2016

Determination Of A Comprehensive Alternative Splicing Regulatory Network And The Combinatorial Regulation By Key Factors During The Epithelial To Mesenchymal Transition

Yueqin Yang
University of Pennsylvania, yueqinyang@gmail.com

Determination Of A Comprehensive Alternative Splicing Regulatory Network And The Combinatorial Regulation By Key Factors During The Epithelial To Mesenchymal Transition

Abstract

The epithelial to mesenchymal transition (EMT) is a process by which epithelial cells transdifferentiate into mesenchymal cells. It is essential for embryonic development and implicated in cancer metastasis. While the transcriptional regulation of EMT has been well-studied, the role of post-transcriptional regulation, particularly alternative splicing (AS) regulation in EMT, remains relatively uncharacterized. We previously showed that the epithelial cell-type-specific proteins ESRP1 and ESRP2 are important for regulation of many AS events that altered during EMT. However, the contributions of the ESRPs and other splicing regulators to the splicing regulatory network in EMT require further investigation.

In the first part of my thesis, we used a robust in vitro EMT model to comprehensively characterize splicing switches during EMT in a temporal manner. These investigations revealed that the ESRPs are responsible for a large number of AS events during EMT. We determined that RBM47 transcript is down-regulated during EMT and RBM47 depletion regulates many EMT-associated AS events. We also determined that Quaking (QKI) broadly promotes mesenchymal splicing patterns for numerous EMT-associated AS events. Our study highlights the broad role of post-transcriptional regulation and the important role of combinatorial regulation by different splicing factors to fine tune gene expression programs during the EMT.

In the second part of my thesis, we determined that Esrp1 generates both nuclear and cytoplasmic isoforms, due to the use of two competing alternative 5' splice sites in exon 12. We carried out a detailed characterization of the Esrp1 nuclear localization signal (NLS) that represents a novel class of NLS. Furthermore, we identified splice variants encoding nuclear and cytoplasmic isoforms of Fusilli, the Esrp1 orthologue in D. Melanogaster. Our observations demonstrate that the production of both nuclear and cytoplasmic Esrp1 isoforms through alternative splicing is highly conserved among species, strongly suggesting it's biologically significant. Thus, while the first part of thesis has characterized extensive regulation by nuclear Esrp1 to promote epithelial splicing patterns, it will be of great interest to study the contribution of cytoplasmic Esrp1 in the maintenance of epithelial cell functions.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Cell & Molecular Biology

First Advisor

Russ P. Carstens

Second Advisor

Kristen W. Lynch

Keywords

Alternative splicing, Epithelial to mesenchymal transition, ESRP1/2, Nuclear localization signal, QKI,

RBM47

DETERMINATION OF A COMPREHENSIVE ALTERNATIVE SPLICING REGULATORY NETWORK AND THE COMBINATORIAL REGULATION BY KEY FACTORS DURING THE EPITHELIAL TO MESENCHYMAL TRANSITION

Yueqin Yang

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

Russ P. Carstens, M.D.

Associate Professor of Medicine

Graduate Group Chairperson

Daniel S. Kessler, Ph.D., Associate Professor of Cell and Developmental Biology

Dissertation Committee

Kristen W. Lynch, Ph.D. (chair), Professor of Biochemistry and Biophysics

Brian D. Gregory, Ph.D., Associate Professor of Biology

Thomas A. Jongens, Ph.D., Associate Professor of Genetics

Stephen A. Liebhaber, M.D., Professor of Genetics

Dedication

For those who allow me to be a better person and for me to know myself better.

ACKNOWLEDGMENT

First and most important, I would like to thank my advisor, Dr. Russ Carstens, for all his guidance throughout my doctoral studies. He gave me the opportunities to develop my own interest and helped me when I came across difficulties. I can't say enough how many times he had made me realized how to improve myself as person.

I would like to thank the current and past members of the Carstens lab for their constant help. Particularly, I would like to thank Claude Warzecha and Thomas Bebee. Claude taught me every basic skill to start my scientific experiment and he was so patient and kind to me. Thomas who I looked up to, showed me what a good scientist is.

I also want to thank my collaborators, Dr, Yi Xing and Dr. Juw Won Park for their computational and bioinformatic analyses. Without them, I wouldn't be able to achieve my thesis.

The members of my thesis committee, Dr. Brian Gregory, Dr. Kristen Lynch, Dr. Thomas Jongens, and Dr. Steve Liebhaber have provided helpful discussions over the past few years.
Dr. Brian Gregory —The energy, the creative thoughts from you have always inspired me to explore more.
Dr. Kristin Lynch — I owe a huge thanks to you and your lab for the help on my experiments
Dr. Thomas Jongens — I see the beauty of being insightful, while I see the strength of being calm and humble.
Dr. Steve Liebhaber — I will always remember that you encouraged me to become a scientist.

Lastly, I would like to thank my parents for their selfless care and support. If there was just one thing I learned from all these years abroad, it is the unparalleled greatness of father and mother.

ABSTRACT

DETERMINATION OF A COMPREHENSIVE ALTERNATIVE SPLICING REGULATORY NETWORK AND THE COMBINATORIAL REGULATION BY KEY FACTORS DURING THE EPITHELIAL TO MESENCHYMAL TRANSITION

Yueqin Yang

Russ P. Carstens

The epithelial to mesenchymal transition (EMT) is a process by which epithelial cells transdifferentiate into mesenchymal cells. It is essential for embryonic development and implicated in cancer metastasis. While the transcriptional regulation of EMT has been well-studied, the role of post-transcriptional regulation, particularly alternative splicing (AS) regulation in EMT, remains relatively uncharacterized. We previously showed that the epithelial cell-type-specific proteins ESRP1 and ESRP2 are important for regulation of many AS events that altered during EMT. However, the contributions of the ESRPs and other splicing regulators to the splicing regulatory network in EMT require further investigation.

In the first part of my thesis, we used a robust *in vitro* EMT model to comprehensively characterize splicing switches during EMT in a temporal manner. These investigations revealed that the ESRPs are responsible for a large number of AS events during EMT. We determined that RBM47 transcript is down-regulated during EMT and RBM47 depletion regulates many EMT-associated AS events. We also determined that Quaking (QKI) broadly promotes mesenchymal splicing patterns for numerous EMT-associated AS events. Our study highlights the broad role of post-transcriptional regulation and the important role of combinatorial regulation by different splicing factors to fine tune gene expression programs during the EMT.

In the second part of my thesis, we determined that Esrp1 generates both nuclear and cytoplasmic isoforms, due to the use of two competing alternative 5' splice sites in exon 12. We carried out a detailed characterization of the Esrp1 nuclear localization signal (NLS) that

represents a novel class of NLS. Furthermore, we identified splice variants encoding nuclear and cytoplasmic isoforms of Fusilli, the Esrp1 orthologue in *D. Melanogaster*. Our observations demonstrate that the production of both nuclear and cytoplasmic Esrp1 isoforms through alternative splicing is highly conserved among species, strongly suggesting it's biologically significant. Thus, while the first part of thesis has characterized extensive regulation by nuclear Esrp1 to promote epithelial splicing patterns, it will be of great interest to study the contribution of cytoplasmic Esrp1 in the maintenance of epithelial cell functions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

1) Chapter 1: Introduction

1.1 Splicing and the underlying reactions

In 1977, several groups reported that certain viral encoded mRNAs are not collinear with the corresponding genomic DNA (Berget et al. 1977; Chow et al. 1977). To account for these split genes, three major hypotheses were proposed, including 1) certain genomic DNA segments are recombined before transcription, 2) certain genomic DNA is selectively transcribed and those transcribed RNA segments are joined together post-transcription, or 3) the complete genomic DNA is transcribed into precursor RNA (pre-RNA) and the non-coding segments are removed post-transcription (Greer and Abelson 1984). The last model was named “RNA splicing” and now accepted as the main mechanism for the split genes. RNA splicing is a process where the non-coding segments (introns) are removed from pre-RNA while the coding segments (exons) are joined to form the mature RNA. It is required for the processing of different types of precursor RNA transcribed from nuclear and organellar genomes, including messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA).

In eukaryotes, at least four distinct mechanisms of splicing reaction are identified, including 1) the spliceosomal splicing for nuclear pre-mRNA; 2) self-splicing for Group I intron; 3) self-splicing for Group II intron; 4) splicing for nuclear tRNA (Phizicky and Greer 1993). The splicing mechanism for nuclear pre-mRNA, Group I and Group II introns shares some similarities. First, each splicing event occurs in two sequential trans-esterification reactions. The first reaction features a nucleophilic attack from a ribose hydroxyl group at the 5’ splice site (5’ ss, the boundary between the upstream exon and intron), which cleaves the upstream exon with 3’ hydroxyl group and produces an intermediate with both intron and downstream exon. The second

step involves the newly formed 3’ hydroxyl group, which attacks and cleaves the 3’ splice site (3’ ss, the boundary between the intron and downstream exon), therefore joining two exons and releasing the intron. Additionally, for all three classes of splicing, the catalytic core is composed of RNA rather than protein. Notably, ribozyme or RNA based enzyme, was first discovered from the self-splicing reaction of Group I introns (Cech 1990). Some major differences also present in the splicing mechanisms for nuclear pre-mRNA, Group I and Group II introns. For nuclear pre-mRNA and Group II intron, the first trans-esterification reaction is initiated by the 2’ hydroxyl group of a branch point nucleotide (usually adenosine) in the intron, which produces a lariat intermediate. Whereas, for Group I introns, it is the 3’ hydroxyl group of a guanosine cofactor, therefore a linear intermediate is produced upon cleavage. Although splicing is catalyzed by RNA in all three classes, the nuclear pre-mRNA splicing is carried out by a large and dynamic RNA-protein complex called spliceosome.

In eukaryotes, there are two types of spliceosomes: the U2-dependent spliceosome (also refer to as major spliceosome), which catalyzes the removal of U2-type introns, and the U12-dependent spliceosome (or minor spliceosome) which splices U12-type introns (Turunen et al. 2013; Lee and Rio 2015). U2-type introns comprise more than 99% of all identified introns while U12-type introns only occur at very low frequency in most of eukaryotes. For U2-type introns, the 5’ ss, branch point sequence (BPS) and 3’ ss are well-established and conform to consensus sequences with various degree of degeneracy (Fig 1-1A). In addition, there is a polypyrimidine tract (Py tract), a 15-20 nucleotide long, pyrimidine-enriched sequence located 5-40 nucleotides upstream of 3’ ss. The dinucleotides at each end of the intron are most conserved, which are “GT” at the 5' end of the intron and “AG” at the 3' end of the intron. Based on the similarity between the splice sites to the consensus sequences, they can be classified as “strong” or “weak” splice sites. In contrast, the U12-type introns were initially identified as a different class of introns utilizing “AT” and “AC” as their terminal dinucleotides. It was soon realized that these ATAC introns are not exclusively U12-dependent. Subsequent study revealed that most U12-type introns actually utilize “GT” and “AG” dinucleotides pair. The defining feature of U12-type introns

are the 5’ ss and the branch point sequence, which are more conserved than the U2-type introns (Fig 1-1B). Each spliceosome consists of five small nuclear ribonucleoproteins (snRNPs) as core splicing factors and many accessory proteins. Each snRNP contains one small nuclear RNA (snRNA) and a number of protein components, including the core Sm proteins in all five snRNPs and many other proteins specific to each snRNP. While U5 snRNP is shared between both spliceosomes, the U2-dependent spliceosome also contains U1, U2, U4/6 snRNP. In the other end, the U12-dependent spliceosome contains four distinct snRNPs, U11, U12, U4atac and U6atac, which are equivalent to their courter-parts in the major spliceosome.

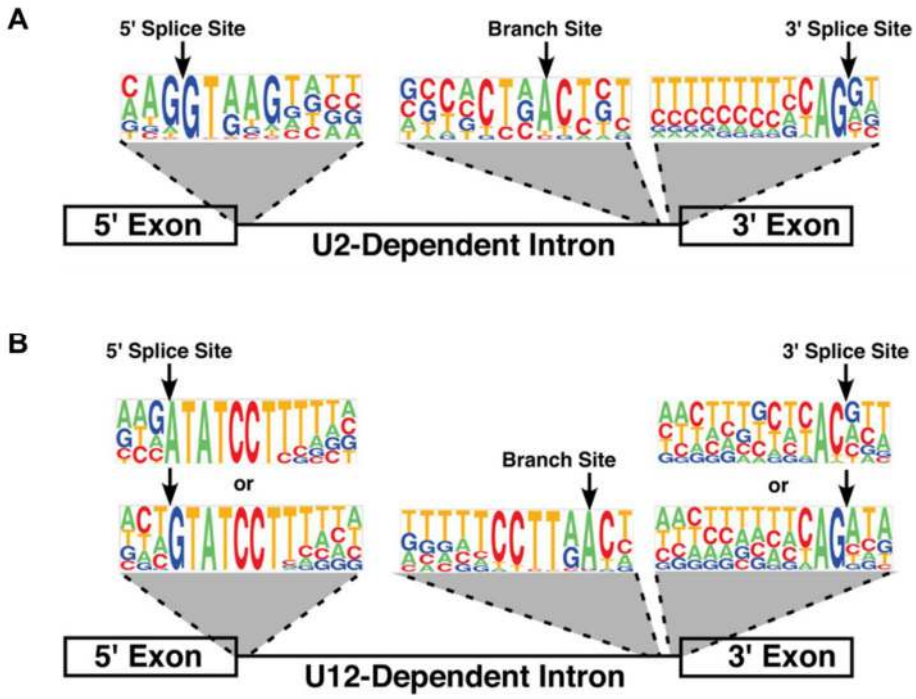


Figure 1-1 Comparison of the consensus sequences for 5’ splice site, branch point sequence and 3’ splice site between U2-dependent and U12-dependent introns.

The size of each letter represents the frequency of each base at each position over all introns with the most frequent base on top.

(A) Splice-site consensus sequences for U2-dependent introns

(B) Splice-site consensus sequences for U12-dependent introns. (Adapted from Padgett, Richard A. "New connections between splicing and human disease." Trends in Genetics 28.4 (2012): 147-154.)

During splicing, the spliceosome assembles and dissembles in a specific manner through the recognition of cis-elements in the pre-mRNA and interaction among different snRNPs (Lee and Rio 2015). *In vitro* studies of spliceosome assembly, using extracts made from whole yeast cells or mammalian cell nuclei, have defined a stepwise, ordered interaction of the snRNPs with the pre-mRNA. Since both spliceosomes are assembled in a similar way and share the overall structure, I will use major spliceosome assembly as an example. In the first step of major spliceosome assembly, the 5' ss is recognized by U1 snRNP, while U2 Auxiliary Factor-65 kD subunit (U2AF65) binds to the Py tract and U2AF35 binds to the 3'ss. This assembly is ATP independent and results in the formation of the E complex or the commitment complex. U2 snRNP then binds to the BPS to form the A complex and the branch point adenosine is bulged to facilitate the nucleophilic attack in the first trans-esterification reaction. Notably, this is the first ATP-dependent step during splicing. Next, the U4/U6, U5 tri-snRNP particle is recruited to form the B complex. Multiple structural rearrangements then take place in a ATP-dependent manner to form C complex, which is endowed with the catalytic activity of splicing. Several key rearrangements include 1) substitution of U1 snRNP with U6 snRNP, which binds downstream of 5' ss, 2) destabilization of U4 and U6 snRNA base pairing, 3) release of U1 and U4 snRNPs, 4) formation of extensive base pairing between U2 and U6 snRNA, which is a pre-requisite for the first step of splicing, 5) binding of U5 snRNA to both exon sequences, which may assist in bringing them close to facilitate the second step of splicing. After exon joining, U2, U5 and U6 snRNPs are released from the intron lariat, which is degraded later. In addition to this traditional model of stepwise assembly, purification of penta-snRNP particle under physiological salt condition also suggests a different model with preassembly of all spliceosomal snRNPs before binding to the pre-mRNA (Stevens et al. 2002). However, these models emerge mainly from *in vitro* studies and subsequent *in vivo* studies support a stepwise assembly of spliceosome in yeast (Görnemann et al. 2005).

