stimulation, SRPK1 levels can be greatly increased in the nucleus, where CLKs are localized (Zhou et al., 2012), an observation that we confirmed through fractionation studies (Figure 1C). This raises the question of how the two kinases function in the same compartment. To begin to address this, we determined whether these kinases can interact with one another in the nucleus by co-immunoprecipitation using specific monoclonal antibodies. We found that anti-CLK1 efficiently co-immunoprecipitates with SRPK1 (Figure 1D, top) and, similarly, anti-SRPK1 can capture CLK1 (Figure 1D, bottom) in EGF-treated HeLa cells. These findings indicate that CLK1 and SRPK1 have the capacity to form a stable complex in the nucleus.

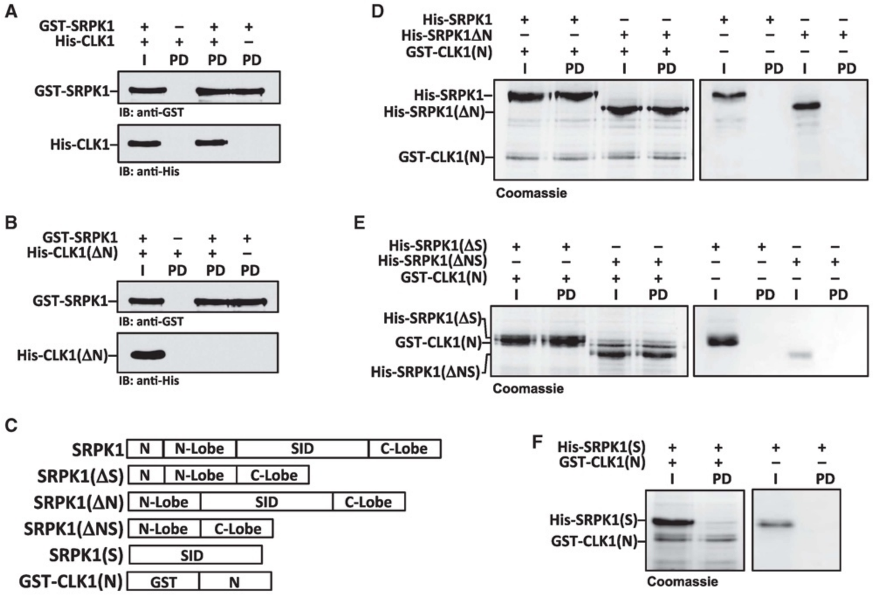


Figure 2. The CLK1 N Terminus Interacts with the SRPK1 Kinase Domain

(A and B) Pull-downs of GST-SRPK1 and His-CLK1 (A) or His-CLK1(ΔN) (B). I, input; PD, pull-down.

(C) Deletion constructs of His-SRPK1 and GST-CLK1(N).

(D–F) Pull-downs of His-SRPK1 deletions using GST-CLK1(N). I, input; PD, pull-down.

The CLK1 N Terminus Interacts with the SRPK1 Kinase Domain

We showed previously that the CLK1 N terminus binds with high affinity to the RS domain of its substrate SRSF1 (Aubol et al., 2014). Furthermore, we showed that the N terminus also interacts with its kinase domain but is not a substrate, suggesting that it can bind different proteins in an intra- and intermolecular fashion (Aubol et al., 2014; Colwill et al., 1996). To determine whether the CLK N terminus also interacts with SRPK, we performed in vitro pull-down experiments using glutathione S-transferase (GST)-tagged SRPK1 with either full-length His-tagged CLK1 or a form lacking its N terminus (CLK1(ΔN)). We found that GST-SRPK1 pulled down full-length CLK1, but not CLK1(ΔN) (Figures 2A and 2B). As controls, we showed that CLK1 and CLK1(ΔN) did not interact with the glutathione-agarose resin. These findings suggest that the N terminus is the principle domain that stabilizes the CLK-SRPK complex.

To determine how the N terminus interacts with SRPK1, we performed a series of pull-down experiments using a GST-tagged form of the CLK1 N terminus (GST-CLK1(N)) and several SRPK1 truncation mutants (Figure 2C). We found that GST-CLK1(N) interacts strongly with SRPK1 and all truncations that include the kinase domain (Figures 2D and 2E). In comparison, GST-CLK1(N) did not pull down SRPK1(S), suggesting that the N terminus does not form a stable complex with the SID (Figure 2F). As controls, we showed that all SRPK1 forms did not interact with glutathione-agarose resin without GST-CLK1(N) (Figures 2D–2F). Together, these results suggest that the CLK1 N terminus interacts with the SRPK1 kinase domain.

CLK1 Regulates Nuclear Levels of SRPK1

Although prior studies showed that SRPKs are maintained in the cytoplasm through chaperone binding (Hsp70/90) (Zhong et al., 2009), it is unclear whether there are nuclear constituents that help maintain nuclear pools of the kinase. We addressed whether complex formation with CLK anchors SRPK1 in the nucleus. By monitoring endogenous SRPK1 in HeLa cells using confocal microscopy, we found SRPK1 predominantly in the cytoplasm with a small fraction in the nucleus (Figure 3A), consistent with our fractionation studies and previous findings (Gui et al., 1994). We then transfected an RFP-tagged CLK1 (CLK1-RFP) that expressed to an observed level six times that of the endogenous kinase and found that it increased the nuclear SRPK1 levels relative to the cytoplasmic pools (Figures 3B and 3E). This result suggests that the SRPK1 cytoplasmic/nuclear levels are controlled by an equilibrium set by chaperones in the cytoplasm and CLK1 in the nucleus. Expressed CLK1 shifts this equilibrium toward greater levels of nuclear SRPK1 in accordance with mass action. To determine whether this phenomenon is dependent on catalytic activity, we expressed a kinase inactive CLK1 (kdCLK1-RFP) and found that it also induced nuclear SRPK1 (Figures 3C and 3E). To determine whether the

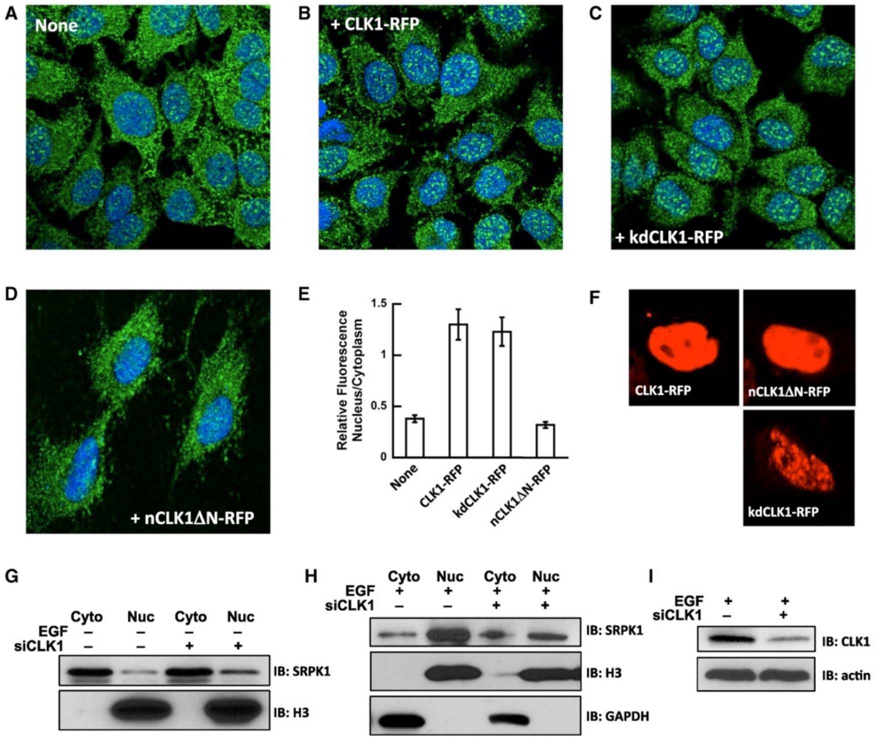


Figure 3. CLK1 Expression and SRPK1 Subcellular Localization

(A–D) Confocal imaging of endogenous SRPK1 (green) in HeLa cells transfected with mock (A), CLK1-RFP (B), kdCLK1-RFP (C), and nCLK1ΔN-RFP (D). Nuclei are stained with DAPI (blue).

(E) Relative amounts of nuclear and cytoplasmic SRPK1 with and without CLK1 expression. Fractional amounts are calculated using ImageJ. The data represent an average of n = 3, and error bars indicate ± SD.

(F) Confocal imaging of CLK1-RFP, kdCLK1-RFP and nCLK1ΔN-RFP showing nuclear localization.

(G and H) Nuclear and cytoplasmic SRPK1 levels in HeLa cells treated with CLK1 siRNA in the absence (G) and presence (H) of EGF stimulation.

(I) CLK1 levels in HeLa cells treated with CLK1 siRNA.

N-terminal RS-like domain of CLK1 plays a role in nuclear retention of SRPK1, we expressed a version of CLK1 that replaces the N terminus with a nuclear localization sequence from nucleoplasmin 2 (KRLVPQKQASVAKKKK) and found that this new construct, nCLK1ΔN-RFP, did not increase nuclear SRPK1 levels (Figures 3D and 3E). CLK1-RFP, kdCLK1-RFP, and nCLK1ΔN-RFP are all exclusively present in the nucleus, indicating that the increased nuclear SRPK1 is due to increased nuclear CLK1 (Figure 3F).

We next addressed whether CLK1 is required for SRPK1 nuclear localization by disrupting CLK1 using small interfering RNA (siRNA) methods. CLK1 depletion did not affect SRPK1 levels in unstimulated cells, where most of the kinase is present in the cytoplasm (Figure 3G). However, in EGF-stimulated cells when SRPK1 is mostly in the nucleus, CLK1 depletion had little effect on the levels of SRPK1 in the cytoplasm but prevented the accumulation of SRPK1 in the nucleus (Figure 3H), likely due to unstable SRPK1 in the nucleus in the absence of CLK1. In control experiments, we showed that siRNA-treated cells displayed a significant decrease in CLK1 (Figure 3I). These observations suggest that CLK1 acts as a nuclear anchor for SRPK1 through complex formation and that the CLK1 N terminus is necessary for this function.

SRPK1 Releases Phospho-SRSF1 from CLK1

Although SRPK1 forms a highly stable complex with its substrate, SRSF1, multi-site RS-domain phosphorylation induces dissociation (Aubol and Adams, 2011; Ma et al., 2010). We verified these results in pull-down assays by incubating GST-SRSF1 with SRPK1 in the absence and presence of ATP. GST-SRSF1, bound to glutathione beads, strongly pulls down SRPK1 in the absence, but not in the presence, of ATP (Figure 4A). In

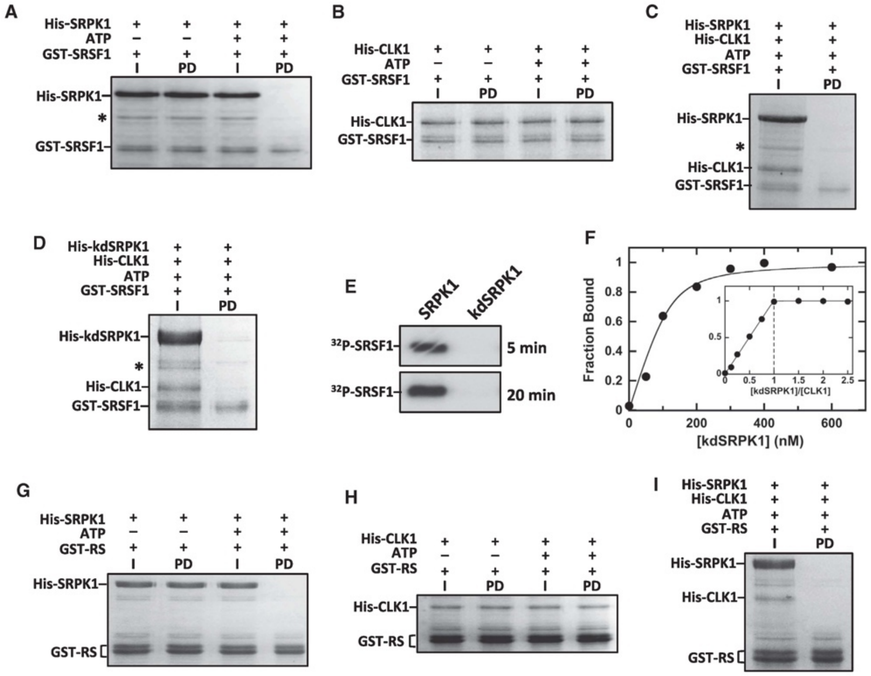


Figure 4. SRPK1 Releases CLK1 from Phosphorylated SRSF1

(A and B) Phosphorylation disrupts SRSF1 interaction with SRPK1, but not CLK1. Immobilized GST-SRSF1 is incubated with His-SRPK1 (A) or His-CLK1 (B) with ATP, washed, and run on SDS-PAGE. I, input; PD, pull-down.