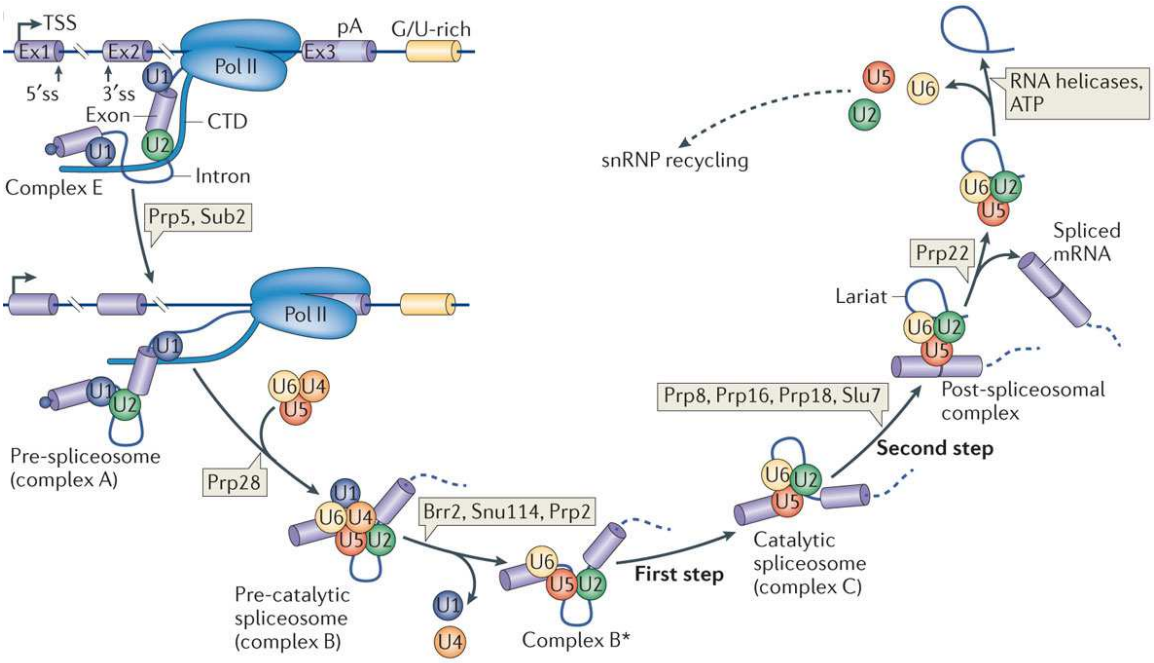


Figure 1-2 The spliceosome assembly and splicing reaction for U2-dependent introns. (Adapted from Matera, A. Gregory, and Zefeng Wang. "A day in the life of the spliceosome." Nature reviews Molecular cell biology 15.2 (2014): 108-121.)

On the other hand, the splicing mechanism is very different for nuclear tRNA splicing, which is best characterized in yeast (Culbertson and Winey 1989). Nuclear tRNA splicing occurs in three steps and each step is catalyzed by a protein enzyme. The first step requires a tRNA endonuclease to recognize and cleave the pre-tRNA at both 5’ splice site and 3’ splice site, which produces two tRNA half-molecules with 2',3' cyclic phosphate and 5' hydroxyl ends and an excised intron with 5' hydroxyl and 3' phosphate ends. The second step involves a tRNA ligase, a heterotetramer protein comprised of four distinct enzymatic activities to catalyze a series of reactions that lead to joining of the two half-molecules with a 2′ phosphate at the splice junction. Therefore, a 2’-phosphotransferase is required for the removal of the 2’ phosphate at the third step.

1.2 Alternative splicing: definition and classification

In eukaryotes, it is very common for one pre-mRNA to produce several different mature mRNA via a process called alternative splicing (AS). For example, it has been estimated that more than 95% of human genes are alternatively spliced (Moroy and Heyd 2007; Pan et al. 2008; Wang et al. 2008). AS not only increases the coding capacity, it also regulates mRNA stability, translational efficiency and subcellular localization. For example, AS can introduce a premature termination codon (PTC) into the mature mRNA which is subjected to nonsense-mediated decay (NMD), therefore decreasing the expression of target gene (Hamid and Makeyev 2014; Jangi et al. 2014). Five major types of alternative splicing have been described so far (Fig 1-1A) (Keren et al. 2010). The first one is skipped exon (SE), also known as cassette exon, in which the exon can be included or excluded in the mature mRNA. Cassette exons account for about 40% of all identified AS events in higher eukaryotes (Sugnet et al. 2004; Alekseyenko et al. 2007). The second and third types are alternative 5' or 3' splice sites (A5SS or A3SS), in which more than one 5' ss or 3' ss can be chosen for splicing of a single exon. The fourth type are mutually exclusive exons, where only one exon can be included in the mature mRNA at a time among a group of adjacent cassette exons. The last one is intron retention, where introns can be included in the mature mRNA. In addition, alternative splicing can be coupled with alternative polyadenylation or alternative promoters, which expands regulatory potential (Fig 1-1B).

Proper splicing is biologically and physiologically important for normal development. Aberrant splicing caused by mutations or mis-regulated splicing is implicated in human diseases, including neurodegenerative diseases and cancers. Mutations in the splice sites and other core cis-elements in pre-mRNAs are among the most common mutations identified in studies of genetic diseases (Cooper et al. 2009; Cieply and Carstens 2015). These mutations can potentially change the coding capacity in different ways, via exon skipping, intron retention, usage of cryptic splice sites or generation of a PTC, which in turn lead to human diseases. Mutations in the core spliceosomal components or other splicing regulatory proteins are also associated with

human diseases. A well-known example is the single nucleotide substitution in exon 7 of the survival of motor neuron 2 (*SMN2*) gene and its critical role in spinal muscular atrophy (SMA) (Cooper et al. 2009). SMN proteins are essential for snRNP biogenesis. In human, there are two *SMN* genes, *SMN1* and *SMN2*, both encoding the same protein. The mutation in SMN2 leads to frequent skipping of exon 7 without changing the encoded amino acid, producing an inactive and unstable protein isoform. In SMA patients, who usually suffer from loss of SMN1, the deficiency in SMN2 to produce fully functional protein cause the disease. In addition, mis-regulation of splicing also plays an important role in human diseases. A well-documented example is the myotonic dystrophy types 1 and 2 (DM1 and DM2) caused in part by the mis-regulation of two AS factors, MBNL1 and CELF1 (also known as CUGBP1). Microsatellite expansions in certain transcripts leads to the formation of nuclear aggregates, which bind to splicing factors and sequester them from being available for normal splicing regulation (Cieply and Carstens 2015). Since splicing is crucial for normal development, it is tightly regulated.

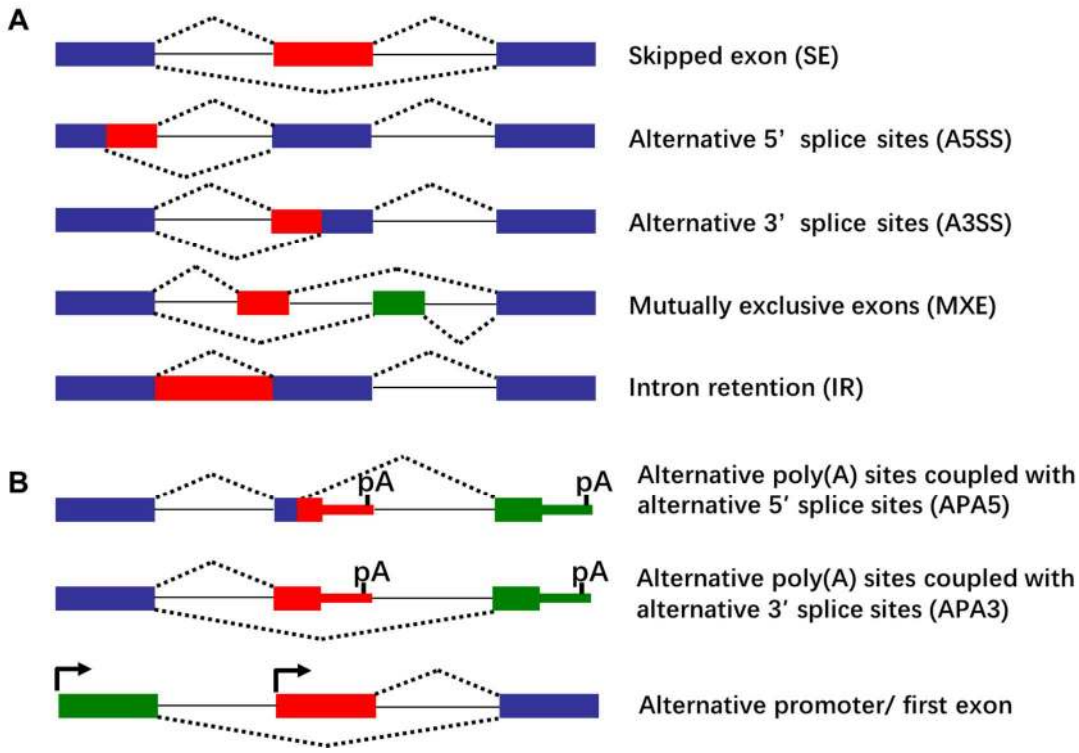


Figure 1-3 Different types of alternative splicing.

Blue boxes indicate constitutive exons. Red and green boxes indicate alternatively spliced sequences. Black lines indicate introns.

(A) Schematic of different types of alternative splicing. A skipped exon (or cassette exon) can either be included or excluded in mature mRNA. Alternative 5' and 3' splice sites occur when adjacent 5' or 3' splice sites are chosen for splicing. Mutually exclusive exons refer to only one exon is included in the mature mRNA at a time while other adjacent cassette exons are excluded. Intron retention occurs when introns remains in the mature mRNA.

(B) Schematic of alternative poly(A) sites or alternative promoters coupled alternative splicing. (adapted from Dr. Russ Carstens splicing lecture)

1.3 Regulation of constitutive and alternative splicing by RNA binding proteins

Splicing is under complex regulation by various cis-elements in pre-mRNAs, and their cognate binding partners, mainly RNA-binding proteins (RBPs). Splicing regulatory RBPs recruited to their respective binding sites can have positive or negative effects on splicing of different exons or splice sites. Based on the position in the pre-mRNA and effect on splicing outcome, the cis-elements (usually short, degenerate sequences) are classified into four distinct groups: exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE) and intronic splicing silencer (ISS). In most cases, the decision of splicing is made during the early spliceosome assembly (formation of the E complex and the A complex) where the recognition of splice sites by U1, U2AF, and U2 snRNPs is the main regulatory step. In general, constitutive exons tend to have strong 5' and 3' ss, which are less dependent on regulatory proteins to recruit U1 and U2 snRNPs (Stamm et al. 1994). In contrast, the 5' and 3' ss of alternative exons are relatively weaker, therefore other splicing factors have substantial effects on splice sites recognition (Stamm et al. 1994). The paring between 5' ss and 3' ss determines the splicing outcome, which can occur in an exon-oriented or intron-oriented manner (Berget 1995; De Conti et al. 2013). In the human genome, a typical exon is around 50-250 nt in size while the introns are usually much longer (frequently 10^3 to 10^4 nt) (Chen and Manley 2009; Lee and Rio 2015). Due to the length of the introns, early spliceosome assembly is believed to occur across the exons though a process referred to as "exon definition" (Robberson et al. 1990; Berget 1995). In contrast, in lower eukaryotes such as yeast and fly where the introns are much smaller, early spliceosome assembly is centered around the introns, which is referred to as "intron definition" (Ruby and Abelson 1991; Talerico and Berget 1994). In general, the size of exon or intron is key to the decision as to which model is preferred and manipulation of the length of an exon or intron may lead to aberrant splicing. Indeed, increasing the length of an internal exon in human genes can lead either to exon skipping or splicing to a cryptic splice site within the expanded exon (Robberson et al. 1990) . Not only there is usually a limitation on the maximum size for internal

exons enforced by the “exon definition” model, the exons can’t be too small where steric hindrance will prevent binding of U1 and U2 snRNPs concomitantly. Similar results were obtained for fly or yeast genes, where expansion of a short intron will prevent it from splicing or result in splicing to a cryptic splice site (Guo et al. 1993; Talerico and Berget 1994). While early spliceosome assembly mainly occurs in an exon-oriented manner in human genes, the two-step splicing reaction requires paring of splice sites in an intron-oriented manner. Subsequent studies propose that a transition from “exon definition” to “intron definition” is essential in the later stage of splicing, where the cross-intron interactions between U1 and U2 snRNPs may play a major role (Lim and Hertel 2004; Kotlajich et al. 2009).

Many RBPs with largely ubiquitous expression in different tissues and cells have been shown to have broad impact on splicing and important cell functions, such as SR and hnRNP (Heterogeneous nuclear ribonucleoproteins) protein families (Nilsen and Graveley 2010). SR proteins are among the most characterized splicing factors. A prototypical SR protein contains one or two N-terminal RNA recognition motifs (RRMs) and a C-terminal RS domain, a region that is enriched with arginine–serine dipeptides repeats (Graveley 2000). The RRM domains are believed to mediate the binding to the ESEs on pre-mRNAs and the RS domains appear to mediate protein-protein interactions among SR proteins and other proteins with RS (or RS-like) domains (Graveley 2000). While it is generally accepted that SR proteins promote exon inclusion through their binding to ESEs, recent studies also suggested they can inhibit splicing as well (Blencowe 2000; Graveley 2000; Pandit et al. 2013). Previous studies have established that SR proteins are essential for early spliceosomal assembly (E complex formation) on certain pre-mRNAs and play an important role in both constitutive and alternative splicing regulation. Various mechanisms were discovered by which SR proteins promote exon inclusion. First, the RS domain of SRSF1 was shown to directly interact with U1 70K, a key component in the U1 snRNP, and U2AF35 through their RS domains, therefore bridges the 5’ ss and 3’ss (Wu and Maniatis 1993; Zuo and Maniatis 1996; Hastings and Krainer 2001). This interaction between U1 and U2 snRNPs mediated by SR protein can occur across exons or introns, which reflects the “exon

definition" or "intron definition" model for early spliceosome assembly respectively (Ast 2004). Further studies showed that phosphorylation status of various serine residues in the RS domains are crucial for protein-protein and protein-RNA interactions (Tacke et al. 1997; Xiao and Manley 1997; Xiao and Manley 1998). The requirement of SR proteins is especially important for those pre-mRNAs that have relatively weak Py tract, where ESE bound SR proteins along with 3'ss bound U2AF35 help to stabilize the binding of U2AF65 to facilitate early spliceosome assembly (Merendino et al. 1999; Wu et al. 1999; Zorio and Blumenthal 1999). While it was believed that the RS domain of SR proteins are key to their role in regulating splicing, other studies shed light on a RS domain-independent role of SR proteins in splicing regulation (Zhu and Krainer 2000). SRSF1 without the RS domain can promote splicing *in vitro* (Zhu and Krainer 2000). Two major mechanisms were proposed. First, other domains can mediate protein-protein interaction to help recruitment of spliceosomal component. Indeed, studies showed that the RRM domain of SRSF1 can bind to U1 70K, which stabilizes the binding of U1 snRNP to the 5'ss (Cho et al. 2011). A second possibility is ESE bound SR proteins compete with splicing inhibitors that bind to the adjacent ESS in the pre-mRNAs, therefore promoting exon inclusion by antagonizing the inhibitive effect (Mayeda and Krainer 1992; Zhu and Krainer 2000). HnRNPs represent a large family of RBPs that contribute to multiple aspects during mRNA processing post-transcription (Chaudhury et al. 2010; Geuens et al. 2016). They were initially identified as binding partners of pre-mRNAs and mature mRNAs. A typical hnRNP usually contain one or more RNA binding domains, such as RRM, K-homology (KH) domain or RGG (Arg-Gly-Gly) box, and some type of auxiliary domains (Chaudhury et al. 2010; Geuens et al. 2016). Like SR proteins, hnRNPs play an important role in both constitutive and alternative splicing. In contrast to SR proteins, hnRNPs generally repress splicing through binding to ESSs or ISSs. For example, HnRNP A1 can inhibit splicing by blocking the binding of U1 snRNP to the 5'ss (Eperon et al. 2000; David et al. 2010). This "bind and block" mechanism is similarly utilized by another hnRNP protein, PTBP1 (polypyrimidine tract binding protein 1, also known as hnRNP I) which binds to the Py tract and competes with U2AF65 during the early spliceosome formation (Lin and Patton 1995; Singh et al.

1995). Subsequent studies suggest models where multiple PTBP1 binding sites in an alternative splicing event can induce propagative binding along the RNA or looping of RNA, therefore blocking splicing factors from binding (Wagner and Garcia-Blanco 2001; Oberstrass et al. 2005). Meanwhile, several other mechanisms by which PTBP1 interferes with the cross-intron or cross-exon interactions are also proposed (Izquierdo et al. 2005; Spellman and Smith 2006; Sharma et al. 2008). One of such suggests PTBP1 antagonizes exon-definition by interfering with the interaction between U1 snRNP and U2AF65, rather than competing with U2AF65 for binding to Py tract (Izquierdo et al. 2005).

In addition to SR proteins and hnRNPs that are mostly ubiquitously expressed across many tissues, recent studies described important roles for tissue-specific RBPs, such as NOVA, PTBP2 (nPTB), MBNL and RBFOX protein families, as essential regulators of tissue- or cell-type-specific splicing (Ule et al. 2003; Gehman et al. 2012; Licatalosi et al. 2012; Lee et al. 2013; Li et al. 2014). For example, NOVA proteins are neuronal tissue-specific RBPs that regulate alternative splicing in brain (Jensen et al. 2000; Ule et al. 2005). Newly developed techniques such as cross-linking immunoprecipitation followed by sequencing (CLIP-seq) and its derivatives have been widely used to determine the mRNA targets and binding motifs for RBPs. Analysis of the distribution of NOVA binding sites on their target transcripts and the splicing outcome revealed an “RNA-map”, in which the position of NOVA binding sites in target transcripts can predict the outcome of splicing to a large extend (Ule et al. 2006; Licatalosi et al. 2008). Specifically, binding in the downstream intron promotes exon inclusion while binding in the upstream intron promotes exon skipping. The RNA map also provides some insight into the mechanistic aspect of NOVA-regulated splicing. It was shown that binding of NOVA to an ESS represses exon inclusion by blocking the binding of U1 snRNP, therefore inhibiting the formation of the E complex (Ule et al. 2006). In contrast, NOVA binding to an ISE downstream of the alternative exon enhances exon inclusion by promoting the formation of the A complex and following complexes (Ule et al. 2006). Subsequent studies showed that other splicing factors, such as RBFOX protein family members, MBNL protein family members, ESRP1/2 and Quaking,

also generally adhere to this position-dependent RNA-map (Zhang et al. 2008; Warzecha et al. 2010; Dittmar et al. 2012; Wang et al. 2012; Hall et al. 2013). The mechanism underlying the position-dependent regulation is largely unknown. The general hypothesis for the silencing effects results from the competition with core splicing factors or other positive regulatory proteins on their binding sites in the upstream intron or exon (Witten and Ule 2011). For example, RBFOX proteins were reported to inhibit splicing at two distinct steps depending on its binding position (Zhou and Lou 2008). When binding to an ISS, they will block the binding of splicing factor 1 (SF1, a core splicing factor) to the branch point sequence at the very beginning of spliceosome assembly (Zhou and Lou 2008). When binding to an ESS, they will prevent two splicing enhancers, namely TRA2 and SRp55, from binding to a nearby ESE, therefore inhibiting the formation of the E complex by blocking U2AF65 from binding (Zhou and Lou 2008). On the other hand, the hypotheses for the enhancing effects, when RBP binds to the downstream intron, are more diverse (Witten and Ule 2011). It can directly bind and recruit core splicing factors to facilitate early spliceosome assembly, or interact with other positive regulatory proteins to stabilize binding of core splicing factors. It can also facilitate the “exon definition” or “intron definition”, or the transition from “exon definition” to “intron definition” through bridging cross exon or intron interactions or modulating the conformation of pre-mRNAs.

While splicing regulation by RBPs were commonly described at the early stages of spliceosome assembly, subsequent studies also showed regulatory roles of RBPs in the later stages of splicing. For example, hnRNP L was shown to prevent the transition from the A complex to B complex during splicing, potentially through interfering with cross-intron paring of U1 and U2 snRNPs (House and Lynch 2006). In the presence of hnRNP L, splicing is stalled at a stage where an A-like complex is formed, which has a similar composition and gel mobility as the A complex, however, progression into the B complex is inhibited. A potential explanation is that the A-like complex represents a cross-exon paring of U1 and U2 snRNPs, where the splicing is stalled since hnRNP L prevents the cross-intron pairing of U1 and U2 snRNPs and subsequent recruitment of tri-snRNP particle. Another example, RBM5 was also shown to inhibit the

recruitment of tri-snRNP particle, therefore blocking the formation of the B complex while the formation of the A complex is intact (Bonnal et al. 2008). Furthermore, RBM17 (or SPF45) was shown to inhibit splicing at the second catalytic step during the splicing reaction, which is at the very late stage of splicing (Lallena et al. 2002). Importantly, splicing of pre-mRNAs is usually under the control of more than one regulatory proteins, where the positive and negative splicing factors work in a cooperative or competitive manner. This combinatorial regulation is emerging as a general rule for splicing. For example, alternative splicing of exon 6 in Fas can generate two distinct mRNAs, which encode a transmembrane receptor that promotes apoptosis when exon 6 is included, or a soluble isoform of the receptor that prevents programmed cell death when skipped (Izquierdo et al. 2005). This splicing event is under regulation of TIA-1, a RNA-binding protein with three RRMs, and PTBP1, TIA-1 binds to an ISE in intron 6 and promotes the inclusion of exon 6 through enhancement of U1 snRNP recruitment to the adjacent 5'ss, which in turn stabilizes the recruitment of U2AF to the upstream 3' ss (Izquierdo et al. 2005). However, PTBP1 binds to an ESS in exon 6 to inhibit the binding of U2AF65, interfering with the "exon definition". Therefore, a balancing between PTBP1 and TIA-1 is required for the proper splicing of exon 6 (Izquierdo et al. 2005). An antagonistic function between SR proteins and hnRNPs have also been described for a few transcripts (Expert-Bezançon et al. 2004; Crawford and Patton 2006). At a genome-wide level, analysis of splicing events regulated by NOVA and RBFOX proteins showed there is a notable overlap in their target transcripts, where motif analysis of NOVA target transcripts presents a significant enrichment of RBFOX binding motifs (Zhang et al. 2010). This study also suggested that both proteins prefer to have an additive or synergistic effect on shared target rather than antagonistic functions. With more studies using genomic approaches including RNA-Seq and CLIP-Seq, we will have a further understanding of how RBPs function both cooperatively and antagonistically to determine the splicing outcome.

1.4 Regulation of constitutive and alternative splicing by other mechanisms

In addition to splicing regulation by RBPs, RNA-RNA interactions or RNA secondary structure can also regulate alternative splicing. An example is the Down syndrome cell adhesion molecule (Dscam) gene in *Drosophila melanogaster* (Graveley 2005; May et al. 2011). Dscam contains 95 alternatively spliced exons that potentially encode 38,016 distinct mRNA variants and protein isoforms. These 95 alternative exons are organized into four clusters, and each cluster is spliced in a mutually exclusive manner. The largest one, exon 6 cluster contains 48 variable exons and only one variable exon is included in each mature mRNA. By comparative genomics, two class of conserved sequence elements were identified in the exon 6 cluster (Graveley 2005). The first one is named “the docking site”, a 66 nt element in the intron downstream of constitutive exon 5. The central 24 nt of the docking site is nearly invariant among 16 insect species. The second class of conserved elements are referred to as “the selector sequences”, which are located upstream of each variable exon in the exon 6 cluster. These selector sequences are similar but not identical to each other, however, they are all complementary to a portion of the docking site. The only one selector sequence is predicted to base pairing with the docking site at a time, and the immediately downstream exon will be included in the mature mRNA. Another example is the regulation of a mutually exclusive splicing event in fibroblast growth factor receptor 2 (FGFR2), a transmembrane protein mediating FGF signaling pathway. While exon IIIb is included in epithelial cells, exon IIIc is included in mesenchymal cells. The mutually exclusive exons encode the second half of the third immunoglobulin domain, therefore alternative splicing will change the ligand binding specificity of FGFR2. This splicing event is tightly regulated and two cis-elements ISE-2 and ISAR located in the intron between exons IIIb and IIIc are required for cell-type specific inclusion of exon IIIb. It was shown that these two cis-elements, while separated by about 700 nt, can form a stem structure, which is essential for its regulatory function (Muh et al. 2002). Introducing mutations which disrupt the stem structure will lead to mis-regulation of exon IIIb, while complementary mutations which restore the stem structure will recover their

function (Muh et al. 2002). Furthermore, substitution of these two cis-elements by randomly selected sequences that also form a stem structure can totally recapitulate their regulatory function (Muh et al. 2002). Subsequent studies identified an ISE downstream of this secondary structure that was bound by ESRP1 and thus a model was proposed where the stem could position ESRP1 closer to exon IIIb to promote its splicing (Warzecha et al. 2009a). A similar stem-loop structure was described for the splicing of cassette exon (E11a) in ENAH, where a "RNA bridge" is formed between a conserved region immediately downstream of E11a to a similarly conserved region near a distal RBFOX cluster, which is located at least 1.8 kb downstream of E11a. It was proposed that the long range RNA structure recruits the distal bound RBFOX close to E11a and promotes its inclusion (Lovci et al. 2013).

Although splicing has been well characterized *in vitro* independent of transcription and polyadenylation, it is not a stand-alone process *in vivo*. In eukaryotes, there is an emerging recognition that pre-mRNA maturation, including RNA processing events such as 5' capping, splicing and 3' end formation, frequently occurs during the same time window as transcription in a higher order complex or "mRNA factory" (Luco et al. 2011; Saldi et al. 2016). Therefore, transcription and some other relevant processes including chromatin structure and histone modification can also regulate (alternative) splicing. Specifically, co-transcriptional splicing means the spliceosome starts to assemble on nascent pre-mRNA during transcription and/or introns are removed before transcription termination and transcript release. It was first suggested by the electron micrographs of early *Drosophila melanogaster* embryonic genes, which illustrated intron loop formation and intron removal on nascent pre-mRNA during transcription (Beyer and Osheim 1988). Later studies using Chromatin Immunoprecipitation (ChIP) assays revealed the recruitment of several core splicing factors including U1, U2 and U5 snRNPs on the nascent pre-mRNA, further supporting the co-transcriptional assembly of spliceosome (Görnemann et al. 2005; Lacadie and Rosbash 2005; Listerman et al. 2006). In addition, studies in human and yeast also confirmed excision of introns are complete before transcript release, suggesting co-transcriptional intron removal is widely present (Pandya-Jones and Black 2009). These studies

also revealed a general 5' to 3' order for co-transcriptional intron removal with some exceptions. A more recent study in yeast using two independent single-molecule nascent RNA-seq methods suggested that intron removal can occur immediately after RNAP II transcribes past the 3' ss (Carrillo Oesterreich et al.). Importantly, co-transcriptional splicing raises the possibility that transcription and splicing interact with each other and they are functionally coupled. Studies using *in vitro* transcription-splicing systems suggested RNA polymerase II (RNAP II) directed transcripts are spliced much more efficiently than pre-synthesized or T7-derived transcripts, suggesting RNAP II mediates the functional coupling of transcription to splicing (Das et al. 2006). In fact, further studies confirmed that splicing factors, including SR protein family members, U1 snRNP proteins and U2AF65 specifically associate with RNAP II, which facilitates spliceosome assembly on the nascent pre-mRNA as it emerges from the polymerase and promotes co-transcriptional splicing (Das et al. 2007; David et al. 2011; Gu et al. 2013). Therefore, RNAP II and potentially other components in the transcription machinery, serve as a landing platform to recruit splicing factors (Hsin and Manley 2012; Saldi et al. 2016). In addition to this spatial coupling between splicing factors and transcription machinery, transcription elongation rate can regulate splicing through kinetic coupling. For example, slower elongation rate led to higher inclusion of E33, a cassette exon in human fibronectin (FN) (de la Mata et al. 2003). Actually, a fast or slow RNAP II mutant has been shown to affect the splicing outcome on a genome-wide scale in human cells (Fong et al. 2014). The potential explanation is changes in the elongation rate affects the "window of opportunity" where positive and/or negative splicing factors recognize and bind to target sequences in nascent transcripts, especially for alternative exons with relatively weaker splice sites.

1.5 A growing number of RNA binding proteins are demonstrated to be multi-functional

RNA binding proteins (RBPs) not only play crucial roles in alternative splicing regulation, they are important for other post-transcriptional RNA processing events, including 5' capping, polyadenylation, mRNA transport, localization, mRNA stability, and translation efficiency. Emerging evidence suggests that many RBPs regulate multiple steps in RNA processing (Keene 2007; Sawicka et al. 2008; Mukherjee et al. 2011; Vanharanta et al. 2014). For example PAB1, the only cytoplasmic poly(A)-binding protein in yeast, is known to play a role in polyadenylation, nuclear export, mRNA turnover as well as translation initiation and termination through binding to mRNA poly(A) tails and protein factors that mediate these regulatory steps (Mangus et al. 2003). In higher eukaryotes such as human, the embryonic lethal abnormal vision like 1 (ELAVL1, also known as HuR) can regulate alternative splicing and promote mRNA stability and/or translation efficiency (Mukherjee et al. 2011; Simone and Keene 2013).

The subcellular localization of RBPs is one of many mechanisms that affect their regulatory functions. In some cases, a single protein has distinct functions depending on the cellular distribution under different environmental conditions. For example, HuR is predominantly nuclear under normal growth conditions but shuttles between nucleus and cytoplasm under certain stress conditions via a nuclear-cytoplasmic shuttling sequence (HuR Nucleocytoplasmic Shuttling sequence or "HNS") (Fan and Steitz 1998). HuR was initially reported primarily for its roles to promote mRNA stability and/or translation efficiency in the cytoplasm via binding to the AU- and U-rich elements (AREs) in the 3' untranslated regions (UTRs) of target transcripts, many of which are involved in various cellular processes, including cell proliferation, survival, differentiation and immune response (Brennan and Steitz 2001; Mukherjee et al. 2009; Abdelmohsen and Gorospe 2010; Srikantan and Gorospe 2012). However, subsequent reports showed that HuR is also involved in regulation of alternative splicing for several mRNA transcripts in the nucleus (Izquierdo 2008; Akaike et al. 2014; Chang et al. 2014; Zhao et al. 2014).

Consistent with roles for HuR in splicing regulation, transcriptome-wide mapping of HuR binding sites revealed numerous intronic binding sites for HuR (Lebedeva et al. 2011; Mukherjee et al. 2011). Another well-studied RBP, PTBP1 is also predominantly nuclear and shuttles between nucleus and cytoplasm (Pérez et al. 1997; Li and Yen 2002). While PTBP1 regulates alternative splicing and 3' end processing in the nucleus, it also plays a role in mRNA localization, mRNA stability and viral translation in the cytoplasm (Sawicka et al. 2008). Interestingly, the translocation from nucleus to cytoplasm of PTBP1 is regulated by phosphorylation within the NLS region (Xie et al. 2003). In some other cases, alternative splicing variants of a single gene encode multiple protein isoforms with different subcellular localization and function. A well-known example is Quaking (QKI), a RNA binding protein belonging to the signal transduction and activation of RNA (STAR) family, which generates three major isoforms (Hardy et al. 1996). QKI-5 is mainly nuclear and known for its role in regulation of alternative splicing, including a recent study suggesting a role for QKI-5 in the formation of circRNAs (Hall et al. 2013; Conn et al. 2015a). In contrast, QKI-6 is primarily cytoplasmic and reported to regulate mRNA stability and act as a translational repressor (Saccomanno et al. 1999; Lakiza et al. 2005; Doukhanine et al. 2010). QKI-7 is also predominantly cytoplasmic and was recently identified to promote poly-A tail extension in the cytoplasm (Yamagishi et al. 2016). Additionally, both QKI-5 and QKI-6 are involved in miRNA processing (Wang et al. 2010; Wang et al. 2013). Rbfox1 also generates both cytoplasmic and nuclear isoforms that are encoded by transcript variants with or without exon 19 respectively. The nuclear isoform of Rbfox1 regulates alternative splicing via binding to the intronic regions of target transcripts, many of which are important for neuronal development and maturation (Fogel et al. 2012; Weyn-Vanhentenryck et al. 2014). Furthermore, mutations in Rbfox1 are implicated in a range of neurodevelopmental diseases, including autism spectrum disorder (Bhalla et al. 2004; Martin et al. 2007; Bill et al. 2013; Lal et al. 2013). A recent study also dissected the role of cytoplasmic Rbfox1 in promoting mRNA stability and translational efficiency via binding to the 3' UTR region (Lee et al. 2016). Notably, target transcripts regulated by cytoplasmic Rbfox1, although different from those regulated by nuclear Rbfox1, are also

enriched for genes involved in synaptic function, calcium signaling and autism, suggesting both isoforms of Rbfox1 coordinately regulate neuronal development in a distinct but coherent manner (Lee et al. 2016). Importantly, enforced translocation of the nuclear Rbfox1 isoform to the cytoplasm by deletion of the NLS, enabled it to regulate stability and translation of target mRNAs indistinguishably from the cytoplasmic Rbfox1, indicating that the differential localizations of the proteins, rather than the differences in the sequence, lead to the different functions. Moreover, the splicing of exon 19 is regulated by nuclear Rbfox1 upon neuronal depolarization which leads to a switch from cytoplasmic to nuclear isoform, adding another layer of complexity and coordination for Rbfox1 regulated network (Lee et al. 2009). Hence, it is increasingly recognized that many other RBPs that have well characterized roles in one specific step of post-transcriptional regulation may have other unrecognized functions in RNA processing. Emerging evidence suggests that integrated post-transcriptional regulation by a specific RBP can function to shape biologically coherent post-transcriptional regulons (Keene 2007).