(C and D) Binding of phosphorylated SRSF1 to His-CLK1 is disrupted by His-SRPK1 (C) and His-kdSRPK1 (D). Immobilized GST-SRSF1 is incubated with His-CLK1 and ATP and treated with His-SRPK1 or His-kdSRPK1. I, input; PD, pull-down.

(E) SRSF1 (0.2 μM) is phosphorylated by SRPK1 (1 μM), but not kdSRPK1 (1 μM).

(F) Complex affinity and stoichiometry. His-CLK1 fraction bound versus His-kdSRPK1 using 100 nM His-CLK1 is fit to Equation 1 to obtain a K_d of 16 ± 6 nM and R_o of 120 ± 17 nM. Fraction bound using 400 nM His-CLK1 is plotted as a ratio of total His-kdSRPK1 to His-CLK1, establishing a complex stoichiometry of 1:1.

(G and H) Phosphorylation disrupts RS domain interactions with His-SRPK1, but not His-CLK1. GST-RS is incubated with His-SRPK1 (G) or His-CLK1 (H) in the presence of ATP, bound to glutathione-agarose resin, washed, and run on SDS-PAGE. I, input; PD, pull-down.

(I) Binding of phosphorylated RS domain to His-CLK1 is disrupted by His-SRPK1. GST-RS is incubated with His-CLK1 and ATP, treated with His-SRPK1, bound to glutathione-agarose resin, washed, and run on SDS-PAGE. I, input; PD, pull-down.

comparison, although CLK1 also forms a very stable complex with SRSF1 similar to that observed with SRPK1, phosphorylation of the RS domain does not reduce the stability of the CLK1-SRSF1 complex (Figure 4B). This suggests that, when acting alone, CLK1 would behave as an inhibitor by titrating free SR proteins in the nucleus.

Given the ability of SRPK1 to interact with CLK1, we were intrigued by the possibility that SRPK1 might serve as a release factor for phospho-SRSF1 from CLK1. To this possibility, we performed pull-down experiments using phosphorylated GST-SRSF1 in the presence of SRPK1. We showed that CLK1-phosphorylated GST-SRSF1 readily dissociates CLK1 in the presence of SRPK1 (Figure 4C). This function does not depend on catalytic activity, because the kinase-inactive SRPK1 (kdSRPK1) also leads to dissociation (Figure 4D). In control experiments, we showed that kdSRPK1 is inactive and does not phosphorylate SRSF1 in the time frame of the exchange reaction (Figure 4E). Thus, while CLK1 forms a complex with phosphorylated SRSF1, SRPK1 facilitates the release of the splicing factor from CLK1. This process is not a simple competitive event with regard to the SR protein, because SRPK1 does not bind phosphorylated SRSF1 under these conditions (Figure 4A).

To measure the affinity of the SRPK1-CLK1 complex, we used the ability of SRPK1 to displace phospho-SRSF1 from CLK1 as a reporter for the kinase-kinase complex. We phosphorylated GST-SRSF1 (500 nM) with His-tagged CLK1 (100 nM) and ^{32}P-ATP, bound the kinase to the Ni-resin, and then added varying kdSRPK1 to displace the phospho-SR protein (Figure 4F). We used kdSRPK1 to avoid any potential additional phosphorylation of the substrate. The fraction bound of CLK1 is plotted against

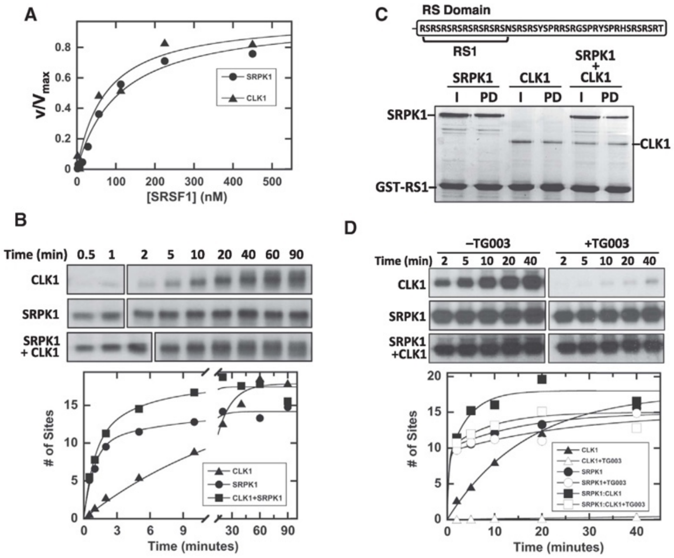


Figure 5. SRPK1-CLK1 Complex Phosphorylates SRSF1

(A) SRSF1 binds with similar, high affinity to CLK1 and SRPK1. Initial velocities with varying SRSF1 are fit to K_m values of 70 ± 10 and 110 ± 10 nM for CLK1 and SRPK1.

(B) Single turnover curves for SRSF1 phosphorylation using SRPK1, CLK1, and the SRPK1-CLK1 complex. Full time courses were analyzed on separate gels as indicated by line breaks. SRSF1 phosphorylation with CLK1 is fit to a rate constant and amplitude of 0.065 ± 0.007 min^{-1} and 18 ± 1 sites. SRSF1 phosphorylation with SRPK1 or the SRPK1-CLK1 complex is fit to rate constants and amplitudes of 1.1 ± 0.1 and 0.11 ± 0.04 min^{-1} and 10 ± 1 and 4 ± 0.5 sites for SRPK1 or 1.1 ± 0.2 and 0.22 ± 0.10 min^{-1} and 10 ± 1 and 7.4 ± 0.6 sites for the SRPK1-CLK1 complex.

(C) SRPK1 and CLK1 bind the RS domain as a complex. Immobilized GST-RS1 is mixed with SRPK1, CLK1, and the SRPK1-CLK1 complex, washed, and run on SDS-PAGE.

(D) Single-turnover curves for SRSF1 phosphorylation using SRPK1, CLK1, and the SRPK1-CLK1 complex with and without TG003.

total kdSRPK1 to obtain an observed K_d of 16 nM using Equation 1, a quadratic function for tight-binding ligands (Taira and Benkovic, 1988) (Figure 4F). We next repeated this experiment using higher CLK1 concentrations (400 nM) and were able to titrate CLK1 with kdSRPK1, establishing a kdSRPK1-CLK1 stoichiometry of 1:1 (Figure 4F, inset). To address whether the RNA recognition motifs (RRMs) are necessary for SRSF1 displacement by SRPK1, we performed pull-down assays using a substrate lacking these domains (GST-RS). We found that although GST-RS binds SRPK1 and CLK1 tightly, phosphorylation leads to dissociation of SRPK1, but not CLK1, similar to that for the full-length substrate (Figures 4 G and 4H). Similar to the full-length SR protein, SRPK1 dissociates the CLK1-pRS complex, suggesting that the RRMs are not involved in the release mechanism (Figure 4I). Overall, the data show that, in addition to phosphorylating Arg-Ser dipeptides in SRSF1, SRPK1 has a secondary function in releasing the tightly bound phospho-SRSF1 from CLK1.

SRPK1-CLK1 Complex Phosphorylates the RS Domain of SRSF1

To determine whether the SRPK-CLK complex functions productively as a dual-kinase system, we performed several kinetic experiments. We initially confirmed that both SRPK1 and CLK1 interact with high affinity to SRSF1 in kinetic assays. We showed that the K_m for SRSF1 to CLK1 (70 nM) is lower than that for SRPK1 (110 nM), consistent with high-affinity binding observed previously (Aubol et al., 2013) (Figure 5A). To address whether these kinases compete for the RS domain of SRSF1, we performed single-turnover experiments in which large amounts of each kinase (3 μM) are used to phosphorylate a lower amount of SRSF1 (0.3 μM). In prior competition studies, we showed that the K_d for SRSF1 is 60 nM to SRPK1 and 6 nM to CLK1, so that no free substrate is present and the reaction is performed under true single-turnover conditions (Aubol et al., 2013). Both kinases achieve high levels of SRSF1 phosphorylation, with SRPK1 attaining multi-site phosphorylation at a faster rate than CLK1 (>10-fold), in keeping with prior studies (Aubol et al., 2014) (Figure 5B). SRPK1 adds ten phosphates very rapidly ($t_{1/2}$ = 0.7 min) and another five phosphates in a second, slower phase ($t_{1/2}$ = 7 min). The phosphorylation of SRSF1 by CLK1 is monophasic and comparatively slow ($t_{1/2}$ = 11 min).

Although CLK1 binds better than SRPK1 to SRSF1 and is a much slower kinase, an equimolar amount of CLK1 did not slow down the SRPK1 reaction (Figure 5B). These data suggest that SRPK1 does not simply compete with CLK1 for the substrate but rather takes on a dominant role in phosphorylating the RS domain when both kinases are present. Furthermore, because the concentrations of CLK1 and SRPK1 are equal and exceed the K_d for the complex by two orders of magnitude (Figure 4F), the dominant catalytic species is the complex under our single-turnover conditions. Although it has no observable effect on the first kinetic phase, SRPK1 increases the net rate of RS-domain phosphorylation, as the second phase rate constant in the complex is ~3-fold larger than the rate constant for CLK1 by itself (0.22 versus 0.065 min^{-1}), suggesting kinase-kinase allosteric interactions. To rule out the possibility that CLK1 and SRPK1 may bind to distinct regions of the RS domain (50 aa), thus allowing two separate reactions on the same substrate, we expressed a short form of the RS domain (16 aa) that should bind only one kinase at a time (Figure 5C). We showed previously that the RS1 segment of SRSF1 binds in both the docking groove

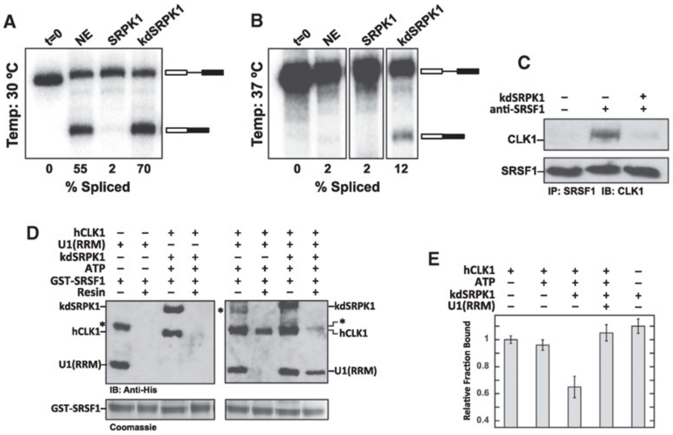


Figure 6. SRPK1-Induced Release Affects Splicing of β-Globin Pre-mRNA

(A and B) Splicing of the β-globin gene in nuclear extracts in the absence and presence of exogenous SRPK1 and kdSRPK1 at 30°C (A) and 37°C (B). Several unrelated lanes were removed from the gel in (B).

(C) SRPK1 induces release of SRSF1 from CLK1 in nuclear extracts. SRSF1 is immunoprecipitated and probed for CLK1 with and without exogneous kdSRPK1. Top (first lane): nuclear lysate with only protein G beads and no SRSF1 antibody.

(D) CLK1 release promotes binding of U1(RRM) to SRSF1. Interaction of U1(RRM) with GST-SRSF1 bound to glutathione-agarose beads is probed using an anti-His antibody. An asterisk represents an impurity in the U1(RRM) preparation.