1.6 Epithelial to mesenchymal transition (EMT) and its regulation

The epithelial to mesenchymal transition (EMT) is a process by which epithelial cells transdifferentiate into mesenchymal cells which involves extensive changes at cellular and molecular levels. Epithelial cells, often with a cobble-stone like morphology, are non-motile with a well-defined apico-basal polarity and are tightly connected to each other laterally by adherens and tight junctions. In contrast, mesenchymal cells usually have an elongated and spindle-like morphology, are motile and invasive with a front-back polarity and a loss of adherens and tight junctions(Ye and Weinberg 2015). This highly conserved biological process, however, is reversible and together with the reverse process, mesenchymal to epithelial transition (MET), plays a fundamental role in the formation of the body plan and differentiation of multiple tissues and organs where several rounds of EMT and MET occur (Lim and Thiery 2012; Nieto and Cano 2012; Nieto et al. 2016). The developmental EMT is a well-coordinated process where a series of events take place in a defined order (Shook and Keller 2003; Nistico et al. 2012). Among those, there are several key steps. First, the regions where the EMT will occur are pre-specified and cells that will undergo EMT are moved to those specified regions. Second, interactions between the epithelial cells and the basement membrane are disrupted. Third, the transitioning cells detach from the epithelial tissue. Fourth, the migratory cells develop mesenchymal phenotypes, which feature altered cytoskeletal organization and cellular metabolism (Nistico et al. 2012). Based on the sequential rounds, the developmental EMTs are classified as primary, secondary, and tertiary EMT. Primary EMT occurs early in embryonic development in processes like gastrulation and the neural crest delamination (Thiery et al. 2009). During gastrulation, a developmental stage where a single layer of epithelial cells called epiblast forms three primary germ layers, the epithelial cells in the primitive streak emerging from the posterior epiblast undergo EMT and delaminate to give rise to mesendermal cells which will form mesoderm and definitive endoderm (Thiery et al. 2009). Further studies revealed a critical role of signaling pathways in the induction of the EMT during gastrulation, including canonical Wnts as well as

fibroblast growth factors (FGFs) (Tam and Behringer 1997; Liu et al. 1999; Skromne and Stern 2001). Subsequent studies identified TGF-β superfamily proteins to mediate the action of Wnts (Kalluri and Weinberg 2009). Transcription factors, such as Snail1, are crucial to orchestrate changes in the transcription program during this EMT (Nieto 2002; Kalluri and Weinberg 2009). After gastrulation, the epithelial cells in the neuroectoderm undergo EMT to give rise to migratory neural crest cells (Duband and Thiery 1982). Those cells then dissociate from the neural folds and migrate to different parts of the embryo before differentiating into various tissues (Thiery et al. 2009). The EMT during neural crest delamination is triggered by signaling pathways similar to those during the EMT associated with gastrulation and transcription factors, such as Sox, Snail1 and Snail2, are responsible for the underlying transcriptional regulatory network (Knecht and Bronner-Fraser 2002; Sauka-Spengler and Bronner-Fraser 2008). Although EMT can be driven by a variety of factors and the signaling transduction cascades are also context dependent, a hallmark of EMT is down-regulation of the epithelial protein CDH1 (E-cadherin), which promotes disassembly of AJs. In addition, mesenchymal proteins, such as CDH2 (N-cadherin) and VIM (vimentin), are also well described standard molecular markers that are up-regulated during EMT.

EMT program can also be activated in the context of human diseases. These pathological EMTs have been the subject of a significant amount of studies as they were proposed to be a major mechanism for tissue fibrosis and cancer metastasis. However, more recent studies have challenged a direct contribution of EMT in some of these disease contexts. Progressive kidney fibrosis is an example where the role of EMT is under debate (Kriz et al. 2011). Initial study using a cell tracking technique proposed that renal tubular epithelial cells can give rise to interstitial fibroblasts through EMT (Iwano et al. 2002). However, subsequent studies using lineage tracing identified either no evidence or evidence of only very few myofibroblasts of epithelial origin (Humphreys et al. 2010; Li et al. 2010; LeBleu et al. 2013). Hence, models proposing a direct role of EMT in kidney fibrosis have become controversial (Kriz et al. 2011). A recent study using a tamoxifen-induced mouse model suggested specific expression of Snail in renal epithelial cells induce fibrosis in adult mice. Since Snail is transcription factor that can

induce EMT *in vitro* in a cell-based model, the study suggested a positive association of EMT *in vivo* in promoting fibrosis (Boutet et al. 2006). However, whether there is a full, complete EMT program in the tubular epithelial cells caused by Snail was not fully addressed. A follow-up study that compared the WT kidney to the mutant where Snail is specifically knockout in the epithelial cell population in a model of UUO-induced fibrosis (UUO stands for unilateral ureteral obstruction, a standard model for progressive renal fibrosis) showed that the renal epithelial cells undergo a "partial EMT" meditated by Snail reactivation, confirmed by the corresponding changes of typical epithelial and mesenchymal markers, including E-cadherin, Vimentin, and an apparent loss of apicobasal polarity of those cells (Grande et al. 2015). Although these epithelial cells didn't delaminate from the tubular basement membrane and migrate to the interstitium to directly contribute to interstitial fibrosis, the partial EMT is proposed to generate signals to cells in the interstitium, including TGF-β and NF-κB to promote fibrogenesis and inflammation (Grande et al. 2015). Similarly, a direct contribution of EMT in cancer metastasis as well as the generation of cancer stem cells (CSCs) or tumor initiating cells (TIC) that are predictive of a poor response to cancer therapies, is also under debate and active research (Singh and Settleman 2010; Scheel and Weinberg 2012; Tiwari et al. 2012; Ye and Weinberg 2015). Studies from *in vitro* data and mouse models of human cancers suggested that the development of cancer metastasis includes an EMT-like process since tumor cells derived from epithelial origin need to infiltrate the basement membrane and intravasate into blood or lymph vessels to be able to travel to distal sites (Scheel and Weinberg 2012; Yu et al. 2013; Ye and Weinberg 2015). For solitary tumor cells to proceed through these steps requires the tumor cell to break up cell-cell junctions with neighboring cells and become mobile in a manner reminiscent of EMT. It was also suggested that an MET-like process can explain the colonization of tumor initiating cells to form the secondary tumor in the distal sties (Scheel and Weinberg 2012). However, due to the technical difficulties in monitoring a single tumor cell *in vivo* and capturing the transient and highly plastic cell status in epithelial or mesenchymal identities during the process of metastasis, there was no direct evidence to convincingly support EMT as a requirement for metastasis. However, two recent

papers using mouse models of mammary tumors as well as pancreatic cancer argued that EMT is dispensable for metastasis but contributes to chemoresistance of tumor cells (Fischer et al. 2015; Zheng et al. 2015). One of the studies took advantage of cell linage tracing technique and showed only a small fraction of primary tumor cells that had undergone EMT (Fischer et al. 2015). While both epithelial and mesenchymal cancer cells can invade the circulating system to derive secondary tumors, it was proposed that the majority of metastatic tumors are initiated by epithelial cells from primary tumors without undergoing EMT (Fischer et al. 2015). They also showed that repression of mesenchymal transcription factors *ZEB1* and *ZEB2* didn't impair metastasis of the mouse mammary tumor to the lung (Fischer et al. 2015). This observation is consistent with a study where tissue specific depletion of Snail or Twist transcription factor didn't affect tumor metastasis in the liver, lung and spleen (Zheng et al. 2015). Importantly, both studies identified EMT as one of the mechanisms conferring chemresistance of tumor metastasis, which suggest a potential EMT-targeting strategy for cancer metastasis treatment (Fischer et al. 2015; Zheng et al. 2015). While these studies provide us with valuable information in how EMT may contribute to cancer metastasis or chemoresistance, their strategy using inhibition of individual EMT transcription factor to block EMT *in vivo* has its own limitations. It remains challenging to study EMT and its underlying mechanism at a genomic level using those mouse models, but further studies on the role of EMT *in vivo* are clearly needed.

To date, molecular analysis of EMT has been done mostly using a small number of *in vitro* models in a very limited number of cell lines in response to several EMT drivers (Taube et al. 2010; Shapiro et al. 2011; Thomson et al. 2011). Among many tested epithelial cell lines, only a few can be induced to undergo EMT. A classical model has been established in a human mammary epithelial (HMLE) cell line, where different drivers such as TGF-β and certain mesenchymal transcription factors, including Snail and Twist, were used to induce EMT *in vitro* (Mani et al. 2008). In a model using human H358 cells, doxycycline inducible expression of Snail1 or ZEB1 as well as TGF-β treatment can induce EMT within 7 days (Thomson et al. 2011). In some cases, a combination of different drivers is preferred to induce a robust and fast EMT *in*

vitro. From these *in vitro* studies, global changes in total gene expression have been characterized during EMT, including those of standard EMT markers (Taube et al. 2010; Thomson et al. 2011). While the relevance of these cell models to *in vivo* EMT is debatable, it provides us with a system to study EMT at molecular level to a greater extent.

2) Chapter 2: Determination of alternative splicing regulatory network during EMT and identification of key splicing regulators

2.1 Background

Most previous studies have mainly focused on signaling pathways and transcriptional regulation of the molecular mechanisms underlying EMT. Until recently, the role of post-transcriptional gene regulation in EMT has been relatively unexplored. We previously identified the Epithelial Splicing Regulatory Proteins 1 and 2 (ESRP1 and ESRP2) as cell-type-specific regulators of an epithelial specific splicing program that is reverted during EMT (Warzecha et al. 2009b; Warzecha et al. 2010). We showed that ESRP1 is transcriptionally inactivated during EMT and studies by other groups have also confirmed that ESRP1 is among the most down-regulated genes in multiple EMT model systems (Onder et al. 2008; Taube et al. 2010; Warzecha et al. 2010; Shapiro et al. 2011; Thomson et al. 2011; Horiguchi et al. 2012; Braeutigam et al. 2014). We also determined that the ESRPs bind to UGG-rich motifs *in vitro* using systematic evolution of ligands by exponential enrichment coupled with high throughput sequencing (SELEX-Seq) (Dittmar et al. 2012). These UGG-rich motifs are enriched in the vicinity of ERSP-regulated exons, consistent with the “RNA map” in which the ESRPs regulate splicing in a position-dependent manner: binding in the downstream intron promotes exon inclusion while binding within or upstream of an exon promotes exon skipping (Warzecha et al. 2010; Dittmar et al. 2012). Conversely, RBFOX2 generally promotes mesenchymal splicing patterns for many gene transcripts that undergo AS during EMT, although in some cases it promotes epithelial splicing (Lapuk et al. 2010; Dittmar et al. 2012; Venables et al. 2013; Braeutigam et al. 2014). Moreover, it was shown that depletion of RBFOX2 impaired the invasiveness of cells that underwent EMT (Shapiro et al. 2011; Braeutigam et al. 2014). Interestingly, we showed that for some EMT-associated AS events, there is a combinatorial regulation between the ESRPs and RBFOX2

(Dittmar et al. 2012). In addition to the ESRPs and RBFOX2, several other RBPs, such as PTBP1 and MNBL have also been implicated in AS regulation during EMT, although their functions are less well-defined (Shapiro et al. 2011; Venables et al. 2013). Additionally, the role of post-transcriptional regulation in EMT has also been shown to involve important epithelial-specific miR-200 family that targets ZEB1 and down-regulation of these micro-RNAs are similarly observed during EMT (Gregory et al. 2008; Park et al. 2008).

To date, there has been only one published study that investigated genome-wide alterations in AS that accompany EMT using RNA-Seq (Shapiro et al. 2011). While this important study identified large scale alterations in AS, many regulated events likely eluded detection due to the limited sequencing depth by current standards and absence of biological replicates. In addition, the combinatorial regulation of AS by different splicing regulators at distinct time points during EMT is still poorly resolved, and contributions by novel EMT regulatory proteins remain to be uncovered. To more comprehensively determine the AS network during EMT, we used a robust inducible *in vitro* EMT model with high sequencing depth and multiple replicates to conduct an in-depth investigation of AS, including a temporal analysis at multiple time points during EMT induction. Although we previously determined genome-wide AS targets of the ESRPs, the degree to which loss of ESRP expression contributes to splicing switches during EMT has not been characterized. We therefore carried out a parallel RNA-Seq analysis of ESRP targets in the same cell line and determined that the ESRPs are major regulators for AS during EMT, but also that other splicing factors contribute to EMT-associated AS switches. For example, we defined novel roles for RBM47 and QKI in combinatorial regulation of AS during EMT. Together, these investigations help lay the groundwork for further studies to define the functional consequences of AS regulation during EMT and how they impact this essential developmental process and relevant disease pathophysiology.

2.2 Results

2.2.1 Comprehensive determination of changes in AS during EMT

In order to study the AS regulatory network during EMT more comprehensively, we adapted a previously described EMT model using the human H358 epithelial non-small cell lung cancer (NSCLC) cell line (Thomson et al. 2011). We generated a H358 clone stably expressing a doxycycline (Dox)-inducible cDNA encoding a ZEB1-mCherry fusion protein. Over a 7-day time course following Dox treatment (1ug/ml), we validated the gradual decrease for the epithelial cell marker CDH1 (E-cadherin) and increase of the mesenchymal cell marker VIM (vimentin) at the protein level, confirming that the cells underwent progressive EMT during the time course (Fig. 2.1A, B). Consistently, expression levels of the ESRPs were also progressively down-regulated at both protein and mRNA level during the time course (Fig. 2.1B, C). To study dynamic changes in splicing during EMT, we isolated total RNA at each day of the EMT time course following Dox treatment and a no Dox-treated control in biological triplicates (Fig. 2.1A). We performed strand-specific 100 bp paired-end RNA-Seq and obtained between 40 million to 70 million read pairs per replicate with a total 1.3 billion read pairs (Table 2-1). The use of biological replicates and high sequencing depth, together with the time course analysis enabled the identification of a comprehensive program of AS switches associated with EMT. To fully characterize the AS program during the process of EMT, our collaborators Dr. Yi Xing and his lab used the rMATS computational pipeline to identify splicing changes comparing each time point to No Dox controls and identified numerous cassette exons (also referred to as “skipped exons” or SE) that show significant splicing changes in at least one rMATS comparison (Table A2-1). We applied a network analysis to identify clusters of SE events that have distinct temporal patterns of splicing changes across the EMT time course (Fig. 2.2A). The two largest clusters consisted of SE that showed either a graded increase (Cluster 1) or decrease (Cluster 2) in splicing across the EMT time course. However, we also identified other clusters representing less straightforward temporal patterns of splicing changes at early or late time points illustrating the complexity of temporally

dynamic regulation of AS during EMT. We identified a total of 1077 significant AS events (|ΔPSI|>5%; FDR <5%) at Day 7 compared to controls, the majority of which were cassette exons (670/1077) (Fig. 2.2B and Table A2-2) (Shen et al. 2014). Gene ontology enrichment analysis using the DAVID functional annotation tool (https://david.ncifcrf.gov/home.jsp) showed that EMT-associated AS genes are involved in biological processes such as RNA splicing, regulation of small GTPase mediated signal transduction, and cytoskeleton organization, which are relevant to EMT (Fig. 2.2C). Using gene expression profiling, we detected more than 2000 genes with an expression change at mRNA level by at least 2-fold after 7 days of Dox treatment compared to controls (Table A2-3). The RNA-Seq results also confirmed the down-regulation of epithelial markers CDH1 and CLDN4 and up-regulation of mesenchymal markers VIM and FN1 (fibronectin 1). Notably, only 74 target transcripts with splicing changes during EMT showed an expression change of at least 2-fold in the whole transcript level, which is consistent with the notion that genes under post-transcriptional regulation during cellular transitions are generally not regulated at transcription level. These findings reinforce the concept that transcription and AS are separate layers of gene expression regulation that are integrated during important developmental and physiological transitions.