(E) Binding of the Ron ESE to SRSF1 with and without CLK1 and kdSRPK1. The data are an average of n = 3. Error bars indicate ± SD.

and active site of SRPK1 and thus cannot permit simultaneous association of both SRPK1 and CLK1 (Ngo et al., 2008). We showed that GST-RS1 binds well to both SRPK1 and CLK1 in separate pull-down experiments (Figure 5C). However, SRPK1 and CLK1 binding was not affected when both were added, a result that is not consistent with a competitive relationship between the kinases. Instead, these findings suggest that both kinases can bind simultaneously to RS1. Given the limited binding surface of RS1, these observations suggest that CLK1 and SRPK1 form a stable complex that fully phosphorylates the RS domain of SRSF1.

SRPK1 Is the Primary Catalyst in the SRPK1-CLK1 Complex

Having shown that SRPK1 and CLK1 phosphorylate the SRSF1 RS domain as a complex, we next wished to determine which of the kinases take the principle role in phosphorylating the RS domain. To address this, we performed single-turnover experiments in which a CLK-specific inhibitor, TG003, was added to the SRPK1-CLK1 complex prior to reaction initiation. If SRPK1 is the principle kinase that binds the RS domain and rapidly phosphorylates the Arg-Ser dipeptides whereas CLK1 mostly modifies the Ser-Pro dipeptides, then we expect that TG003 addition will not affect the rapid, initial kinetic phase in the progress curve for the complex. Indeed, we observed that TG003 did not impact the fast phase for the complex but instead slowed down later phosphorylation events, suggesting that SRPK1 is the primary kinase that binds the RS domain and performs rapid Arg-Ser phosphorylation (Figure 5D). To confirm inhibitor efficacy, we showed that TG003 addition to CLK1 by itself significantly inhibited SRSF1 phosphorylation (Figure 5D). These data suggest that SRPK1 takes on a dominant role in binding the RS domain in the SRPK1-CLK1 complex, rapidly phosphorylating Arg-Ser dipeptides. By itself, CLK1 can also phosphorylate Arg-Ser dipeptides, but in the complex, is likely to take on the specialized function of phosphorylating Ser-Pro dipeptides.

SRPK1-Induced Release Enhances Splicing of β-Globin Pre-mRNA

Having demonstrated that SRPK1 is a release factor for CLK1 in vitro, we wished to determine whether SRPK1 could also serve a similar role in nuclear extracts to regulate splicing. To address this possibility, we studied splicing of β-globin pre-mRNA in HeLa cell nuclear extracts. We found that the addition of SRPK1 to the extracts significantly inhibited β-globin pre-mRNA splicing (Figure 6A). This result was previously ascribed to a misbalance in the phosphorylation-dephosphorylation cycle necessary for spliceosome assembly and subsequent catalysis (Gui et al., 1994). As a control, we added kdSRPK1 to the nuclear extracts and, unexpectedly, found that it enhanced splicing of β-globin pre-mRNA (Figure 6A).

In vitro splicing assays are typically performed at 30°C, an optimum temperature for nuclear extracts (Movassat et al., 2014; Xiao and Manley, 1998). Given that SRPKs are part of the spliceosome (Mathew et al., 2008), we wished to determine whether the addition of exogenous SRPK1 could stabilize the spliceosome, thus enabling splicing at non-permissible temperatures. We showed that while nuclear extracts are not capable of facilitating splicing at 37°C, kdSRPK1 addition indeed stimulated splicing function (Figure 6B). These findings indicate that SRPK1 can facilitate splicing in a phosphorylation-independent manner. To determine whether the general splicing enhancement at either temperature could be linked to SRPK1-dependent substrate release, we monitored the interaction of CLK1 and SRSF1 in nuclear extracts. We found that although endogenous CLK1 interacts with endogenous SRSF1 in nuclear extracts, the addition of recombinant kdSRPK1 dissociates this complex (Figure 6C). These findings strongly suggest that SRPK1-induced release of SRSF1 from CLK1 may account for enhanced splicing.

CLK1 Dissociation Induces U1 RRM Binding to SRSF1

Previous studies from the Ghosh lab showed that a phosphomimetic form (poly-serine-to-glutamate mutant) representing the

CLK-phosphorylated state of SRSF1 associated favorably with the RRM from the 70K subunit of U1 small nuclear ribonucleoprotein particle (snRNP) (U1(RRM)) whereas the unphosphorylated, wild-type SRSF1 did not (Cho et al., 2011). These studies suggest that CLK-phosphorylated SRSF1 supports establishment of the 5′ splice site through interactions between the RRMs from SRSF1 and U1 snRNP. Given the pivotal role for CLK1 in establishing this interaction, we wished to determine whether SRPK1-induced release of SRSF1 could play a role in this key step of splicing initiation. We performed pull-down assays and found that U1(RRM) did not interact with unphosphorylated GST-SRSF1, consistent with the previous study (Cho et al., 2011) (Figure 6D, left, second lane). Surprisingly, we found that CLK1-phosphorylated GST-SRSF1 also did not interact with U1(RRM), suggesting that phosphorylation alone is not sufficient to initiate binding (Figure 6D, right, second lane). Strikingly, kdSRPK1 addition dissociates CLK1 and strongly promotes U1(RRM) binding to SRSF1 (Figure 6D, right, last lane). These findings indicate that SRPK1-induced release of CLK1-phosphorylated SRSF1 is needed to support the binding of U1(RRM) to SRSF1.

U1 RRM Binding and SRPK1-Induced Release Alters RNA Binding to SRSF1

SR proteins promote pre-mRNA splicing by binding to exonic splicing enhancers (ESEs) in pre-mRNA. We wished to define the potential role of CLK1 phosphorylation and SRPK1-induced release on this recognition mechanism. To accomplish this, we monitored SRSF1 binding to the ESE from the Ron proto-oncogene (5′-AGGCGGAGGAAGC-3′) using a filter-binding assay (Aubol et al., 2014). We found that SRSF1 binds the ESE similarly in the presence of CLK1 whether or not the RS domain was phosphorylated (Figure 6E, first two columns), indicating that CLK1 does not interfere with the RNA binding property of the SR protein. In contrast, kdSRPK1 addition to the CLK1-pSRSF1-RNA complex reduced RNA binding by ~35%, suggesting that CLK1 release causes the splicing factor to adopt a less productive conformation for RNA binding (Figure 6E, third column). As control, we showed that kdSRPK1 alone does not reduce ESE affinity to SRSF1 in the absence of phosphorylation (Figure 6E, last column). Interestingly, U1(RRM) addition to CLK1-phosphorylated SRSF1 lacking bound CLK1 led to restored high-affinity binding of the SR protein to the ESE (Figure 6E, fourth column). Together, these findings suggest that, while SRPK1-induced release generates a form of SRSF1 that binds poorly to RNA in isolation, the released SR protein favors the formation of the ternary complex with U1 and ESE, which likely reflects the active process in 5′ splice site recognition in the cell.

DISCUSSION

Enzymes have highly evolved active sites containing residues that specifically bind substrates with high affinity and position select functional groups for the ensuing catalytic steps. Indeed, the front end and internal portion of the enzymatic reaction is typically so fine-tuned that the product is oftentimes not easily released, thus leading to buildup of an inhibitory complex. Enzymes have accordingly developed strategies to destabilize the product so that catalytic cycling can proceed unimpeded. The splicing kinase SRPK1 binds the RS domain of the SR protein SRSF1 with unusually high affinity and then initiates a series of fast phosphoryl transfer steps. During these modifications, significant negative charge develops that clashes with the negatively charged docking groove in the kinase domain. The result is a progressive destabilization of the phospho-RS domain and rapid release of the SR protein. Transient-state kinetic studies showed that the binding affinity of the product is reduced by about two orders of magnitude after eight phosphate additions, the minimum number required for translocation into the nucleus (Aubol and Adams, 2011). In contrast, CLK1 incorporates an alternative strategy for SR protein recognition that uses a disordered N-terminal extension rather than a docking groove in the kinase domain (Aubol et al., 2014). This thematic variation is highly effective for binding SR proteins with very high affinity but does not offer a discriminatory mechanism between substrate and product. This observation raises the question of how CLKs release phospho-SR proteins for splicing.

Our results describe a unique paradigm in which two splicing kinases (CLK and SRPK) develop a symbiotic relationship to functionally compensate for their limitations and cooperatively facilitate the release of phospho-SRSF1. In the nucleus, SRPK1 strips the N terminus from the phospho-RS domain, thereby destabilizing its interaction with CLK1 to generate free SR protein (Figure 7A). This mechanism bears some resemblance to a classic exchange process akin to the guanine exchange factors (GEFs) that destabilize and promote GDP release from the Rho GTPases (Rossman et al., 2005). However, unlike a simple inert exchange factor, SRPK1 is also a catalyst and works symbiotically with CLK1 to yield phosphorylated SR proteins in the nucleus. This symbiosis generates an active SRPK1-CLK1 complex that sheds the limitations of each individual kinase by incorporating the favorable attributes of both kinases (Figure 7B). Although SRPK1 is a very agile kinase that rapidly phosphorylates Arg-Ser dipeptide repeats in a C-to-N-terminal direction, it cannot modify Ser-Pro dipeptides that are common in SR proteins and whose phosphorylation mobilizes the splicing factors and alters splicing patterns (Keshwani et al., 2015a). CLK1 can phosphorylate Ser-Pro dipeptides along with Arg-Ser repeats, but only at very reduced rates. In the hetero-kinase complex, SRPK1 takes on a superior position by binding and rapidly phosphorylating the lengthy Arg-Ser repeats in SRSF1. With its N terminus re-purposed for interaction with SRPK1, the kinase domain of CLK1 can freely pivot within the complex and modify the individual Ser-Pro dipeptides in the RS domain. We believe that Ser-Pro dipeptide modification occurs later in the multisite reaction, because a CLK1-specific inhibitor does not affect the rapid initial phase of the progress curves. There is likely to be some allosteric effects in the complex, as the second reaction phase is much faster than the net rate of SR phosphorylation by CLK1 alone. In all, the heterocomplex avoids a competitive roadblock by melding the speed and dissociative mechanism of SRPK1 with the dipeptide specificity of CLK1 to achieve a fully phosphorylated SR protein that participates in spliceosome assembly.

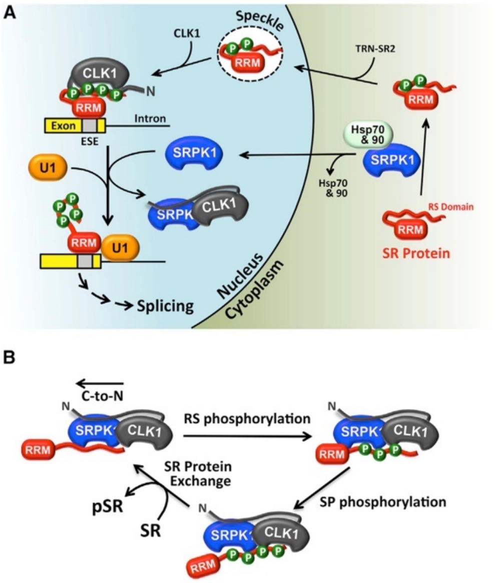


Figure 7. Role of Nuclear SRPK1 for SR Protein Phosphorylation and Splicing

(A) SRPK1 is drawn into the nucleus through contacts with CLK1. The SRPK1-CLK1 complex releases the SR protein from CLK1 and promotes splicing activity through enhanced contacts with U1.

(B) Catalytic cycle for SR protein phosphorylation by the SRPK1-CLK1 complex. The complex is held together by interactions between the N terminus of CLK1 and the kinase domain of SRPK1. SRPK1 is the primary catalyst that phosphorylates Arg-Ser repeats in a C-to-N-terminal direction, and CLK1 primarily phosphorylates Ser-Pro dipeptides. SRPK1 induces SR protein release by prohibiting contacts with the RS domain.

In addition to demonstrating how the dual-kinase mechanism solves the biochemical release problem for CLK1, the SRPK1-CLK1 complex also provides a framework for understanding the biological role of these kinases in splicing regulation. Although we showed previously that SRPKs localize in the cytoplasm through interactions with molecular chaperones (Zhong et al., 2009), we could not explain why SRPKs are also found in the nucleus without a localization sequence. Furthermore, we could not explain why removal of the SID that interacts with the cytoplasmic chaperones leads to complete nuclear localization of SRPK1 (Ding et al., 2006). These phenomena can now be explained by the ability of CLK1 to retain SRPK1 in the nucleus, balancing the cytoplasmic tethering of the chaperones (Figure 7A).