Table 2-1 Sequencing statistics for the EMT experiment.

Sample	Total read pairs	Mapped pairs	%	Genome reads	%	Junction reads	%	% exon	% intron	% intergenic
No dox-1	72,116,801	54,315,032	75.3%	71,658,960	49.7%	36,971,104	25.6%	65.9%	4.8%	4.6%
No dox-2	63,670,441	47,501,976	74.6%	62,955,271	49.4%	32,048,681	25.2%	65.2%	4.7%	4.6%
No dox-3	59,743,765	45,728,417	76.5%	60,461,232	50.6%	30,995,602	25.9%	66.9%	4.9%	4.7%
Day 1-1	61,110,098	45,018,366	73.7%	59,575,815	48.7%	30,460,917	24.9%	64.5%	4.7%	4.5%
Day 1-2	50,969,965	36,501,262	71.6%	48,398,540	47.5%	24,603,984	24.1%	62.6%	4.5%	4.5%
Day 1-3	46,400,841	33,946,390	73.2%	44,996,721	48.5%	22,896,059	24.7%	64.0%	4.8%	4.4%
Day 2-1	61,657,740	44,790,006	72.6%	60,441,967	49.0%	29,138,045	23.6%	62.8%	5.2%	4.7%
Day 2-2	60,834,809	43,794,339	72.0%	59,008,752	48.5%	28,579,926	23.5%	62.2%	5.2%	4.6%
Day 2-3	52,400,786	37,975,296	72.5%	51,385,265	49.0%	24,565,327	23.4%	62.5%	5.2%	4.7%
Day 3-1	46,321,820	34,118,857	73.7%	45,763,495	49.4%	22,474,219	24.3%	63.7%	5.2%	4.7%
Day 3-2	51,267,703	35,039,382	68.3%	46,737,929	45.6%	23,340,835	22.8%	59.9%	4.1%	4.4%
Day 3-3	53,184,696	39,101,259	73.5%	52,555,999	49.4%	25,646,519	24.1%	63.8%	5.0%	4.7%
Day 4-1	70,251,123	52,370,822	74.5%	70,994,400	50.5%	33,747,244	24.0%	63.9%	5.8%	4.9%
Day 4-2	51,587,561	37,415,656	72.5%	50,658,929	49.1%	24,172,383	23.4%	62.3%	5.5%	4.7%
Day 4-3	52,883,952	38,150,316	72.1%	51,554,820	48.7%	24,745,812	23.4%	62.0%	5.5%	4.7%
Day 5-1	53,512,878	38,455,251	71.9%	51,572,321	48.2%	25,338,181	23.7%	64.3%	2.9%	4.6%
Day 5-2	51,041,512	37,482,560	73.4%	49,894,282	48.9%	25,070,838	24.6%	64.9%	4.0%	4.5%
Day 5-3	50,343,790	36,678,264	72.9%	48,979,183	48.6%	24,377,345	24.2%	64.5%	3.9%	4.5%
Day 6-1	49,043,336	35,410,615	72.2%	48,292,197	49.2%	22,529,033	23.0%	63.2%	4.4%	4.7%
Day 6-2	63,724,659	45,913,753	72.1%	62,102,543	48.7%	29,724,963	23.3%	63.0%	4.5%	4.6%
Day 6-3	47,301,339	33,821,298	71.5%	45,819,434	48.4%	21,823,162	23.1%	63.2%	3.7%	4.7%
Day 7-1	53,632,960	39,854,597	74.3%	54,104,633	50.4%	25,604,561	23.9%	63.9%	5.8%	4.6%
Day 7-2	55,587,194	41,736,520	75.1%	56,886,659	51.2%	26,586,381	23.9%	64.5%	5.9%	4.7%
Day 7-3	57,773,955	43,201,972	74.8%	58,622,162	50.7%	27,781,782	24.0%	64.4%	5.8%	4.6%

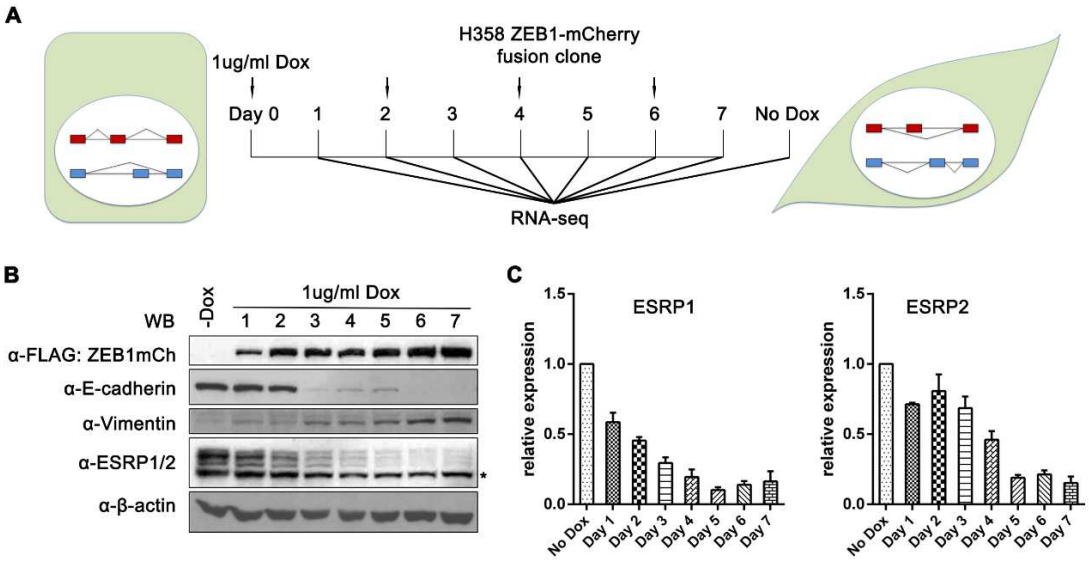


Figure 2-1 Establishment of an in vitro EMT model in human epithelial H358 cells.

(A) Schematic of Doxycycline-induced EMT in H358 ZEB1-mCherry stable clone over a seven-day time course. 1ug/ml Doxycycline was added to the cells on Day 0, 2, 4, and 6 to induce and maintain the expression of ZEB1.

(B) Validation of ZEB1 induction upon Dox treatment and decreased expression of epithelial marker (E-cadherin, or CDH1), increased expression of mesenchymal marker (vimentin, or VIM) as well as decreased expression of ESRP1/2 during the time course by Western blot. β-actin is used as a loading control. Multiple bands in the anti-ESRP1/2 blot represent different splicing isoforms for ESRP1. The asterisk indicates a non-specific cross-reacting band.

(C) Quantitative RT-PCR (qRT-PCR) analysis of *ESRP1* and *ESRP2* showed decreased transcript level during the time course (in biological triplicate). For all qRT-PCR analyses, graphs are shown as mean with error bars representing standard deviation.

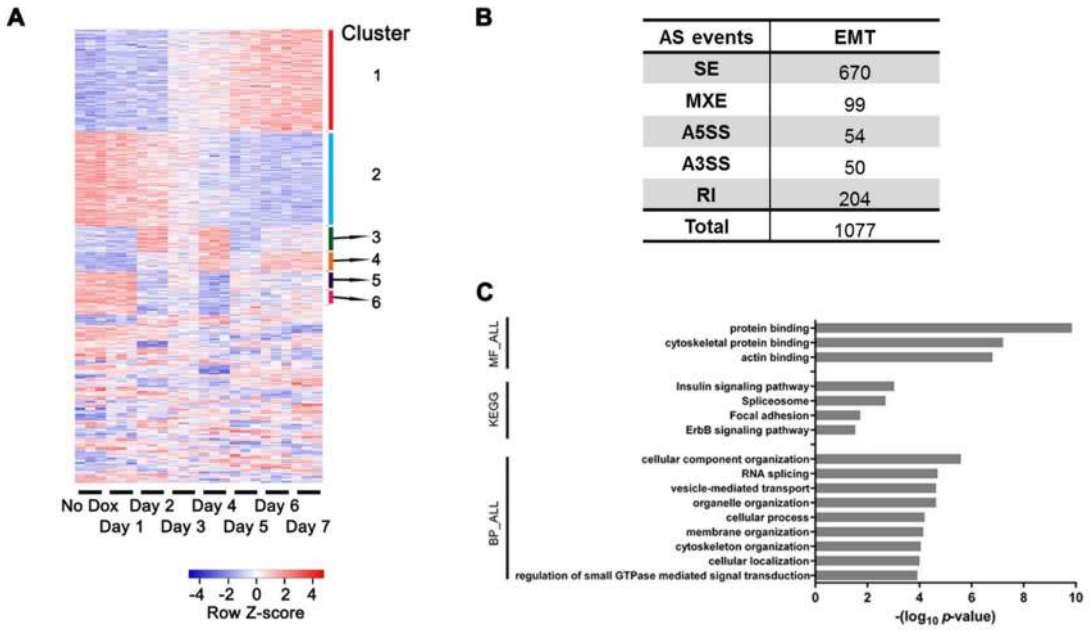


Figure 2-2 Comprehensive determination of changes in AS during EMT.

(A) Cluster analysis for all identified SE that significantly changed in at least one rMATS comparison between each time point to controls revealed two major clusters of SE events. Clusters with at least five SE events are shown (complete clusters and more details in Table A2-1).

(B) Summary of different types of significant AS events identified from Day 7/No Dox comparison. SE: skipped exon; MXE: mutually exclusive exon; A5SS: alternative 5' splice site; A3SS: alternative 3' splice site; RI: retained intron.

(C) Gene ontology analysis for EMT-associated SE events showed an enrichment of EMT-relevant processes and pathways.

2.2.2 Determination of changes in AS regulated by the ESRPs

Our previous studies showed that the ESRPs are important for regulation of many splicing events that switched in a different EMT model (Warzecha et al. 2010). However, the extent to which the ESRPs contribute to the AS regulatory network in EMT required further investigation. We therefore used lentiviral short hairpin RNAs (shRNAs) to deplete both ESRP1 and ESRP2 in H358 cells (Fig. 2.3A-C). We generated cDNA libraries with total RNA isolated from ESRP1/2 knockdown and control knockdown in biological triplicate, conducted strand-specific 100 bp paired-end RNA-Seq, and obtained about 50 million read pairs per replicate and 320 million read pairs in total (Table 2-2). Using the same rMATS pipeline, we identified 235 significant AS events (|ΔPSI|>5%; FDR <5%) regulated by the ESRPs including 179 cassette exons, some of which were previously reported ESRP targets in other cell lines (Fig. 2-4A, B and Table A2-4) (Warzecha et al. 2009a; Warzecha et al. 2010; Dittmar et al. 2012). Gene ontology enrichment analysis revealed that ESRP-regulated AS transcripts are enriched for genes involved in cell junction organization and cytoskeleton organization, consistent with our previous study (Fig. 2-4C) (Warzecha et al. 2010; Dittmar et al. 2012). Fewer than 50 genes showed an expression change of at least 2-fold upon ESRP1/2 knockdown (29 down-regulated and 18 up-regulated) (Table A2-5). Consistent with our previous study, there was an enrichment of ESRP1 binding motifs downstream of ESRP-enhanced exons and upstream of ESRP-silenced exons in H358 cells (Fig. 2-4D) (Warzecha et al. 2010). We also observed an enrichment of RBFOX2 binding motif upstream of ESRP-enhanced exons and downstream of ESRP-silenced exons, further supporting that the combinatorial regulation of splicing between the ESRPs and RBFOX2 is more often antagonistic than cooperative.

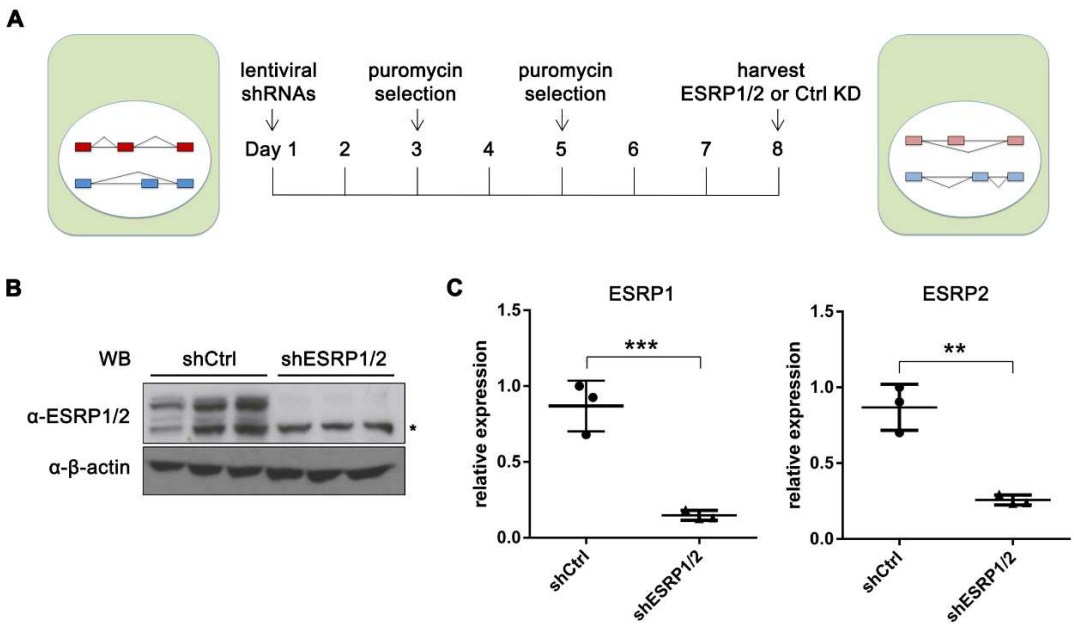
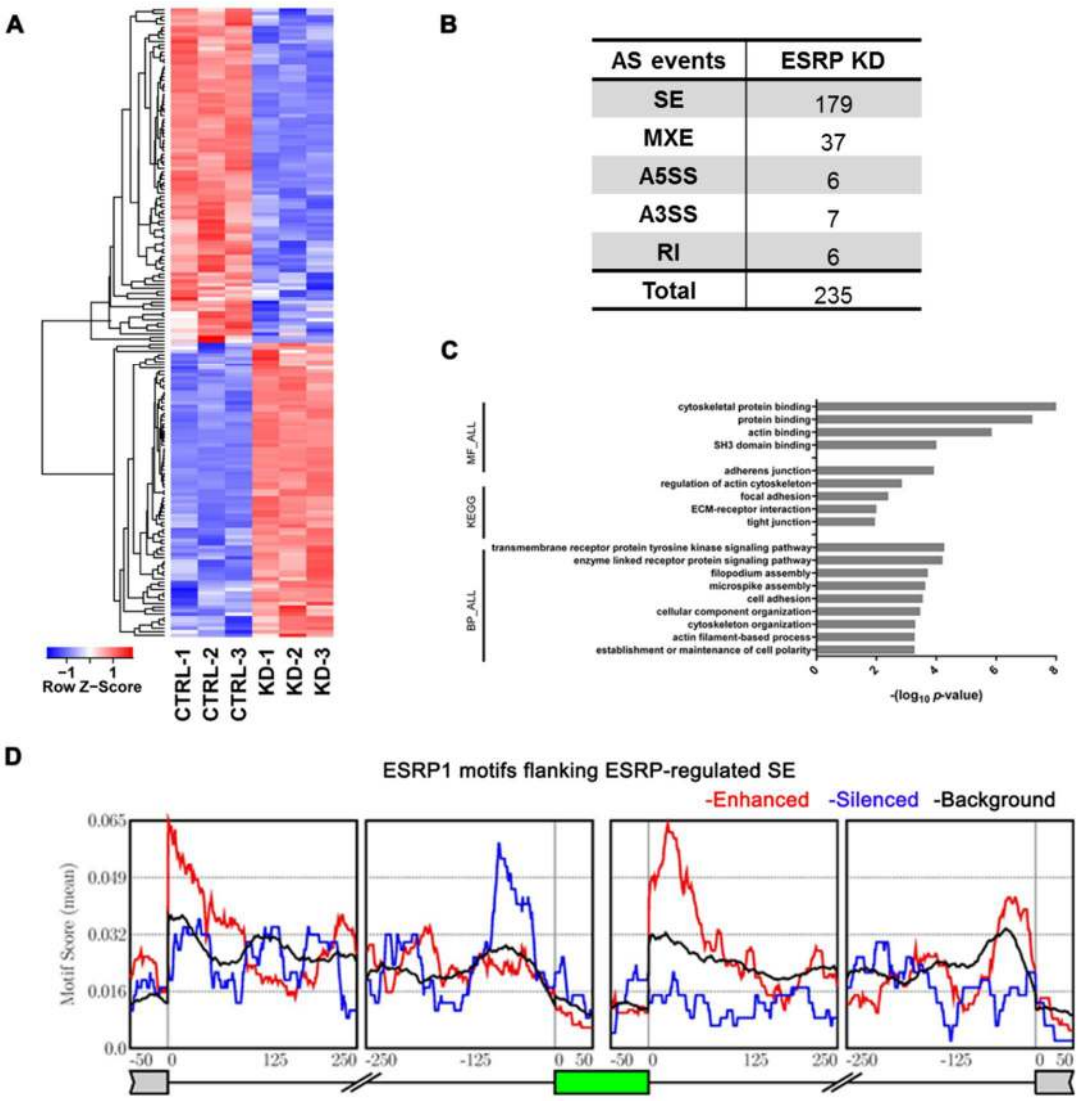


Figure 2-3 Knockdown of ESRP1 and ESRP2 in H358 cells using shRNAs.

(A) Schematic of ESRP1/2 knockdown experiment in H358 cells using lentiviral shRNAs. After infection, cells were selected with 5ug/ml puromycin and 20mg/ml blasticidin. At the end of selection, mock transfection control cells were all dead.

(B) Western blot validation of ESRP1/2 protein knockdown.

(C) Validation of *ESRP1*/2 mRNA knockdown by qRT-PCR (in biological triplicate). A one-tailed unpaired *t*-test was used to calculate the p-value. The same statistical test was used in all subsequent qRT-PCR analyses in subsequent figures.



AS events	ESRP KD
SE	179
MXE	37
A5SS	6
A3SS	7
RI	6
Total	235

Figure 2-4 Identification of global changes in AS after ESRP1/2 depletion.

(A) Heat map for ESRP-regulated cassette exons showed high consistency across replicates.

(B) Summary of different types of significant AS events identified upon ESRP1/2 depletion in H358 cells.

(C) Gene ontology analysis for ESRP-regulated SE events showed an enrichment of EMT-relevant processes and pathways.

(D) ESRP1 binding motifs are enriched upstream of ESRP-silenced exons and downstream of ESRP-enhanced exons.

Table 2-2 Sequencing statistics for the ESRP1/2 knockdown experiment.

Sample	Total read pairs	Mapped pairs	%	Genome reads	%	Junction reads	%
CTRL-1	41,442,942	30,214,803	72.9%	40,587,147	49.0%	19,842,459	23.9%
CTRL-2	67,887,074	49,831,244	73.4%	67,080,914	49.4%	32,581,574	24.0%
CTRL-3	67,933,853	49,551,067	72.9%	66,386,087	48.9%	32,716,047	24.1%
KD-1	40,653,351	29,902,177	73.6%	39,921,169	49.1%	19,883,185	24.5%
KD-2	57,451,860	42,855,527	74.6%	57,205,354	49.8%	28,505,700	24.8%
KD-3	51,094,649	37,614,990	73.6%	50,920,409	49.8%	24,309,571	23.8%
Total	**326,463,729**	**239,969,808**	**73.5%**	**322,101,080**	**49.3%**	**157,838,536**	**24.2%**

2.2.3 ESRP1 and ESRP2 are major regulators of epithelial splicing patterns during EMT

To define the degree to which ESRP down-regulation contributes to the global changes in AS during EMT, we analyzed the overlap between ESRP-regulated AS events and EMT-associated AS events (Fig. 2-5 and Table A2-6). Of the 670 cassette exons that showed splicing switches during EMT, a highly significant number of events (122/670, p=8.55E-117) are regulated by the ESRPs based on the rMATS analysis. As predicted, for all the 122 events, the ESRPs promote the epithelial splicing patterns that are abrogated during EMT. A recent study proposed that HNRNPM is also a key splicing factor involved in AS regulation during EMT (Xu et al. 2014). However, when we compared the 324 HNRNPM-regulated cassette exons identified by rMATS using published RNA-Seq data to the 670 EMT-associated events identified here, we found fewer (38/670, p=1.50E-10) shared events. As I will discuss in later sections, although RBM47 and QKI both are splicing regulators during EMT, the number of AS events regulated by either factor during EMT is less significant than that of the ESRPs. In the previous study of EMT-associated splicing changes Gertler and colleagues implicated RBFOX2 and PTBP1 as regulators of EMT splicing switches along with the ESRPs. However, they noted a statistically greater overlap with ESRP-regulated targets identified using splicing sensitive microarrays (p=9.27E-16) than with

RBFOX2 or PTBP1-regulated events defined in CLIP-Seq studies (p=8.58E-5 and p=0.0013, respectively) (Shapiro et al. 2011). Taken together, these observations suggest that ESRP1 and ESRP2 are major regulators of AS during EMT, exclusively promoting the epithelial splicing patterns. However, the fact that we see some EMT-associated AS events are not significantly regulated by the ESRPs prompted us to pursue roles for additional splicing factors that contribute to AS programs during EMT.

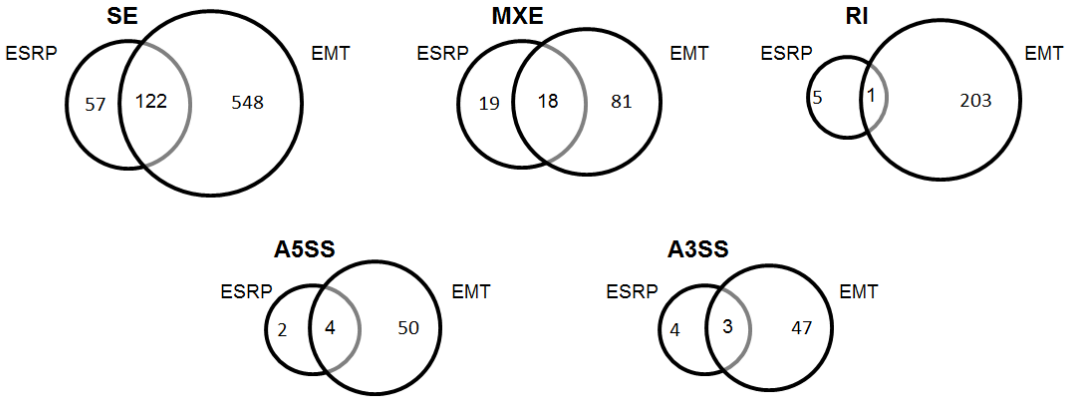


Figure 2-5 Venn diagram of the overlap between EMT-associated AS events and those regulated upon ESRP1/2 knockdown based on rMATS analysis.

2.2.4 Validation of cassette exons regulated by the ESRPs or EMT

To validate these RNA-Seq results, we used semi-quantitative RT-PCR for some representative cassette exons whose inclusion or exclusion was promoted by EMT or ESRP1/2 knockdown.

From 122 cassette exons that are ESRP-regulated, EMT-associated as predicted by rMATS analysis, seven were tested and five were validated both in EMT and upon ESRP depletion (Fig. 2-6). In the case of *CARD19*, the splicing change was validated in EMT and a similar change was also observed upon ESRP depletion, but did not reach statistical significance. For *RPS24* the splicing change was not validated in either setting. Importantly, knockdown of

ESRP often induced splicing changes that were less than those observed during EMT (e.g. *PLOD2*, *INF2* and *NFYA*), despite similar levels of reduction for ESRP total transcripts in both settings, indicating that ESRP depletion alone often can't fully account for the degree of splicing changes we observed during EMT.

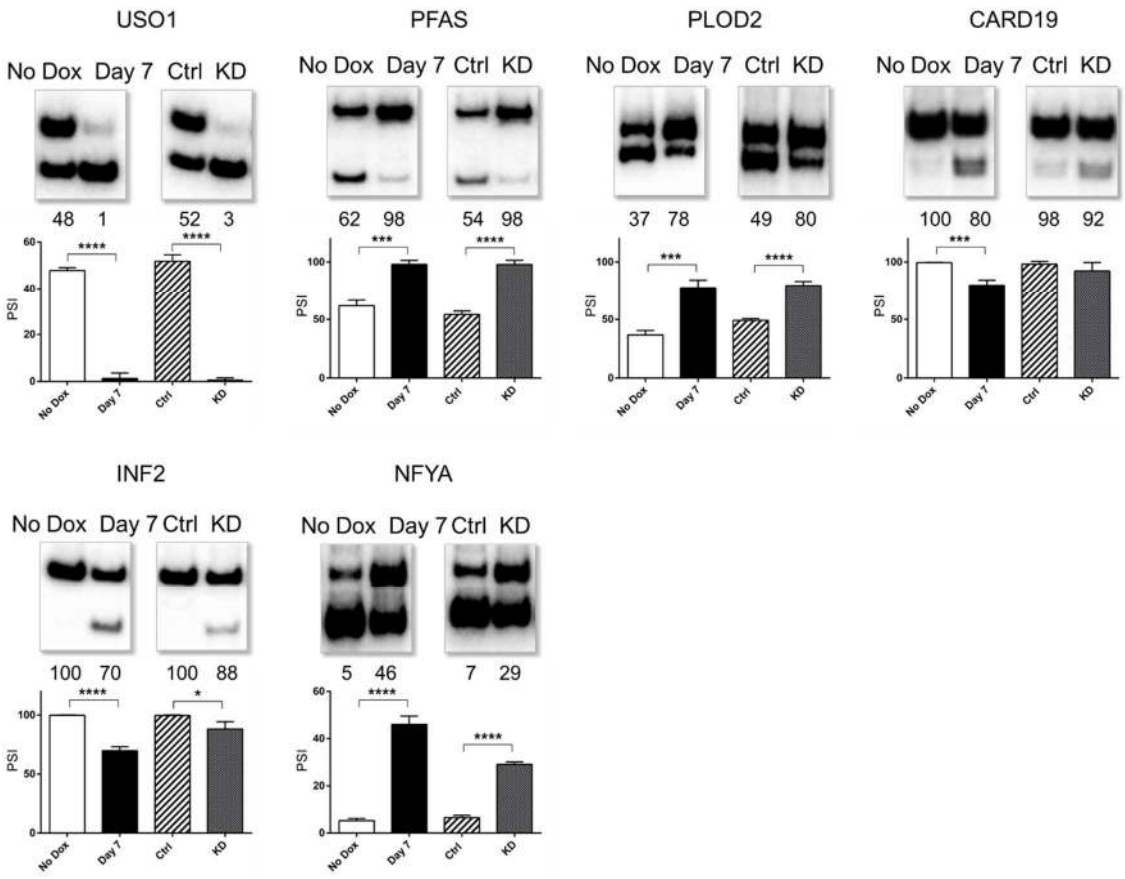


Figure 2-6 RT-PCR validation of exons predicted to switch during EMT and with ESRP1/2 knockdown.

For each target a representative gel is shown on top and the graph with the averaged PSI (Avg. PSI) and standard deviation (S.D.) from biological triplicate is shown on bottom. A one-tailed unpaired *t*-test was used to calculate the p-value for each comparison. The same statistical test was used in all subsequent RT-PCR analyses in subsequent figures.

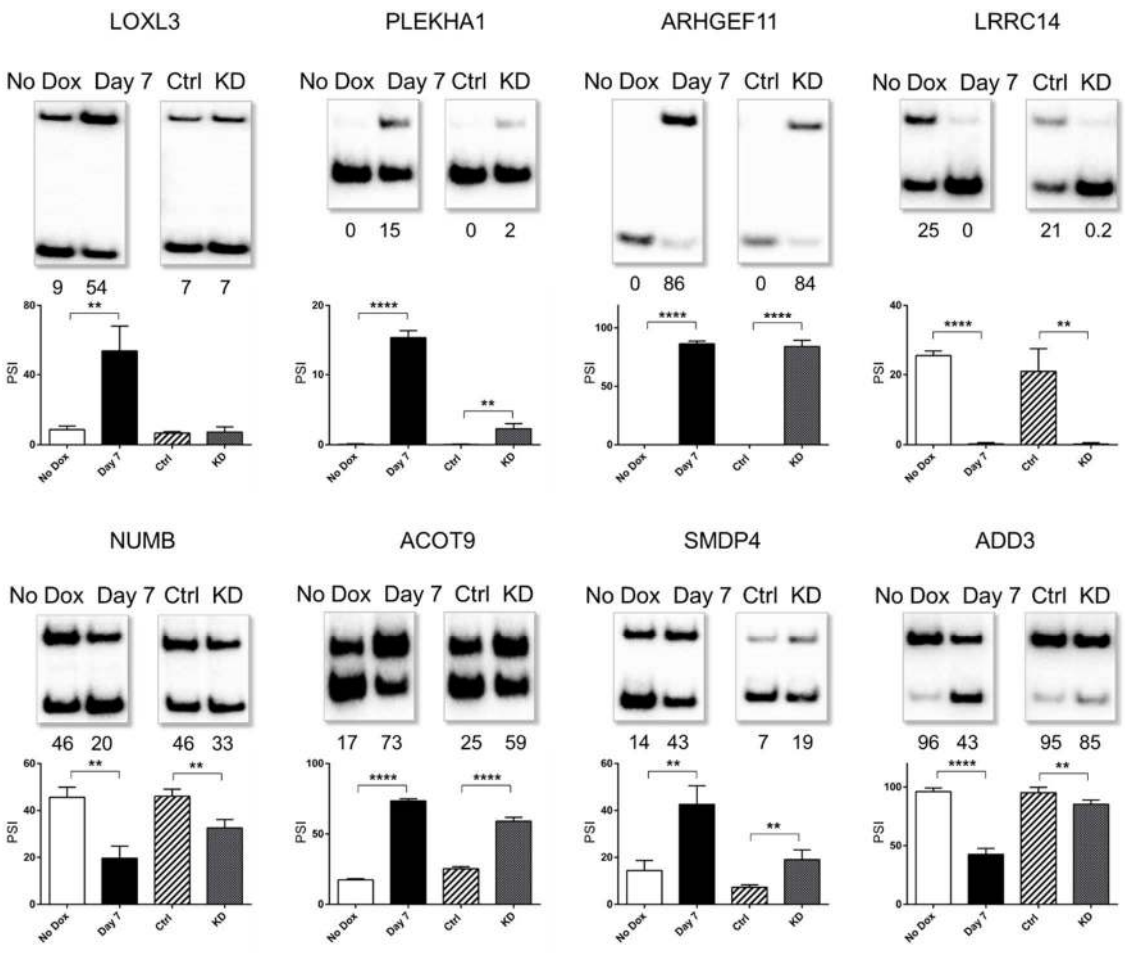


Figure 2-7 RT-PCR validation of exons predicted to switch during EMT, but not ESRP1/2 knockdown.

In addition to cassette exons with altered splicing during EMT as well as with ESRP depletion according to the rMATS analysis, there were a significant number of EMT-associated splicing events that were not identified as ESRP targets using the same statistical thresholds (88%, 548/670). However, we noted that many of those AS events in this analysis included exons previously shown to be ESRP targets in other systems or that were picked up in the rMATS analysis, but below the statistical cutoffs used to confidently identify ESRP-regulated targets (e.g. *MBNL1*, *ARHGEF11*, *EXOC7*). Given the known limitations of RNA-Seq for detecting AS changes in less highly expressed genes and those with smaller delta PSI, we suspect that the number of ESRP-regulated events was underestimated using the stringent criteria we applied. To further test this hypothesis, we selected 12 targets from a total of 51 AS events that showed at least 30% change of PSI during EMT while not identified upon ESRP depletion. For these events we analyzed the splicing changes during EMT as well as upon ESRP1/2 knockdown (Fig. 2-7). While four targets (*SPATS2L*, *IMPA1*, *WARS* and *KIAA1468*) showed no significant splicing changes during EMT or upon ESRP1/2 knockdown, eight were validated as EMT-associated AS targets. However, six out of these eight validated EMT-associated cassette exons (75%) are also regulated by the ESRPs, suggesting there is indeed an underestimation of the number of AS events regulated by the ESRPs during EMT. Notably, for those seven events, the change in splicing was usually smaller upon ESRP depletion compared to that during EMT.