Once in the nucleus, SRPK1 serves a vital function in splicing. Prior studies using phospho-mimics of SRSF1 suggest that CLK1 phosphorylation promotes interaction of a component of the 70K subunit of U1 snRNP with the SR protein and subsequent formation of the 5′ splice site in pre-mRNA (Cho et al., 2011). We now show that CLK1-dependent phosphorylation is not sufficient for this process. Instead, CLK1 behaves as an inhibitor of splicing when bound to phosphorylated SRSF1. This explains the early observation that overexpressed CLK1 gave rise to similar functional consequences to SR protein depletion in transfected cells (Prasad et al., 1999). Moreover, formation of the SRPK1-CLK1 heterocomplex shifts internal binding contacts and promotes the association of the U1 RRM with phospho-SRSF1, a key step in establishing the 5′ splice site. Signals that re-attach SRPK1 to cytoplasmic chaperones (e.g., downregulation of Akt) are then expected to increase cytoplasmic SRPK1 levels, leading to dissociation of the SRPK1-CLK1 complex and resumption of CLK1-bound SR proteins.

Traditionally, SRPKs and CLKs have been viewed in light of their spatial orientation in eukaryotic cells. SRPKs phosphorylate Arg-Ser dipeptides in RS domains, releasing phospho-SR proteins for attachment to TRN-SR and nuclear import. Partially phosphorylated SR proteins are further phosphorylated at Ser-Pro dipeptides by nuclear CLKs, a modification that shunts the splicing factors toward the splicing machinery. We have now uncovered an arrangement of these two kinases in the nucleus that helps explain the mechanism of SR protein-dependent splicing that needs both kinases in the nucleus. The SRPK1-CLK1 complex brilliantly avoids the dilemma of establishing two kinases in a single compartment that, owing to their similar high affinities, would compete for the SR protein pool. Such elegant intracellular symbiosis between kinases in the nucleus explains how the inhibitory effects of CLK1 are removed to facilitate spliceosome assembly and the splicing reaction upon signaling-induced nuclear translocation of SRPK1.

EXPERIMENTAL PROCEDURES

Materials

ATP, Mops, Tris, $MgCl_2$, NaCl, EDTA, glycerol, sucrose, acetic acid, lysozyme, DNase, RNase, Phenix imaging film, BSA, Protein G agarose, Ni-resin, and liquid scintillant were obtained from Fisher Scientific. ^{32}P-ATP was obtained from NEN Products. RNA (5′-AGGCGGAGGAAGC-3′) was purchased from Integrated DNA Technologies. siRNA for *CLK1* was obtained from Bioneer. Protease inhibitor cocktail was obtained from Roche, and TG003 was obtained from Sigma. Anti-CLK1 monoclonal antibody was purchased from Aviva Systems Biology, and anti-SRPK1 monoclonal antibody was purchased from BD Biosciences. InstantBlue was purchased from Expedeon, Hybond ECL nitrocellulose blotting membrane was purchased from Amersham, and the KinaseMax Kit was purchased from Ambion.

Expression and Purification of Proteins

SRPK1, SRSF1, and CLK1(ΔN) (residues 148–484) were expressed and purified from pET19b vectors with an N-terminal His tag, and GST-SRSF1, GST-RS1, and GST-CLK1(N) (residues 1–160) were expressed and purified from a pGEX vector as previously described (Aubol et al., 2014). kdSRPK1, inactive SRPK1 containing K109M, was expressed and purified from a pET19b vector. Deletion constructs SRPK1(ΔN), SRPK1(ΔS), and SRPK1(S) (residues 222–492) were expressed and purified as previously described (Aubol et al., 2012, 2014). CLK1 virus was transfected and expressed in Hi5 insect cells, and CLK1 was purified with a nickel resin and a previously described procedure (Keshwani et al., 2015b).

Cell Fractionation Studies

HeLa cells were harvested and lysed in 10 mM HEPES (pH 7.9), 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, 0.05% Triton, and protease inhibitor cocktail. The lysates were centrifuged at 228 × *g* for 10 min to pellet nucleus, and the supernatant was retained as the cytoplasmic fraction. The pellet was washed with lysis buffer and re-suspended in 200 μl of 0.25 mM sucrose and 10 mM $MgCl_2$. The nuclear suspension was layered on 0.88 mM sucrose and 0.5 mM $MgCl_2$ and spun at 2,800 × *g* to obtain a nuclear pellet. The pellet was re-suspended in 1X RIPA buffer, spun at 2,800 × *g*, and the supernatant was retained as nuclear fraction.

Immunoprecipitation Experiments

HeLa cell nuclear lysates (200 μL) were pre-cleared with Protein A beads and incubated overnight in the cold room with 25 μL Protein A beads and either 3 μl rabbit anti-CLK1 (Aviva systems) or mouse anti-SRPK1 antibody (BD Biosciences). The beads were spun at 2,052 × *g* and washed with 200 μL lysis buffer followed by the addition of 2× SDS loading buffer. The heated slurry was run on a 12% SDS-PAGE followed by immunoblot analysis. SRSF1 was immunoprecipitated from nuclear lysates (100 μL) using 3 μL SRSF1 antibody with and without 5 μM kdSRPK1 and probed using anti-CLK1 antibody.

Confocal Imaging

HeLa cells were plated on 2.5-cm^2 MatTek poly-D-lysine plates and transfected with mock, CLK1-RFP, kdCLK1-RFP, and nCLK1ΔN-RFP DNA constructs (2 μg) for 24 hr. The cells were washed with PBS and fixed with 2% paraformaldehyde in PBS for 20 min followed by a 2× PBS wash. Cell permeabilization was done with 0.25% Triton in PBS for 10 min followed by a 3× PBS wash. Cells were then blocked in 10% donkey serum PBS for 1 hr at room temperature followed by overnight incubation with SRPK antibody at 4°C. Cells were washed three times with PBS, incubated with secondary fluorescent anti-mouse antibody conjugated with Alexa 488 (Life Technologies) for 1 hr at room temperature, followed by 3× PBS washes for 5 min. Cells were mounted with DAPI-containing mounting medium (Vector Laboratories). For live-cell imaging, transfected HeLa cells were analyzed using an Olympus FV1000 as described previously (Keshwani et al., 2015a).

Pull-Down Assays

GST- and His-tagged proteins (4 μM) were incubated in 40 μL binding buffer (0.1% NP40 [Nonidet P40], 20 mM Tris/HCl [pH 7.5], and 75 mM NaCl) for 30 min before incubating with 25 μL glutathione–agarose resin for 30 min at room temperature. Where phosphorylation was performed, GST- and His-tagged enzymes (4 μM) were mixed with and without 100 μM ATP in 10 mM Mg^{2+}, 20 mM Tris/HCl (pH 7.5), and 75 mM NaCl at 37°C for 30 min, followed by incubation with 25 μL glutathione-agarose resin for 30 min at room temperature. In some cases, 4 μM kdSRPK was added and incubated at room temperature for 10 min prior to wash steps. Resin was washed four times with 200 μL binding buffer, and the bound proteins were eluted with SDS quench buffer and boiled for 5 min. Retained protein was resolved by SDS-PAGE (12% gel) and visualized by Instant Blue Coomassie stain or western blotting with a mouse anti-His antibody.

Phosphorylation Assays

Single-turnover assays were performed in assay buffer (100 mM Mops [pH 7.4], 10 mM Mg^{2+}, and 5 mg/mL BSA), at 37°C using 3 μM enzyme, 300 nM SRSF1 and 100 μM ^{32}P-ATP (4,000–8,000 cpm/pmol) with and without 50 μM TG003 and 5% DMSO. Steady-state assays using 5 nM SRPK1 or 25 nM CLK1 were performed in assay buffer with 100 μM ^{32}P-ATP (4,000–8,000 cpm $pmol^{-1}$) at 23°C (SRPK1) and 37°C (CLK1) and quenched with 10 μL SDS/PAGE loading buffer at 2 and 20 min, respectively. Phosphorylated SR proteins were cut from a dried 12% SDS-PAGE and counted in liquid scintillant. The amount of kdSRPK1 bound to CLK1 was measured by phosphorylating GST-SRSF1 (500 nM) with CLK1 (100 or 400 nM) and 100 μM ^{32}P-ATP for 90 min at 37°C and binding to the Ni-resin. Varying kdSRPK1 (0–1,000 nM) was added to CLK1-bound resin and washed two times with 400 μl binding buffer. The CLK1 fraction bound (FB) was measured from the CPMs on the resin and fit to Equation 1:

$$FB = \frac{R_o + L_o + K_d - \sqrt{(R_o + L_o + K_d)^2 - 4K_dL_o}}{2R_o} \qquad \text{(Equation 1)}$$

where R_o is the total CLK1, L_o is the total kdSRPK1, and K_d is the complex dissociation constant.

RNA Preparation

In vitro transcription was performed using 0.5 μg/μL linear β-globin DNA in a reaction buffer containing 0.67 μM (α-^{32}P) UTP, 0.4 mM ATP, 0.4 mM CTP, 0.1 mM UTP, 0.1 mM GTP, 2 mM m7G(5′)ppp(5′G) (cap analog), 2 mM DTT, 10 U/μL ribonuclease inhibitor, 7.5 U/μL T7 RNA polymerase (Promega), and 1× transcription buffer (Promega). Reactions were incubated at 37°C for 2 hr, gel purified using denaturing PAGE, eluted from the gel (elution buffer: 0.5 mM NaOAc [pH 5.6], 0.1% SDS, 10 mM Tris [pH 7.5], and 1 mM EDTA), ethanol precipitated, and resuspended in nuclease-free water.

In Vitro Splicing

Splicing reactions were performed using α-^{32}P-labeled β-globin RNA in 30% HeLa nuclear extract with 500 nM SRPK1 or kdSRPK1 in reaction buffer as previously described (Movassat et al., 2014) with final 70 mM NaCl and incubated at 30°C for 90 min. Reactions were digested using proteinase K, extracted with phenol chloroform, precipitated with ethanol, and separated on denaturing PAGE. The gel was exposed to BaFBr:Eu screen and imaged on PhosphorImager (Bio-Rad). Appearance of final spliced product was determined by taking the percent of the sum of the adjusted intensity for the spliced band divided by the total signal for the spliced and unspliced product band (Movassat et al., 2014).

RNA Binding Assays

Pull-down experiments were performed with a ^{32}P-labeled RNA oligomer based on the Ron ESE (5′-AGGCGGAGGAAGC-3′). Labeling was carried out using the KinaseMax kit from Ambion and confirmed by 12% urea PAGE. GST-tagged SRSF1 proteins (4 μM) were incubated with 3 pmol ^{32}P-labeled Ron ESE at 23°C for 30 min in 20 μL binding buffer and mixed with 25 μL glutathione-agarose resin for 30 min at room temperature. Where phosphorylation was required, GST- and His-tagged enzymes (4 μM) were incubated with and without 100 μM ATP and 10 mM Mg^{2+} in 20 mM Tris/HCl (pH 7.5) and 75 mM NaCl at 37°C for 30 min prior to the RNA addition step. In some cases, 4 μM kdSRPK or buffer blank was added and incubated at room temperature for 10 min prior to the RNA addition step. The resin was washed three times with 200 μL binding buffer, and the retained RNA was counted in liquid scintillant.

AUTHOR CONTRIBUTIONS

B.E.A., G.W., M.M.K., L.F., and M.M. performed all experiments. K.J.H. and X.-D.F. helped with data analysis and manuscript editing. J.A.A. planned the experiments and wrote the paper.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health (NIH) grants GM67969 (to J.A.A.), GM52872 (to X.-D.F.), and GM62287 (to K.J.H.). We would like to thank Dr. Gourisankar Ghosh for providing U1(RRM).

Received: October 9, 2015
Revised: April 26, 2016
Accepted: May 25, 2016
Published: July 7, 2016

REFERENCES

Aubol, B.E., and Adams, J.A. (2011). Applying the brakes to multisite SR protein phosphorylation: substrate-induced effects on the splicing kinase SRPK1. Biochemistry *50*, 6888–6900.