We also selected a subset of targets from the 57 significant splicing events that were identified upon ESRP1/2 knockdown, but not during EMT based on the rMATS analysis. From six tested cassette exons, five were indeed regulated by the ESRPs, but of these *CEACAM1* and *RNF231* also showed an apparent splicing change during EMT (Fig. 2-8). In the case of *GRHL1* and *EPN3* we noted dramatic decrease in the total transcript level during the EMT, which made identification of isoform ratios difficult after EMT. We therefore suspect that some other targets eluded detection due to reduced total transcript level during EMT. Furthermore, among the 50 genes in which these 57 AS events occur, 17 had at least a 1.5-fold decrease in the total transcript level during the EMT. Specifically, for *GRHL1* we also note that the skipped isoform

results in a premature stop codon that renders the transcript sensitive to nonsense-mediated decay (NMD) that likely leads to under-detection of the degree of induced exon skipping.

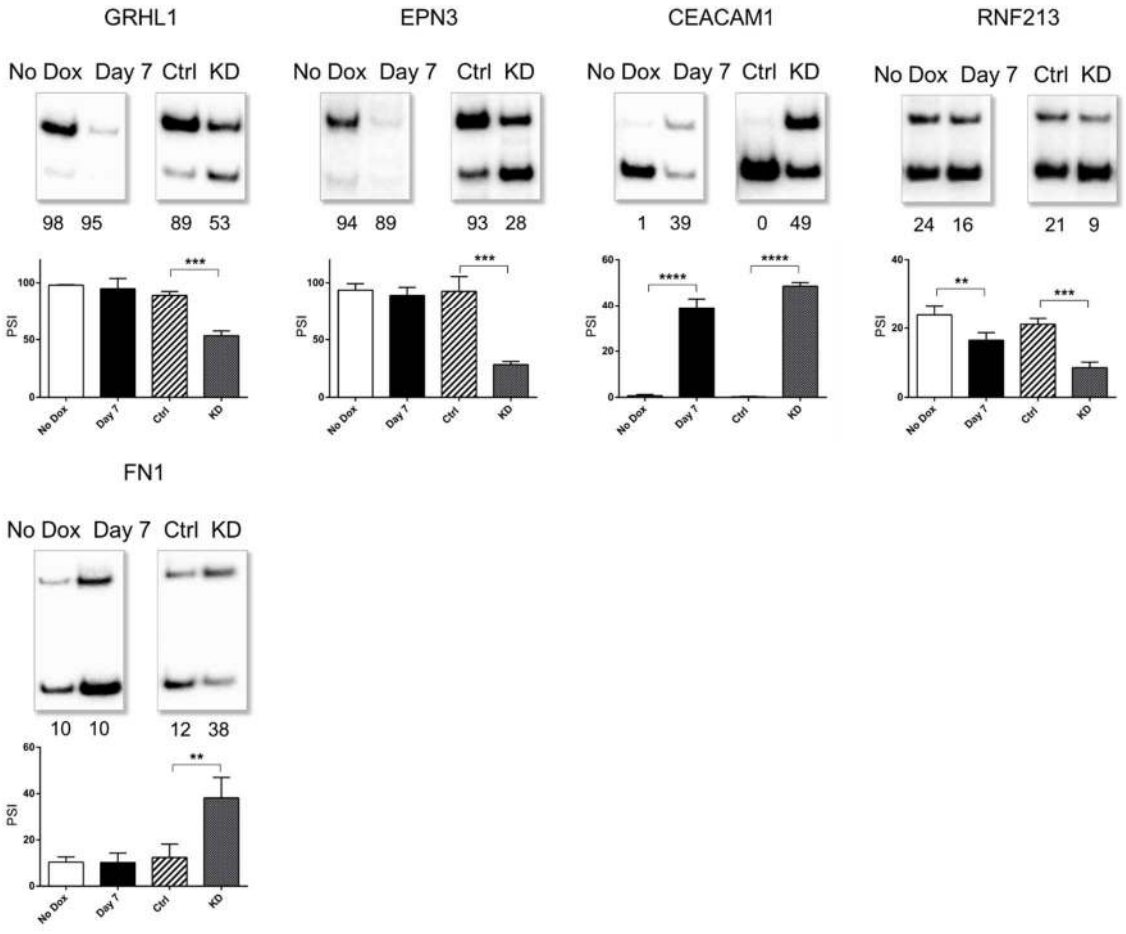


Figure 2-8 RT-PCR validation of exons predicted as ESRP-regulated events that were not identified during EMT.

2.2.5 A subset of RNA binding proteins showed expression changes during EMT

The observation that some events that switch only during EMT or upon ESRP depletion, but not both conditions and other events with smaller changes in AS upon ESRP depletion than that observed during EMT indicates that contributions by other splicing factors are involved in EMT. To identify additional splicing factors that play a role in regulation of AS during EMT, we evaluated changes in expression for 446 RNA binding proteins after 7 days of Dox treatment compared to No Dox controls, including all proteins that contain the most common canonical RNA binding domains, as well as additional proteins with known or putative roles in AS such as those with arginine-serine (RS) rich domains (Fig. 2-9A). We identified 68 RBPs with an expression change of at least 1.5 fold comparing Day 7 to controls, among which 48 were expressed at FPKM>1 in at least one condition (Table A2-7). As expected, ESRP1 and ESRP2 showed most dramatic change in total expression with a 10-fold decrease for ESRP1. Notably, we did not observe a substantial change in RBFOX2 expression level (less than 10% increase) in this EMT model. To further extend this analysis, we also carried out a cluster analysis for a subset of RBPs (average FPKM>5 in at least one condition) to identify temporal expression patterns that might correspond to those observed in certain splicing clusters (Fig. 2-9B and Table A2-1). Five major clusters with at least five RBPs were obtained. In cluster I, 57 RBPs showed an overall decrease at the total transcript level with some fluctuations during the time course. In cluster II, 34 RBPs had a dramatic decrease at total transcript level at Day 5 and 6, which however was converted to a level similar to controls at Day 7. MBNL1, MBNL2, MBNL3 and RBFOX2 are among this cluster. In cluster III, 28 RBPs showed a graded increase in total transcript level, which was peaked at Day 6 and had a slight drop at Day 7. In cluster IV, 22 RBPs showed a graded increase in total transcript level. Both ESRP1 and ESRP2 were in this cluster. The graded change in total transcript level was similar to the temporal patterns of splicing changes observed for the top two splicing clusters (Cluster 1 and 2). These observations suggested that events in these two clusters are more likely to be regulated by the ESRPs. Consistent with this prediction, we noted that ESRP-regulated exons from rMATS analysis were statistically enriched in Clusters 1 or

2 compared to a background set of EMT-associated SE that were not in either cluster (p=4.95e-28). In cluster V, 16 RBPs showed an overall increase at the total transcript level with a drop at Day 5 and 6. Among all five clusters, at least three (cluster I, II and V) showed an apparent switch at Day 5 that converted back at Day 7, which we found difficult to explain. Based on the expression of typical epithelial and mesenchymal markers, cells have completed EMT around Day 5, therefore the cellular status should be similar during the later time points. We then suspect that some of the unusual expression change are likely due to cell passaging and splitting on Day 2 and Day 4.

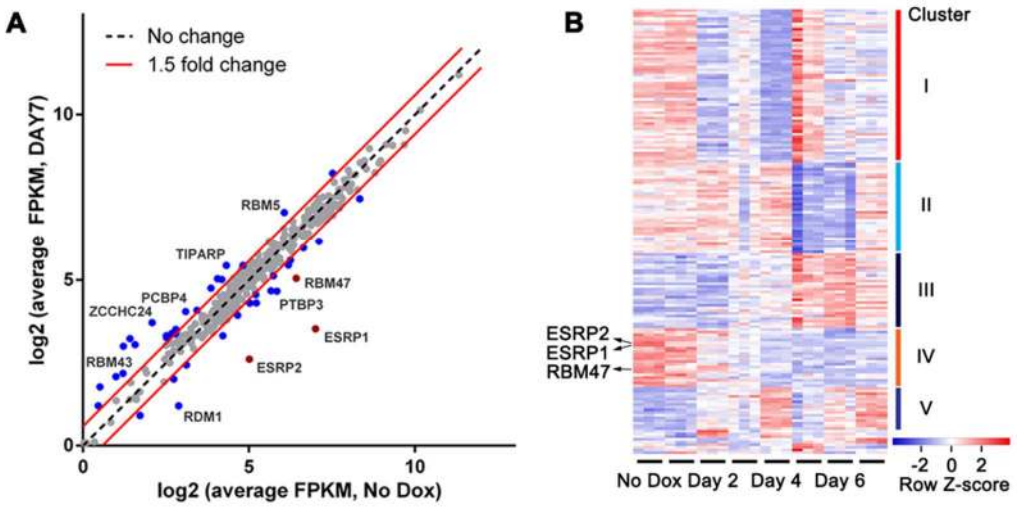


Figure 2-9 A subset of RNA binding proteins showed expression changes during EMT.

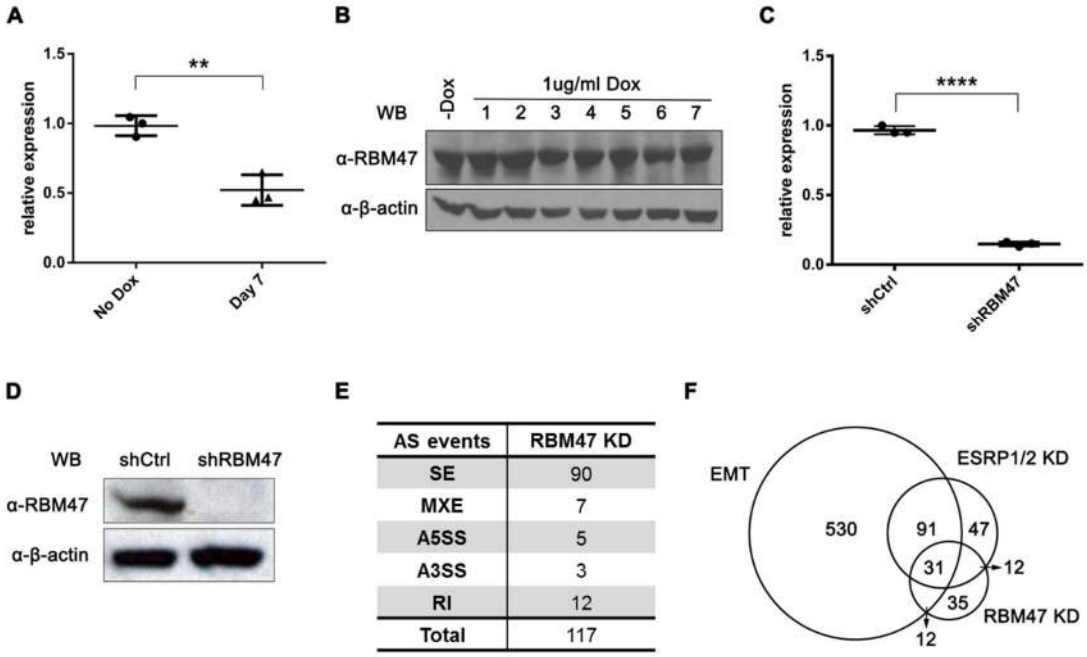
(A) Among 446 splicing factors, 48 (FPKM>1, highlighted in blue) showed at least 1.5-fold change in expression level comparing Day 7 to controls, including ESRP1, ESRP2 and RBM47 (highlighted in red). (B) Heat map of cluster analysis revealed five major temporal patterns of expression changes for 166 RBPs (see Table A2-1 for more details).

2.2.6 The splicing factor RBM47 regulates a subset of AS events during EMT

In the same cluster of the ESRPs, we identified RBM47 with a similar graded reduction in the expression level during the time course and a 50% decrease comparing Day 7 to No Dox controls. While the reduction at the mRNA level was confirmed by qRT-PCR, only a modest change at the protein level was apparent using standard Western blot (Fig. 2-10A, B). RBM47 was recently shown to regulate pre-mRNA splicing, mRNA stability as well as RNA editing (Fossat et al. 2014; Vanharanta et al. 2014). Among a set of the published cassette exon events that changed upon overexpression of RBM47 in MDA-MB-231-BrM2 cells, we noted many of those are either known ESRP targets or among cassette exons that switch splicing during EMT (Vanharanta et al. 2014). The ESRPs and RBM47 were predicted to regulate splicing in the same direction for the majority of common target events, suggesting RBM47 is likely to promote epithelial splicing patterns. For example, both RBM47 and the ESRPs promote exclusion of the 36nt exon 7 in MBNL1, while both promote inclusion of the 81nt exon 12 in MAP3K7. Notably, both cassette exons were also predicted as EMT-associated splicing events, suggesting potential combinatorial regulation during EMT. We therefore further investigated potential roles for RBM47 in the regulation of AS during EMT.

To further investigate the role of RBM47 in enforcing epithelial splicing patterns that are altered during EMT, we depleted RBM47 in H358 cells with an shRNA and generated cDNA libraries from RBM47 knockdown and control in biological triplicate (Fig. 2-10C, D). We then determined the AS program regulated by RBM47 using strand-specific 100 bp paired-end RNA-Seq and rMATS pipeline (Fig. 2-10E and Table 2-3). We identified 117 significant AS events (|ΔPSI|>5%; FDR <5%) including 90 cassette exons (Fig. 2-10D and Table A2-8). We evaluated the overlap of RBM47-regulated cassette exons with those that switch during EMT and found 43 out of 90 RBM47-regulated cassette exons that were also predicted to change during EMT (p=1.37E-29) (Fig. 2-10F and Table A2-9). Although RBM47 promotes epithelial splicing patterns for most of these targets (27/43), it was also predicted to promote mesenchymal splicing patterns in some cases. We tested a number of these RBM47-regulated targets by RT-PCR and

confirmed that indeed down-regulation of RBM47 likely promotes epithelial splicing patterns of some, but not all shared targets during EMT (Fig. 2-11). Similar results were obtained for all those RBM47-regulated AS events using an siRNA targeting a different region in the RBM47 transcript other than that of the shRNA used in the RNA-Seq experiment (Fig. A2-1).



AS events	RBM47 KD
SE	90
MXE	7
A5SS	5
A3SS	3
RI	12
Total	117

Figure 2-10 The splicing factor RBM47 regulates a subset of AS events during EMT.

(A-B) qRT-PCR validated a 2-fold decrease at the mRNA level for RBM47 at Day 7 versus No Dox, while Western blot only showed a modest change at the protein level for RBM47 during EMT.

(C-D) Validation of RBM47 depletion at the mRNA level and protein level in H358 cells using lentiviral shRNAs (in biological triplicate).

(E) Summary of different types of significant AS events identified upon RBM47 depletion in H358 cells.

(F) Venn diagram of the overlap between EMT-associated cassette exons and those regulated upon ESRP1/2 and RBM47 knockdown.

Table 2-3 Sequencing statistics for RBM47 knockdown experiment.