Aubol, B.E., Plocinik, R.M., McGlone, M.L., and Adams, J.A. (2012). Nucleotide release sequences in the protein kinase SRPK1 accelerate substrate phosphorylation. Biochemistry *51*, 6584–6594.

Aubol, B.E., Plocinik, R.M., Hagopian, J.C., Ma, C.T., McGlone, M.L., Bandyopadhyay, R., Fu, X.D., and Adams, J.A. (2013). Partitioning RS domain phosphorylation in an SR protein through the CLK and SRPK protein kinases. J. Mol. Biol. *425*, 2894–2909.

Aubol, B.E., Plocinik, R.M., Keshwani, M.M., McGlone, M.L., Hagopian, J.C., Ghosh, G., Fu, X.D., and Adams, J.A. (2014). N-terminus of the protein kinase CLK1 induces SR protein hyperphosphorylation. Biochem. J. *462*, 143–152.

Cho, S., Hoang, A., Sinha, R., Zhong, X.Y., Fu, X.D., Krainer, A.R., and Ghosh, G. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. Proc. Natl. Acad. Sci. USA *108*, 8233–8238.

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C., and Duncan, P.I. (1996). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. EMBO J. *15*, 265–275.

Ding, J.H., Zhong, X.Y., Hagopian, J.C., Cruz, M.M., Ghosh, G., Feramisco, J., Adams, J.A., and Fu, X.D. (2006). Regulated cellular partitioning of SR protein-specific kinases in mammalian cells. Mol. Biol. Cell *17*, 876–885.

Endicott, J.A., Noble, M.E., and Johnson, L.N. (2012). The structural basis for control of eukaryotic protein kinases. Annu. Rev. Biochem. *81*, 587–613.

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

Ghosh, G., and Adams, J.A. (2011). Phosphorylation mechanism and structure of serine-arginine protein kinases. FEBS J. *278*, 587–597.

Gui, J.F., Lane, W.S., and Fu, X.D. (1994). A serine kinase regulates intracellular localization of splicing factors in the cell cycle. Nature *369*, 678–682.

Kataoka, N., Bachorik, J.L., and Dreyfuss, G. (1999). Transportin-SR, a nuclear import receptor for SR proteins. J. Cell Biol. *145*, 1145–1152.

Keshwani, M.M., Aubol, B.E., Fattet, L., Ma, C.T., Qiu, J., Jennings, P.A., Fu, X.D., and Adams, J.A. (2015a). Conserved proline-directed phosphorylation regulates SR protein conformation and splicing function. Biochem. J. *466*, 311–322.

Keshwani, M.M., Hailey, K.L., Aubol, B.E., Fattet, L., McGlone, M.L., Jennings, P.A., and Adams, J.A. (2015b). Nuclear protein kinase CLK1 uses a nontraditional docking mechanism to select physiological substrates. Biochem. J. *472*, 329–338.

Koizumi, J., Okamoto, Y., Onogi, H., Mayeda, A., Krainer, A.R., and Hagiwara, M. (1999). The subcellular localization of SF2/ASF is regulated by direct interaction with SR protein kinases (SRPKs). J. Biol. Chem. *274*, 11125–11131.

Lai, M.C., Lin, R.I., and Tarn, W.Y. (2001). Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. Proc. Natl. Acad. Sci. USA *98*, 10154–10159.

Lavoie, H., Li, J.J., Thevakumaran, N., Therrien, M., and Sicheri, F. (2014). Dimerization-induced allostery in protein kinase regulation. Trends Biochem. Sci. *39*, 475–486.

Lemmon, M.A., and Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. Cell *141*, 1117–1134.

Ma, C.T., Ghosh, G., Fu, X.D., and Adams, J.A. (2010). Mechanism of dephosphorylation of the SR protein ASF/SF2 by protein phosphatase 1. J. Mol. Biol. *403*, 386–404.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499–506.

Mathew, R., Hartmuth, K., Möhlmann, S., Urlaub, H., Ficner, R., and Lührmann, R. (2008). Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. Nat. Struct. Mol. Biol. *15*, 435–443.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688–700.

Movassat, M., Mueller, W.F., and Hertel, K.J. (2014). In vitro assay of pre-mRNA splicing in mammalian nuclear extract. Methods Mol. Biol. *1126*, 151–160.

Nayler, O., Stamm, S., and Ullrich, A. (1997). Characterization and comparison of four serine- and arginine-rich (SR) protein kinases. Biochem. J. *326*, 693–700.

Ngo, J.C., Giang, K., Chakrabarti, S., Ma, C.T., Huynh, N., Hagopian, J.C., Dorrestein, P.C., Fu, X.D., Adams, J.A., and Ghosh, G. (2008). A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1. Mol. Cell *29*, 563–576.

Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature *463*, 457–463.

Pandit, S., Wang, D., and Fu, X.D. (2008). Functional integration of transcriptional and RNA processing machineries. Curr. Opin. Cell Biol. *20*, 260–265.

Parsons, S.J., and Parsons, J.T. (2004). Src family kinases, key regulators of signal transduction. Oncogene *23*, 7906–7909.

Poulikakos, P.I., Zhang, C., Bollag, G., Shokat, K.M., and Rosen, N. (2010). RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. Nature *464*, 427–430.

Prasad, J., Colwill, K., Pawson, T., and Manley, J.L. (1999). The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. Mol. Cell. Biol. *19*, 6991–7000.

Rossman, K.L., Der, C.J., and Sondek, J. (2005). GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. Nat. Rev. Mol. Cell Biol. *6*, 167–180.

Shin, C., and Manley, J.L. (2004). Cell signalling and the control of pre-mRNA splicing. Nat. Rev. Mol. Cell Biol. *5*, 727–738.

Stamm, S. (2008). Regulation of alternative splicing by reversible protein phosphorylation. J. Biol. Chem. *283*, 1223–1227.

Taira, K., and Benkovic, S.J. (1988). Evaluation of the importance of hydrophobic interactions in drug binding to dihydrofolate reductase. J. Med. Chem. *31*, 129–137.

Taylor, S.S., and Kornev, A.P. (2011). Protein kinases: evolution of dynamic regulatory proteins. Trends Biochem. Sci. *36*, 65–77.

Taylor, S.S., Ilouz, R., Zhang, P., and Kornev, A.P. (2012). Assembly of allosteric macromolecular switches: lessons from PKA. Nat. Rev. Mol. Cell Biol. *13*, 646–658.

Wang, G.S., and Cooper, T.A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. Nat. Rev. Genet. *8*, 749–761.

Wang, H.Y., Lin, W., Dyck, J.A., Yeakley, J.M., Songyang, Z., Cantley, L.C., and Fu, X.D. (1998). SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. J. Cell Biol. *140*, 737–750.

Xiao, S.H., and Manley, J.L. (1998). Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. EMBO J. *17*, 6359–6367.

Yun, C.Y., Velazquez-Dones, A.L., Lyman, S.K., and Fu, X.D. (2003). Phosphorylation-dependent and -independent nuclear import of RS domain-containing splicing factors and regulators. J. Biol. Chem. *278*, 18050–18055.

Zhong, X.Y., Ding, J.H., Adams, J.A., Ghosh, G., and Fu, X.D. (2009). Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones. Genes Dev. *23*, 482–495.

Zhou, Z., Qiu, J., Liu, W., Zhou, Y., Plocinik, R.M., Li, H., Hu, Q., Ghosh, G., Adams, J.A., Rosenfeld, M.G., and Fu, X.D. (2012). The Akt-SRPK-SR axis constitutes a major pathway in transducing EGF signaling to regulate alternative splicing in the nucleus. Mol. Cell *47*, 422–433.

APPENDIX C

Preparation of Splicing Competent Nuclear Extract from Mammalian Cells and In Vitro Pre-mRNA Splicing Assay

Chapter 2

Preparation of Splicing Competent Nuclear Extract from Mammalian Cells and In Vitro Pre-mRNA Splicing Assay

Maliheh Movassat, Hossein Shenasa, and Klemens J. Hertel

Abstract

The ability to perform in vitro splicing assays has paved the way for in-depth studies of the mechanisms and machinery involved in the process of splicing. The in vitro splicing assay is a valuable experimental approach that combines the complexity of the spliceosome and regulatory systems with the flexibility of performing endless splicing and alternative splicing reactions. Through the use of crude nuclear extract and radiolabeled pre-mRNA, spliced mRNAs can be visualized using autoradiography for downstream analysis. This chapter describes the necessary steps to perform an in vitro splicing reaction, including the generation of the key components necessary for the splicing reaction; nuclear extract.

Key words Nuclear extract, HeLa cells, In vitro splicing, Splicing, Alternative splicing, mRNA processing, Pre-mRNA substrate, In vitro transcription, RNA extraction and purification

1 Introduction

RNA splicing is an essential co-transcriptional feature of gene expression in eukaryotes [1]. The spliceosome assembles in a stepwise manner on pre-mRNA transcripts as RNA polymerase elongates the nascent chain. RNA splicing entails the simultaneous excision and ligation of gene coding exons and the eventual degradation of introns [2]. Alternative splicing involves the selective excision and ligation of gene coding exons and is responsible for much of the genetic diversity seen in higher order organisms [2, 3].

In the field of RNA biology, cell-free or in vitro-based assays [4] have played a key role in providing mechanistic insights into the molecular workings of the cell. One such assay has been the in vitro splicing assay [5, 6], which allows for an experimenter controlled method for the investigation of splicing mechanisms, including

Maliheh Movassat and Hossein Shenasa contributed equally to this work.

Yongsheng Shi (ed.), *mRNA Processing: Methods and Protocols*, Methods in Molecular Biology, vol. 1648,
DOI 10.1007/978-1-4939-7204-3_2, © Springer Science+Business Media LLC 2017

spliceosomal assembly, splicing kinetics, and splicing regulatory processes. In vitro splicing reactions require two critical components: minigene constructs and nuclear extract. Minigene constructs are small fragments of DNA that, at minimum contain two exons, an intron and a phage promoter sequence upstream of the coding strand [7]. Minigene constructs can be transcribed from DNA to RNA using commercial phage polymerases in an in vitro transcription reaction. Nuclear extract preparations were originally developed to study RNA polymerase II transcription in the test tube [8]; however, these extracts were also shown to support intron excision [9]. Over the years many forms of splicing competent nuclear extract have been developed [10–15], all of which contain the essential components required for an efficient splicing reaction of pre-synthesized pre-mRNAs. The extract contains: spliceosomal components, splicing regulatory proteins, ATP, and other components of the nucleus [14, 16]. HeLa cells are the most common cells from which nuclear extract is prepared; however, nuclear extract from other cells can also be prepared in the same manner. In vitro splicing reactions generally use radiolabeled transcription of minigene constructs and subsequent incubation with nuclear extract [6, 16]. Proteins are digested using Proteinase K and the RNA is then separated from the proteins using phenol/chloroform extraction. After RNA precipitation, the spliced RNAs can be electrophoretically separated on a denaturing polyacrylamide gel and analyzed by autoradiography.

In vitro splicing reactions can be used to study many aspects of the splicing reaction such as splice site strength of the exon/intron junctions, influence of splicing regulatory elements, or the molecular interactions during splice site pairing. The ease with which the experimenter can manipulate the spliceosomal assembly pathway renders this assay an invaluable tool.

2 Materials and Reagents

All reagents and materials should be of high quality, RNase free, and molecular biology grade. In vitro splicing steps usually require work with radioactive isotopes; therefore, all necessary precautions must be taken. Carefully follow all local hazardous and radioactive waste disposal regulations when carrying out experiments using radiolabeled RNA.

2.1 Nuclear Extract Components and Reagents

2.1.1 Cells

Spinner cultured suspension HeLa-S3 cells (*see* **Note 1**).

2.1.2 Reagents

1. 1 M dithiothreitol (DTT).
2. 100 mM phenylmethanesulfonyl fluoride (PMSF) in isopropanol (*see* **Note 2**).

3. 1 M 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) buffer, pH 7.9 at 4 °C with KOH. Store at 4 °C (*see* **Note 3**).
4. 1 M magnesium acetate ($Mg(OAc)_2$) (*see* **Note 4**).
5. 2.5 mM potassium acetate (KOAc) (*see* **Note 4**).
6. 0.5 mM EDTA.
7. 10× Phosphate-buffered saline (PBS), pH 7.4. Store at 4 °C.
8. Glycerol (at least 2 L needed for a 30 L culture of cells).
9. Autoclaved or double-deionized water.
10. Trypan blue.

2.1.3 Buffer Recipes

All reagents should be prepared with autoclaved or double-deionized water, followed by filter sterilization with a 0.22 μm filter or sterilized by autoclaving. All reagents should be stored at 4 °C and be cold prior to use. DTT and PMSF should be made fresh and only added to each corresponding buffer prior to use, not ahead of time.

1. Hypotonic buffer: 10 mM HEPES-KOH at pH 7.9, 1.5 mM $Mg(OAc)_2$, 10 mM KOAc, 0.5 mM DTT, 0.2 mM PMSF.
2. Low-salt buffer: 20 mM HEPES-KOH at pH 7.9, 1.5 mM $Mg(OAc)_2$, 20 mM KOAc, 0.2 mM EDTA, 25% glycerol, 0.5 mM DTT, 0.2 mM PMSF.
3. High-salt buffer: 20 mM HEPES-KOH at pH 7.9, 1.5 mM $Mg(OAc)_2$, 1.4 M KOAc, 0.2 mM EDTA, 25% glycerol, 0.5 mM DTT, 0.2 mM PMSF.
4. Dialysis buffer: 20 mM HEPES-KOH at pH 7.9, 100 mM KOAc, 0.2 mM EDTA, 20% glycerol, 0.5 mM DTT, 0.2 mM PMSF.

2.1.4 Equipment

All glassware (including dounce homogenizer) and bottles/tubes should be sterilized and autoclaved ahead of time. Make sure all glassware and plasticware that is used does not contain detergent residue.

1. Dialysis tubing: 10,000 molecular weight cutoff (MWCO).
2. Glass dounce homogenizer with a tight clearance pestle (*see* **Note 5**).
3. Centrifuge with swinging bucket rotor capable of speeds up to 3500 × *g*.
4. Centrifuge bottles: polypropylene, conical bottom with graduations, wide mouth with sealing caps.
5. Ultracentrifuge with fixed angle rotor capable of speeds up to 25,000 × *g*.
6. Centrifuge tubes: polycarbonate with polypropylene screw caps, 30 mL volume.

7. 25 mL serological glass pipettes.
8. 1 L glass bottles.
9. 200 mL glass beaker.
10. 4 L beaker/buckets for dialysis.
11. 1.5 mL tubes.
12. Column to aid in drip addition of high-salt buffer (optional).
13. Glass slide(s).
14. Phase-contrast microscope.
15. Magnetic stir bar.
16. Magnetic stir plate.
17. Ice.
18. Dry ice.
19. −80 °C Freezer.

2.2 In Vitro Splicing Reaction

2.2.1 Splicing Reaction Components

1. Radiolabeled pre-mRNA generated from an in vitro transcription reaction (*see* **Note 6**).
2. Splicing competent nuclear extract (*see* Subheading 3.1).
3. 25 mM adenosine triphosphate (ATP).
4. 0.5 M creatine phosphate (CP).
5. 80 mM $Mg(OAc)_2$ (*see* **Note 4**).
6. RNase inhibitor, 40 units/μL.
7. 100 mM DTT.
8. 13% polyvinyl alcohol (PVA) (optional).
9. 1 M KOAc (*see* **Note 4**).
10. 0.5 M HEPES-KOH, pH 7.9.
11. Wet ice and dry ice.
12. Water bath.

2.2.2 Splicing gel Components

1. Upright vertical gel electrophoresis system.
2. Electrophoresis power supply with temperature probe.
3. 8″ × 8″ glass plates or equivalent.
4. 0.4 mm spacer set.
5. 0.4 mm gel comb.
6. Aluminum plate: 8″ × 8″ or longer and precooled.
7. 1¼″ binder clips (at least four).
8. 10× Tris-Borate-EDTA (TBE): 1 M Tris Base, 1 M boric acid, 20 mM EDTA.
9. 7 M Urea prepared in 1× TBE.
10. 20% (19:1) acrylamide:bis-acrylamide solution: 210.2 g solid urea, 50 mL 10× TBE, 250 mL 40% acrylamide:bis-acrylamide

solution, adjust to 500 mL with sterile water to yield a final concentration of 20% acrylamide/7 M urea/1× TBE.

11. *N,N,N′,N′*-Tetramethylethylenediamine (TEMED).
12. 10% ammonium persulfate (APS).
13. Stop dye: 98% formamide, 0.1% bromophenol blue, 0.1% xylene cyanol, 10 mM EDTA at pH 8.
14. Silicone-based coating solution.
15. 70% ethanol.
16. Putty knife/gel spatula.
17. Filter paper cut in 8″ × 8″ squares.
18. Plastic wrap.
19. Gel dryer.
20. PhosphorImager System. Film may also be used.
21. Lint-free tissue.
22. 30 mL syringe and needles.

2.2.3 RNA Purification

1. Proteinase K at 10 mg/mL.
2. 2× Proteinase K buffer: 20 mM Tris Base, 2% SDS, 200 mM NaCl, 2 mM EDTA, pH 7.5.
3. 100% ethanol.
4. Glycogen.
5. Phenol, chloroform, isoamyl alcohol solution (25:24:1 pH 8.0).

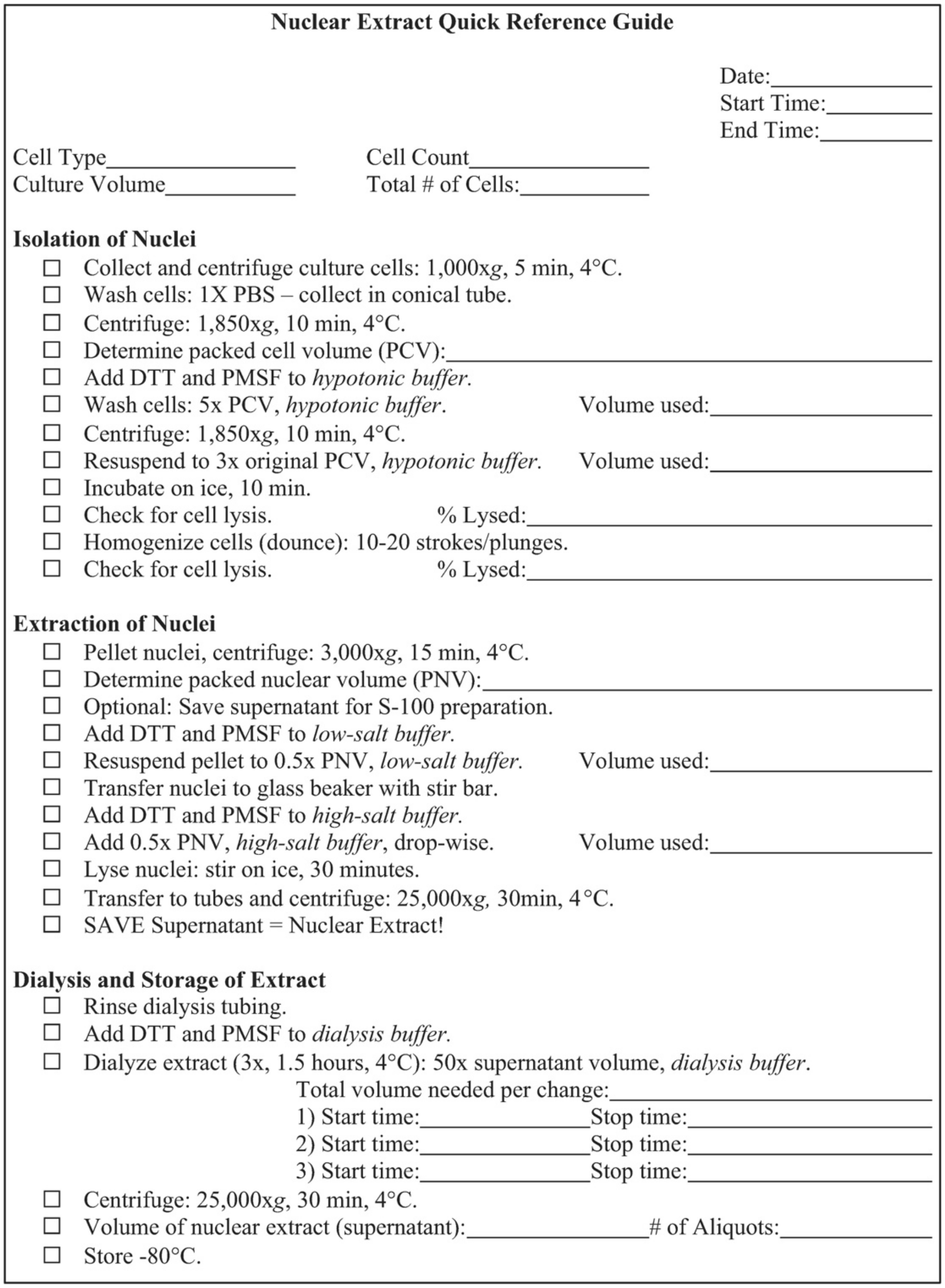
3 Methods

3.1 Nuclear Extract Preparation

To prevent the denaturation of RNA and proteins, all the extraction steps should be carried out on ice (or in a coldroom). All reagents and buffers should be equilibrated to 4 °C and centrifuge rotors should be precooled to 4 °C. Once initiated, this protocol should be carried all the way to completion. *See* Fig. 1 for a quick reference guide to nuclear extract preparation.

3.1.1 Isolation of Nuclei

1. Transfer HeLa cells to a conical centrifuge bottle to pellet the cells. Centrifuge at 1000 × *g* for 5 min at 4 °C. Decant the supernatant carefully so as not to disturb the pellet.
2. Wash the cells by resuspending the pellet with ice-cold 1× PBS. Add PBS at 5× the volume of cells (*see* **Note 7**).
3. Pellet cells post-wash by centrifugation at 1850 × *g* for 10 min at 4 °C. Remove the supernatant carefully. From these pelleted cells, determine the packed cell volume (PCV), using the graduations on the centrifuge bottle.
4. Wash the cells GENTLY by resuspending the pellet with 5× PCV hypotonic buffer. Immediately centrifuge cells to pellet



Nuclear Extract Quick Reference Guide

Date:_____________
Start Time:_________
End Time:_________

Cell Type________________ Cell Count_______________
Culture Volume___________ Total # of Cells:___________

Isolation of Nuclei

- □ Collect and centrifuge culture cells: 1,000x*g*, 5 min, 4°C.
- □ Wash cells: 1X PBS – collect in conical tube.
- □ Centrifuge: 1,850x*g*, 10 min, 4°C.
- □ Determine packed cell volume (PCV):_______________________________________
- □ Add DTT and PMSF to *hypotonic buffer.*
- □ Wash cells: 5x PCV, *hypotonic buffer*. Volume used:__________________
- □ Centrifuge: 1,850x*g*, 10 min, 4°C.
- □ Resuspend to 3x original PCV, *hypotonic buffer.* Volume used:__________________
- □ Incubate on ice, 10 min.
- □ Check for cell lysis. % Lysed:_________________________________
- □ Homogenize cells (dounce): 10-20 strokes/plunges.
- □ Check for cell lysis. % Lysed:_________________________________

Extraction of Nuclei

- □ Pellet nuclei, centrifuge: 3,000x*g*, 15 min, 4°C.
- □ Determine packed nuclear volume (PNV):____________________________________
- □ Optional: Save supernatant for S-100 preparation.
- □ Add DTT and PMSF to *low-salt buffer.*
- □ Resuspend pellet to 0.5x PNV, *low-salt buffer.* Volume used:__________________
- □ Transfer nuclei to glass beaker with stir bar.
- □ Add DTT and PMSF to *high-salt buffer.*
- □ Add 0.5x PNV, *high-salt buffer*, drop-wise. Volume used:__________________
- □ Lyse nuclei: stir on ice, 30 minutes.
- □ Transfer to tubes and centrifuge: 25,000x*g,* 30min, 4 °C.
- □ SAVE Supernatant = Nuclear Extract!

Dialysis and Storage of Extract

- □ Rinse dialysis tubing.
- □ Add DTT and PMSF to *dialysis buffer.*
- □ Dialyze extract (3x, 1.5 hours, 4°C): 50x supernatant volume, *dialysis buffer*.

  Total volume needed per change:_______________________
  1) Start time:______________Stop time:____________________
  2) Start time:______________Stop time:____________________
  3) Start time:______________Stop time:____________________
- □ Centrifuge: 25,000x*g*, 30 min, 4°C.
- □ Volume of nuclear extract (supernatant):_______________# of Aliquots:____________
- □ Store -80°C.

Fig. 1 Nuclear extract preparation guide for quick referencing and data recording

at 1850 × *g* for 10 min at 4 °C and decant the supernatant (*see* **Note 8**).

5. Resuspend the packed cells in hypotonic buffer to a final volume of 3× the original PCV (the volume of the cells and the hypotonic buffer combined should be 3× PCV). Incubate cells on ice and allow the cells to swell for 10 min (*see* **Note 9**).
   (a) Check for cell lysis of pre-dounced cells. Lysis can be determined by visualizing stained cells under the microscope by the addition of trypan blue. For a dense concentration of cells, dilute with 1× PBS (*see* **Note 10**).
6. Transfer the cells to a dounce homogenizer to aid in cell lysis. Homogenize the cells with 10–20 plunges/strokes using an up and down motion (*see* **Note 11**).
   (a) Monitor cell lysis by checking post-dounced cells as previously described in **step 5a** (*see* **Note 12**).

3.1.2 Extraction of Nuclei

1. Transfer dounced cells to clean centrifuge bottles and spin to pellet by centrifugation at 3300 × *g* for 15 min at 4 °C. Determine and record the packed nuclear volume (PNV), using the graduations on the centrifuge bottle. At this point the supernatant can be saved for cytoplasmic S-100 extract preparation (*see* **Note 1**3).
2. Resuspend the pellet (which now contains nuclei) by adding 0.5× PNV low-salt buffer. Transfer the resuspension to a glass beaker with a magnetic stir bar and gently stir on ice (*see* **Note 14**).
3. Gently release the soluble proteins from the nuclei by adding 0.5× PNV high-salt buffer in a dropwise fashion. Continue to stir on ice for 30 min to complete the extraction of the nuclei (*see* **Note 15**).
4. Transfer nuclei to centrifuge tubes to pellet by centrifugation at 25,000 × *g* for 30 min. Save the supernatant, as this is the nuclear extract.

3.1.3 Dialysis and Storage of Extract

1. Prepare the dialysis tubing by rinsing in distilled water while samples are in the centrifuge (*see* **Note 16**).
2. Desalt the nuclear extract by dialyzing the supernatant in the dialysis tubing dialysis buffer at 50× supernatant volume. Dialyze for 1.5 h at 4 °C, while stirring (*see* **Note 17**).
3. Change the dialysis buffer two additional times and dialyze for 1.5 h at 4 °C each time.
4. Transfer the nuclear extract to centrifuge tubes and remove the precipitate by centrifugation at 25,000 × *g* for 30 min at 4 °C. Save the supernatant.

5. Aliquot the supernatant into 1 mL fractions and freeze on dry ice immediately (*see* **Note 18**).
6. Store the nuclear extract aliquots at −80 °C (*see* **Note 19**).
7. Validate the activity of the nuclear extract by performing an in vitro splicing reaction.

3.2 In Vitro Splicing Assay

3.2.1 Splicing Reaction

1. Determine the volume for a master mix. The master mix volume can be determined with the following equation:
   Reaction volume × (# of reactions + 1) = master mix volume
   (a) For example, 25 μL × (4 reactions + 1) = 125 μL. The extra (+1) volume is to account for pipetting errors. Typical splicing reaction volumes range between 10 and 25 μL.
2. Thaw nuclear extract on ice.
3. Allow the following reagents to thaw at room temperature and then immediately place them on ice: ATP, CP, $Mg(OAc)_2$, DTT, HEPES, and KOAc. Thaw radiolabeled RNA and nuclear extract on ice. Keep RNase inhibitor on ice as well (*see* **Note 4**).
4. Combine the following reagents to obtain a final concentration of: 1 mM ATP, 20 mM CP, 3.2 mM $Mg(OAc)_2$, 10 units RNase inhibitor, 1 mM DTT, 3% PVA (optional), 12 mM HEPES, 72.5 mM KOAc, and 10–50% nuclear extract (percent of nuclear extract to use should be optimized for each extract and substrate that will be used). Add sterile water to bring up the master mix volume if necessary (*see* **Notes 20**–**22**).
5. For each experimental condition, add the following: appropriate volume of master mix, 0.01–0.1 nM RNA (~1000 cpm), experimental variant (such as protein), and/or sterile water to bring up the volume. Add nuclear extract last to initiate the reaction. Gently pipette up and down to mix the reaction. Keep all reaction tubes on ice. Once nuclear extract has been added, quickly take a time 0 ($T = 0$) aliquot from each reaction tube and immediately freeze on dry ice (*see* **Note 23**).
6. Incubate all reactions (except the time 0 control) at 30 °C for 90 min in a water bath (*see* **Note 24**).
7. While the splicing reaction is incubating, prepare the 6% polyacrylamide gel.

3.2.2 Splicing Gel Preparation

1. Prepare a 6% polyacrylamide solution in a 50 mL conical tube. For the gel size specified above, 25 mL is sufficient. Dilute 20% acrylamide/7 M urea/1× TBE with 7 M urea/1× TBE to obtain a 6% acrylamide solution.
2. Carefully clean the glass plates by running them under deionized water (*see* **Note 25**).

3. Place both glass plates on a flat surface with the inside of each plate facing upward. Wipe away any residual water with lint-free paper towels.
4. Spray the glass plates with 70% ethanol and wipe them dry with lint-free paper towels (*see* **Note 26**).
5. Assemble the gel cassette (*see* **Note 27**) and add binder clips to the edges and sides of the glass plates to hold the plates together, taking care to not move the spacers (*see* **Note 28)**.
6. Once the gel cassette is ready, add a 1:1000 volume of TEMED and a 1:100 volume of 10% APS to the 6% acrylamide solution (*see* **Note 29)**.
7. Aspirate the polyacrylamide solution with a 30 mL syringe that does not have a needle. Hold the gel at a 45° angle and place the syringe tip such that it makes firm and direct contact with the non-siliconized plate. Apply constant pressure to dispense the acrylamide solution. Once the cassette is full, place it on a flat elevated surface such as a test tube rack (*see* **Note 30)**.
8. Insert the gel comb (*see* **Note 31**).
9. Leave the cassette flat on a bench top. Flush out any remaining un-polymerized gel from the syringe back into the conical tube. This can be used as a marker to confirm that the gel has polymerized. Let the gel polymerize for ~30 min.
10. Remove the gel comb and bottom spacer. Clamp the gel cassette to the upright gel electrophoresis apparatus using 1¼″ binder clips (*see* **Note 32**).
11. Tilt the electrophoresis apparatus to one side, making a 45° angle between the bench top and the bottom of the electrophoresis apparatus, slowly pour 1× TBE buffer down the raised end while incrementally lowering the apparatus. This will minimize the formation of bubbles between the glass plates (*see* **Note 33**).
12. Make sure the bottom of the gel is submerged in buffer. Fill the top compartment of the electrophoresis apparatus with 1× TBE, until the wells in the gel are filled with buffer.
13. Prerun the gel for 15 min at 30 W. Attach a temperature probe and set the temperature limit to 45 °C to ensure the glass plates do not break.

3.2.3 Proteinase K Digest

1. Once the splicing reaction has reached completion, immediately place the tubes on dry ice.
2. Determine the final volume for the Proteinase K master mix using the following equation:

   Reaction volume × (# of reactions + 1) = master mix volume
3. Mix reagents to yield a final master mix volume: 1× Proteinase K buffer at final desired volume, 0.25 mg/mL glycogen (*see* **Note 34**) and 0.25 mg/mL Proteinase K. Use sterile water to

adjust the volume of the master mix (if necessary) to a final volume of 180 μL per reaction.

4. Add 175 μL of Proteinase K master mix to each tube and incubate at 37 °C for 10–15 min (*see* **Note 35**).

3.2.4 RNA Purification

1. Once the Proteinase K digest is complete, add 200 μL of phenol:chloroform to each reaction tube.
2. Vortex the tubes on high speed for 30s and centrifuge at 16,500 × *g* for 5 min.
3. Carefully remove the aqueous phase (top layer), taking care not to remove any of the organic phase.
4. Pipette the aqueous phase of each tube into a fresh tube and add 3 volumes of 100% ice-cold ethanol. Incubate the tubes at −20 °C for 15 min.
5. Centrifuge the tubes at 16,500 × *g* for 10 min at room temperature (*see* **Note 36**).
6. Remove the ethanol supernatant, taking care not to disturb the pellet. Allow the pellet to air dry for no more than a few minutes. Add a small volume of stop dye (10 μL or less) and pipette up and down to mix. At this point the RNA is ready to be loaded onto the gel.

3.2.5 Electrophoresis and Visualization

1. Load samples onto the gel and run at 30 W (100 V), 45 °C, for 60–90 min or until the bromophenol blue dye reaches the bottom of the gel. The length of time samples that are run on the gel should be optimized based on the size of expected products (*see* **Notes 37** and **38**).
2. Remove the buffer from the top and bottom compartments of the upright electrophoresis apparatus as well as the temperature probe. Carefully detach the 1¼″ binder clips.
3. Cut filter paper into 8″ × 8″ pieces.
4. Pull the vertical spacers out of the gel cassette and use the putty knife to wedge open the glass plates. Remove the siliconized (notched) plate.
5. Place a piece of 8″ × 8″ filter paper on the gel and gently press down. Invert the gel such that the filter paper is on the bench and the glass plate is on top.
6. Use the putty knife to remove the remaining glass plate by slowly lifting one corner. The gel should adhere to the filter paper.
7. Place a piece of plastic wrap on the top of the gel. Take care to avoid creases in the wrap.
8. Dry the gel for 20 min at 80 °C using a gel dryer. Make sure the suction pump is turned on for the duration of drying.
9. Expose a phosphorscreen to the gel for 1 h-overnight (*see* **Notes 39** and **40**).

10. Use a PhosphorImager to obtain an image of the spliced products. This image can be used to quantify parameters such as percentage of splicing or splice site preference in a minigene that contains competitive splice sites.
11. Quantify results using a gel analysis software (*see* **Note 41**).

4 Notes

1. This nuclear extract protocol starts with a large volume (~30 L) of purchased spinner cultured HeLa cells. HeLa cells can also be cultured in the lab.
2. PMSF is dissolved in anhydrous isopropanol and should be prepared fresh and added to corresponding buffers just prior to use (similar to addition of DTT). Store the freshly prepared PMSF solution on ice or at 4 °C during duration of the nuclear extract prep.
3. HEPES solution should be brought to pH 7.9 while at 4 °C or on ice.
4. $MgCl_2$ and KCl can be used as well; however, a previous study has shown that the use of acetate as a counter ion enhances the splicing reaction [17].
5. Use a pestle that has a tight fit in the mortar (approximately 0.025–0.076 mm): Kontes brand homogenizers have a tight clearence in their type-B pestle, alternatively, Wheaton dounce homogenizers have a tighter clearance in their type-A pestle.
6. Pre-mRNAs are generally transcribed with commercially available phage polymerases in the presence of UTP that contains phosphorous-32 at the alpha position (^{32}P α-UTP). The radioactive nucleotide is generally in the 0.3–3 nM range and usually leads to 100,000 cpm/μL incorporation. Other nucleotide triphosphates that have a radioactive α-phosphate can be used instead of ^{32}P α-UTP if needed.
7. It is suggested to use 25 mL serological glass pipettes to wash the cells. Make sure to mix gently and not expel volume from the pipette completely while mixing, so as to prevent the creation of a vacuum.
8. This step should be performed quickly. The hypotonic buffer swells the cells and could potentially cause the cells to leak or burst, leading to loss of protein into the supernatant or cell death. Monitoring cell lysis with trypan blue allows for visualization of lysed cells, which take up the dye, as compared to intact cells, which do not.
9. The previous wash step with hypotonic buffer may have initiated swelling of the cells. Therefore, the PCV may have increased. When determining how much hypotonic buffer to

add in this step, refer only to the initial PCV that was recorded. For example, the PCV determined in **step 4** is 15 mL, yet after **step 5** it has increased to 25 mL. In **step 6** add hypotonic buffer such that the final volume of cells and buffer is 45 mL (3 × 15 mL). Add the hypotonic solution gently while mixing with a serological pipette.

10. Check for the lysis of pre-dounced cells by adding trypan blue, 1:2 dilution. At this point, you want a minimum of 50% of the cells alive. Greater than 80% cell survival is ideal. Use a plastic pipette with a cut tip to check for lysis to prevent shearing of the cells.
11. Perform the dounce homogenization step with gentle strokes maintaining a constant plunging motion. Do not remove the pestle from the dounce until douncing is complete. This will ensure efficient cell breakage.
12. Visualize cell lysis with a 1:2 dilution of cells to trypan blue. 80–90% cell lysis should be expected. A good sampling of the lysed cells can be found from further down the dounce tube, and not from the top.
13. S-100 cytoplasmic fraction extraction can be performed at this step. For further instructions, *see* ref. 11.
14. The lysate can be homogenized again by douncing if the cell solution is chunky. At this point, if there are multiple tubes, combine the nuclei into one beaker.
15. The dropwise addition of the high-salt buffer is vitally important because rapidly increasing the salt concentration may lead to nuclear lysis and precipitation of nuclear components. High-salt buffer permeabilizes the nuclear membrane to allow for release of necessary components. Nuclei can be further homogenized with 5 dounce strokes to prevent clumping when nuclei are in large volumes.
16. Alternatively, dialysis tubing can be rinsed in dialysis buffer.
17. Avoid the presence of bubbles in the dialysis tubing by gently squeezing the bubbles out of the tubing with your fingers. Remember to clamp one end of the tubing prior to addition of sample.
18. 30 L of a HeLa cell culture with a 4–6 × 10^5 cells/mL density should yield about 45 mL of nuclear extract.
19. Freeze/thaw cycles of the nuclear extract aliquots should be limited to avoid compromising extract activity. The non-thawed extracts can be stored up to 2 years at −80 °C without loss of activity. However, the half-life at 4 °C is only 12 h [18].
20. The optimum potassium concentration for an in vitro splicing reaction is around 30 mM; however, splicing can be observed between 2 mM up to 100 mM. Splicing efficiency is reduced

drastically at the high and low extremes. Splicing efficiency may be increased by optimizing the potassium ion concentration [9].

21. When adding KOAc and HEPES it is important to account for the potassium ion and HEPES concentration already present in the nuclear extract. For example, if the nuclear extract already contains 100 mM KOAc and 20 mM HEPES it is possible to calculate how much additional KOAc and HEPES to add to the reaction with the following calculation:

$$\begin{gathered} \textit{For a } 25\,\mu L \textit{ reaction containing } 30\%\ NE: \\ (7.5\ \mu L\, NE)(100\ mM\ KOAc) = (25\ \mu L)(\mathrm{X}) \\ \mathrm{X} = 30\, mM\ KOAc. \end{gathered}$$

The addition of nuclear extract to the reaction yields an initial 30 mM KOAc concentration. This amount must be subtracted from the desired final concentration

$$\begin{gathered} 72.5\ mM\ KOAc - 30\ mM\ KOAc = 42.5\ mM\ KOAc \\ (1000\ mM\ KOAc)(\mathrm{X}) = (25\,\mu L)(42.5\ mM\ KOAc) \\ \mathrm{X} = 1.06\,\mu L\, KOAc. \end{gathered}$$

PVA is a concentration-enhancing polymer that may increase reaction efficiency. It may increase splicing, but is not essential to the reaction [19].

22. Taking a $T = 0$ aliquot and freezing it on dry ice stalls the reaction from proceeding. The $T = 0$ aliquot can be used as a zero time point to mark the initiation of the reaction.

23. Reaction times can be varied and optimized depending on the experimental conditions.

24. If there is residual debris stuck to the plates, water will be forced to flow around the debris. Hold the gel at a 45° angle and run water over it. Move glass plates slowly and scan for debris. Gently wipe debris away if any is present.

25. Cleaning glass plates with 100% ethanol will lead to faster removal of the silicone based coating; therefore, it is best to use 70% ethanol.

26. Place two spacers on the vertical edges of the non-siliconized glass plate. Make sure the spacers are aligned. Place the third spacer horizontally at the bottom of the glass plate. Make sure spacers are flush with each other and place the second (notched) glass plate onto the first. Make sure the siliconized side is facing toward the inside of the gel cassette.

27. If the spacers move during the attachment of 1¼″ binder clips, gaps can appear between them, which will lead to leakage from the gel cassette.

28. Once APS and TEMED have been added it is important to work quickly because the acrylamide will start to solidify. Invert the conical tube three times to ensure mixing.

29. It is critical to avoid bubbles. If a bubble appears inside the gel, tilt the gel cassette to one side and then bring it back to its original position with a rapid and continuous movement.

30. For the best results, press the gel comb flat against the non-siliconized glass plate. Insert the comb by placing thumbs on both sides of the comb and gently pushing down. Do not push the comb down past the top of the wells.

31. To remove the comb, hold the gel at a 45° angle (alternatively, place the gel flat on raised surface such as test tube rack) and place both thumbs at the edges of the comb. Gently push upward, taking care to distribute force evenly.

32. Ensure there are no bubbles on the bottom of the gel cassette. The presence of bubbles will cause uneven current distribution and may cause the gel box to shut off. Take a syringe with a bent needle and draw up buffer. Slide the needle along the bottom of the gel cassette in the opening between the two glass plates and simultaneously dispense buffer with minimal force to move the bubbles out.

33. Glycogen is polysaccharide carrier that aids in the precipitation and visualization of nucleic acids in ethanol precipitation protocols.

34. For 25 μL reactions, add 175 μL of Proteinase K master mix; however, this amount should be scaled up or down depending on the volume of the splicing reaction used.

35. Place tubes in a uniform fashion so that the pellet will appear on the same side for each tube. It is important to minimize loss of the RNA during the ethanol precipitation. If tubes are placed in a uniform manner such that the pellet appears in the same spot, the experimenter can assume the pellet is there even if it is faint or not visible.

36. Wash out the wells of the gel prior to loading samples to remove any residual urea that has settled at the bottom of the wells. Using a syringe with a straight needle attached aspirate some buffer from the top compartment of electrophoresis apparatus and dispense buffer into the wells with mild force to displace the urea.

37. Use 1¼″ binder clips to clamp a precooled aluminum plate to the front of the gel cassette. This will help dissipate heat and prevent the glass plates from breaking.

38. Phosphor storage technology makes use of phosphorscreens that are composed of BaFBr:Eu^{2+} crystals immobilized in an organic matrix. Phosphorscreens can be used for autoradiography and have many advantages over the traditional method of

exposing x-ray film to radioactive gels. For example, phosphorscreens have between 10 and 250 times higher sensitivity and a linear dynamic range that spans five orders of magnitude. High-energy radiation emitted from radioactive atoms, such as ^{32}P, oxidizes Eu^{2+} to Eu^{3+} and leads to an electron being trapped in the BaFBr complex. The reduced BaFBr^{-} has a unique absorbance in the 600 nm range. Exposure of the phosphorscreen to a 633 nm wavelength scanning laser results in the oxidation of the BaFBr^{-} complex and reduction of Eu^{3+} back to Eu^{2+}. The reduction of Eu^{3+} to Eu^{2+} leads to the emission of a photon with a wavelength of 390 nm. These photons can be detected with a photomultiplier instrument as the laser scans the phosphorscreen [20].

39. A 1 h exposure to the phosphorscreen is sufficient for pre-mRNA that has a specific activity of 800 cpm/μL or higher. Longer exposure times may be needed to visualize less abundant splicing intermediates; an optimum exposure time can be found through trial and error.

40. There are many forms of analysis for autoradiograms of in vitro splicing reactions. One of the simplest forms of analysis is to calculate the percent of spliced product (Fig. 2). Calculate the sum of the volume intensity for the spliced band and the unspliced band taking into account background and time 0 (total signal). Divide the intensity of the band corresponding to the spliced product by the total signal intensity and then multiply by 100 to obtain a percentage.

$$\%\,Spliced\ = \frac{Signal\ from\ final\ spliced\ product}{Total\ signal\ in\ lane}$$

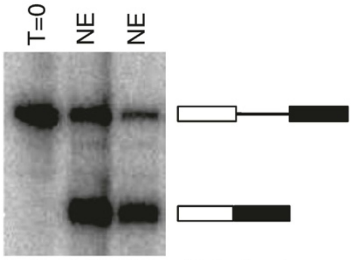


Fig. 2 Autoradiogram of radiolabeled β-globin minigene pre-mRNA splicing reaction (from *left* to *right*) at time 0 (*T* = 0) (*lane 1*) or incubated for 90 min in 30% nuclear extract (NE) from two different extract preparations (*lane 2* and *3*). Reaction products were separated on a 6% polyacrylamide gel. The splicing efficiency (% spliced) was determined using software from Bio-Rad Quantity One® (*see* **Note 41**)

Acknowledgments

Research in the Hertel laboratory is supported by NIH (GM062287 and GM110244). Maliheh Movassat and Hossein Shenasa contributed equally to this work.

References

1. Merkhofer EC, Hu P, Johnson TL (2014) Introduction to cotranscriptional RNA splicing. Methods Mol Biol 1126:83–96
2. Moore MJ, Query CC, Sharp PA (1993) 13 Splicing of precursors to mRNA by the Spliceosome. Cold Spring Harb Monogr Arch 24:303–357
3. Hertel KJ (2008) Combinatorial control of exon recognition. J Biol Chem 283: 1211–1215
4. Roca X, Karginov FV (2012) RNA biology in a test tube – an overview of in vitro systems/assays. Wiley Interdiscip Rev RNA 3:509–527
5. Hicks MJ, Lam BJ, Hertel KJ (2005) Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays. Methods 37:306–313
6. Movassat M, Mueller WF, Hertel KJ (2014) In vitro assay of pre-mRNA splicing in mammalian nuclear extract. Methods Mol Biol 1126:151–160
7. Cooper TA (2005) Use of minigene systems to dissect alternative splicing elements. Methods 37:331–340
8. Dignam JD, Lebovitz RM, Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res 11:1475–1489
9. Krainer AR, Maniatis T, Ruskin B, Green MR (1984) Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell 36:993–1005
10. Pugh BF (1995) Preparation of HeLa nuclear extracts. Methods Mol Biol 37:349–357
11. Mayeda A, Krainer AR (1999) Preparation of HeLa cell nuclear and cytosolic S100 extracts for in vitro splicing. Methods Mol Biol 118:309–314
12. Abmayr SM, Yao T, Parmely T, Workman JL (2006) Preparation of nuclear and cytoplasmic extracts from mammalian cells. Curr Protoc Mol Biol Chapter 12:Unit 12.1
13. Kataoka N, Dreyfuss G (2008) Preparation of efficient splicing extracts from whole cells, nuclei, and cytoplasmic fractions. Methods Mol Biol 488:357–365
14. Webb C-HT, Hertel KJ (2014) Preparation of splicing competent nuclear extracts. Methods Mol Biol 1126:117–121
15. Nilsen TW Preparation of nuclear extracts from HeLa cells. Cold Spring Harb Protoc 2013, 2013:579–583
16. Hernandez N, Keller W (1983) Splicing of in vitro synthesized messenger RNA precursors in HeLa cell extracts. Cell 35:89–99
17. Reichert V, Moore MJ (2000) Better conditions for mammalian in vitro splicing provided by acetate and glutamate as potassium counterions. Nucleic Acids Res 28:416–423
18. Carey MF, Peterson CL, Smale ST (2009) Dignam and Roeder nuclear extract preparation. Cold Spring Harb Protoc 4:1–4
19. Mayeda A, Krainer AR (1999) Mammalian in vitro splicing assays. Methods Mol Biol 118:315–321
20. Johnston RF, Pickett SC, Barker DL (1990) Autoradiography using storage phosphor technology. Electrophoresis 11:355–360