Sample	Total read pairs	Mapped pairs	%	Genome reads	%	Junction reads	%	% Exon	% Intron	% Intergenic
CTRL-1	51,743,743	25,926,653	50.1%	38,765,602	37.5%	13,087,704	12.6%	39.3%	6.4%	4.4%
CTRL-2	50,480,995	25,225,392	50.0%	37,242,131	36.9%	13,208,653	13.1%	39.9%	5.7%	4.4%
CTRL-3	53,953,222	27,081,890	50.2%	40,182,119	37.2%	13,981,661	13.0%	39.7%	6.0%	4.5%
KD-1	46,287,754	23,275,027	50.3%	34,659,995	37.4%	11,890,059	12.8%	39.1%	6.8%	4.4%
KD-2	47,479,380	24,016,488	50.6%	35,865,749	37.8%	12,167,227	12.8%	39.2%	6.9%	4.6%
KD-3	48,932,320	25,087,600	51.3%	37,814,836	38.6%	12,360,364	12.6%	39.4%	7.3%	4.6%
Total	**298,877,414**	**150,613,050**	**50.4%**	**224,530,432**	**37.6%**	**76,695,668**	**12.8%**	**39.4%**	**6.5%**	**4.5%**

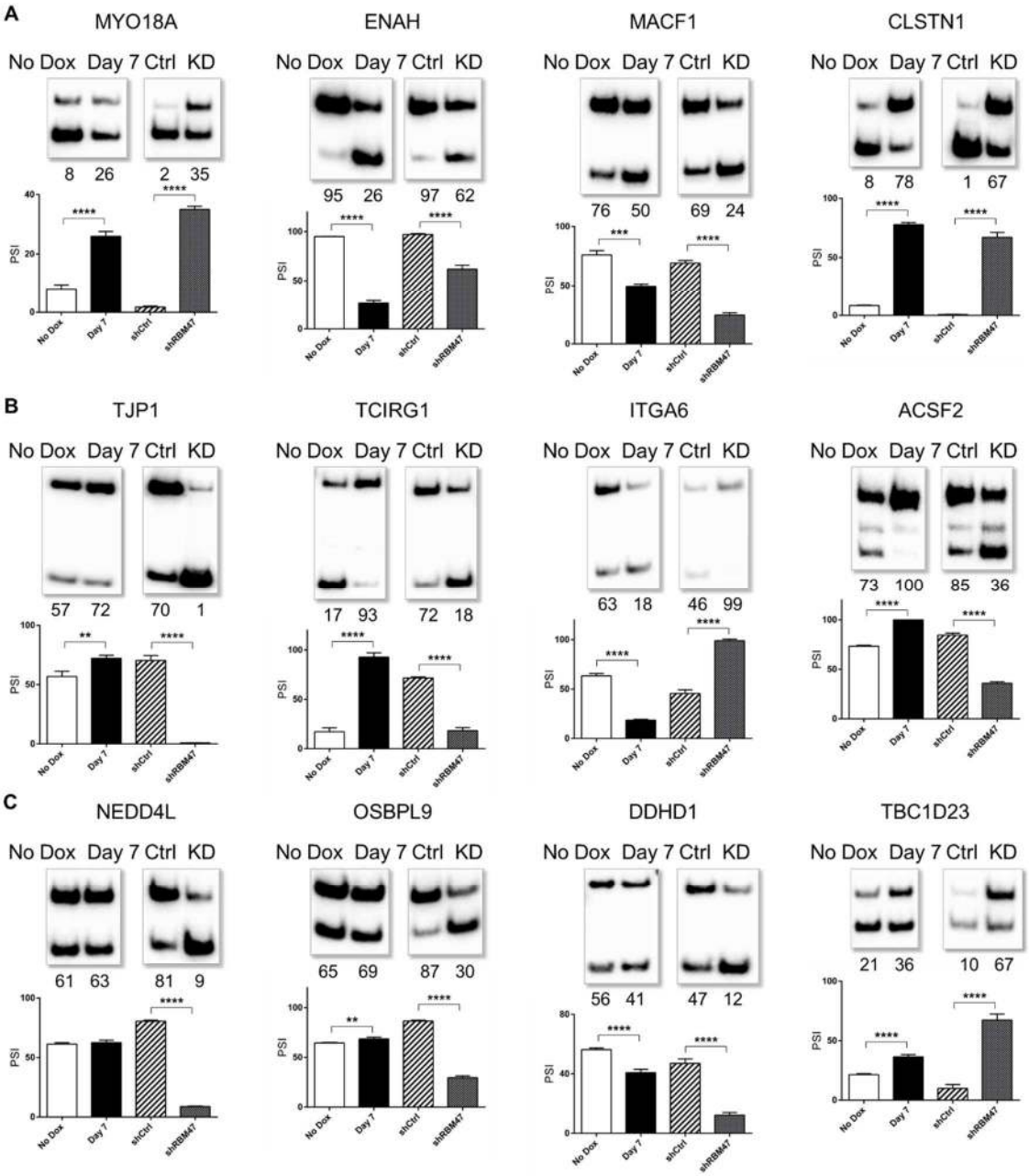


Figure 2-11 RT-PCR validation of exons predicted as RBM47-regulated events.

(A) Validation of representative exons where RBM47 promotes epithelial splicing patterns.

(B) Validation of representative exons where RBM47 promotes mesenchymal splicing patterns. For *ACSF2*, there is an alternative 5' splice site at the end of the cassette exon, which yields the middle band in between the included/skipped PCR products. Only the top and bottom bands are quantified to calculate the PSI.

(C) Validation of RBM47-regulated exons that were not predicted to switch during EMT.

2.2.7 The ESRPs and RBM47 regulate a subset of EMT-associated splicing events in a combinatorial manner

Among the 43 RBM47-regulated, EMT-associated targets, 31 events (72.1%) are also regulated by the ESRPs based on the rMATS analysis, indicating a significant overlap of co-regulated AS events, of which the majority (19/31) are regulated in the same direction by the ESRPs and RBM47 (Fig. 2-10F). This complex combinatorial regulation between the ESRPs and RBM47 on AS is similar to that previously described for the ESRPs and RBFOX2 where they usually, but not always, promote opposite changes in splicing (Dittmar et al. 2012). This is consistent with a function for RBFOX2 to usually, but not always, promote mesenchymal splicing patterns. To further investigate the combinatorial effect of the ESRPs and RBM47 on their common targets, we tested AS of several targets following knockdown of ESRP1/2 or RBM47 alone and with combined knockdown of both regulators in H358 cells (Fig. 2-12A, B). For MBNL1, CLSTN1, MACF1 and MAP3K7, knockdown of ESRP1/2 or RBM47 alone induced a partial splicing switch towards the mesenchymal splicing pattern, whereas the combined knockdown revealed an additive switch in splicing, consistent with each contributing to the loss of epithelial-specific splicing during EMT (Fig. 2-12C). In contrast for TCIRG1, ACSF2, MYO1B and ITGA6, ESRP1/2 knockdown induced a change in splicing consistent with that during EMT which is opposite to that observed in RBM47 knockdown, indicating that the effect of ESRP down-regulation during EMT predominated over that for RBM47 (Fig. 2-12D). Based on these collective observations, we suspect that for exons where the ESRPs and RBM47 have opposing functions, the reduction in ESRP expression plays a more prominent role in the splicing switches that occur during EMT. Hence, whereas the down-regulation of both the ESRPs and RBM47 together can account for the degree of some splicing switches observed during EMT, the effect of RBM47 to maintain epithelial splicing is less consistent than that of the ESRPs. Although the mechanism underlying the complicated combinatorial regulation is largely unclear, these results further indicate the EMT-associated AS network is fine-tuned by multiple regulators.

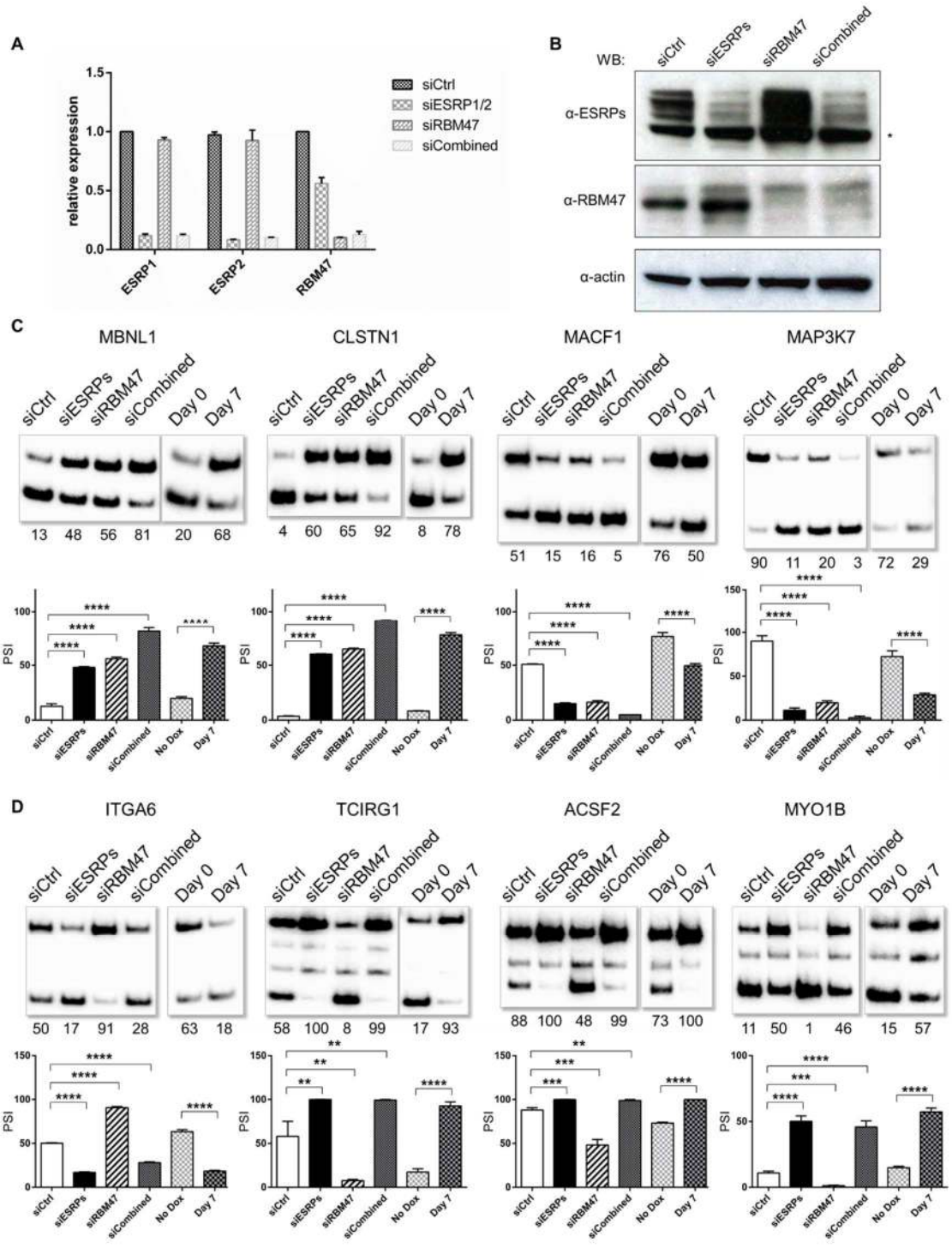


Figure 2-12 Combinatorial regulation of AS during EMT by the ESRPs and RBM47.

(A) qRT-PCR validation of ESRP1, ESRP2, and RBM47 depletion at mRNA level using siRNAs (in biological triplicate). Graph is shown as mean with error bars representing standard deviation.

(B) Western blot validation of ESRP1, ESRP2 and RBM47 protein knockdown. Note that, RBM47 mRNA showed a 50% reduction upon ESRP1/2 knockdown while the protein level didn't appreciably change.
(C) Validation of representative exons where ESRP1/2 and RBM47 knockdown have additive functions to promote splicing changes that occur during EMT.
(D) Validation of representative exons where ESRP1/2 and RBM47 promote opposite changes in splicing. For MYO1B, there are two consecutive cassette exons (both are 87nt in length) that can be included individually (middle band), together in tandem (top band) or skipped together (bottom band). Only the top and bottom bands are quantified to calculate the PSI.

2.2.8 Motif enrichment analysis revealed additional potential splicing regulatory proteins during EMT

While expression analysis of RBPs during EMT is helpful to identify potential splicing regulators with expression change, it is important for us to also consider contributions by splicing factors that did not demonstrate substantial changes in expression level during EMT. For example, RBFOX2 promotes predominantly mesenchymal splicing despite not showing a substantial change in total transcripts in our EMT model. To uncover other EMT splicing regulators, including those without clear expression changes, we performed a motif enrichment analysis for 115 known splicing factor binding motifs flanking cassette exons that switch during EMT (Fig. 2-13A and Table A2-10). Binding motifs of several SR proteins were enriched flanking EMT-associated cassette exons. Consistent with the "RNA map", ESRP1 binding motifs were specifically enriched downstream of cassette exons with increased skipping during EMT, indicating the presence of the ESRPs in epithelial cells promotes inclusion of those exons while the loss of the ESRPs in mesenchymal cell promotes skipping (Fig. 2-13B). Binding motifs of several other RBPs, including MBNL1/2 and CELF protein family member, were also specifically enriched downstream of cassette exons with increased skipping during EMT. RBFOX2 binding sites are specifically enriched downstream of cassette exons whose splicing increases during EMT, indicating the presence of RBFOX2 in mesenchymal cells promotes the inclusion of those exons. Similar to RBFOX2, QKI binding sites are enriched downstream of cassette exons with increased inclusion following EMT. Previous reports showed QKI also adheres to an RNA-map in which binding in the downstream intron promotes exon inclusion while binding in the upstream intron promotes skipping (Hall et al. 2013). Therefore, our motif analysis suggested a potential role for QKI in promoting inclusion of mesenchymal specific exons.

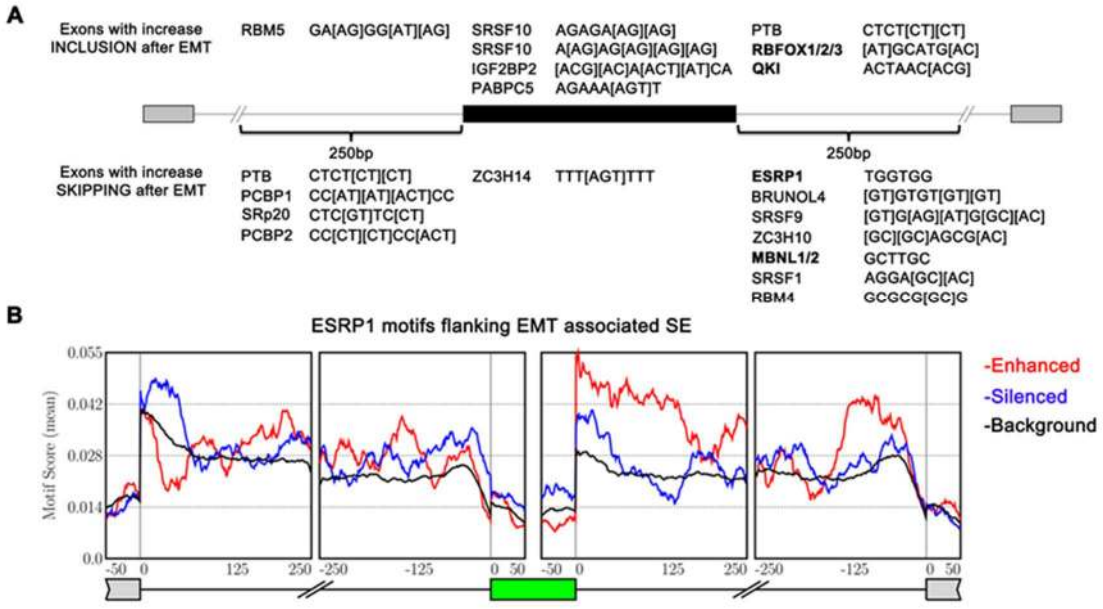


Figure 2-13 Identification of motifs enriched flanking EMT-associated cassette exons.

(A) Motifs and the corresponding RPBs that are enriched flanking exons that switch splicing during EMT.

(B) ESRP1 binding motifs is enriched downstream of exons with increased skipping after EMT.

2.2.9 QKI promotes mesenchymal splicing patterns for AS events during EMT

In H358 cells, QKI-5 is the most abundant isoform and we did not observe major isoform switches or significant expression changes during EMT (Fig. 2-14A). A recent study characterized the splicing events that change upon QKI knockdown in a lung cancer cell line using RNA-Seq (Zong et al. 2014). We noted that several of the QKI-regulated cassette exons identified in that study showed the opposite change in splicing from what we observed with ESRP depletion suggesting opposing roles in splicing regulation during EMT. We therefore applied the rMATS pipeline that we used to define splicing switches during EMT as well as ESRP1/2 and RBM47 knockdown to the same dataset and identified 70 QKI-regulated SE events (Table A2-11). Compared to EMT-associated cassette exons, there was a significant overlap (p=1.17E-24) with 35 shared target events despite the fact that they were identified in different cell lines (Table A2-

11). Strikingly, for 33 out of 35 QKI-regulated, EMT-associated AS events, QKI was predicted to promote the mesenchymal splicing patterns, including 12 splicing events that are co-regulated by the ESRPs. Notably, the two target events where QKI seemed to promote the epithelial patterns, the splicing change was relatively subtle. To further validate the RNA-Seq results, we knocked down QKI in a human mesenchymal breast cancer cell line MDA-MB-231 using two different siRNAs (siQKI #6 and #7) (Fig. 2-14B, C). We tested eight out the 33 AS events using semi-quantitative RT-PCR. All eight AS events were validated and QKI indeed promoted the mesenchymal splicing patterns for all tested AS events, suggesting QKI is a novel regulator of EMT-associated splicing switches, promoting mesenchymal splicing (Fig. 2-14D and Fig. A2-2).

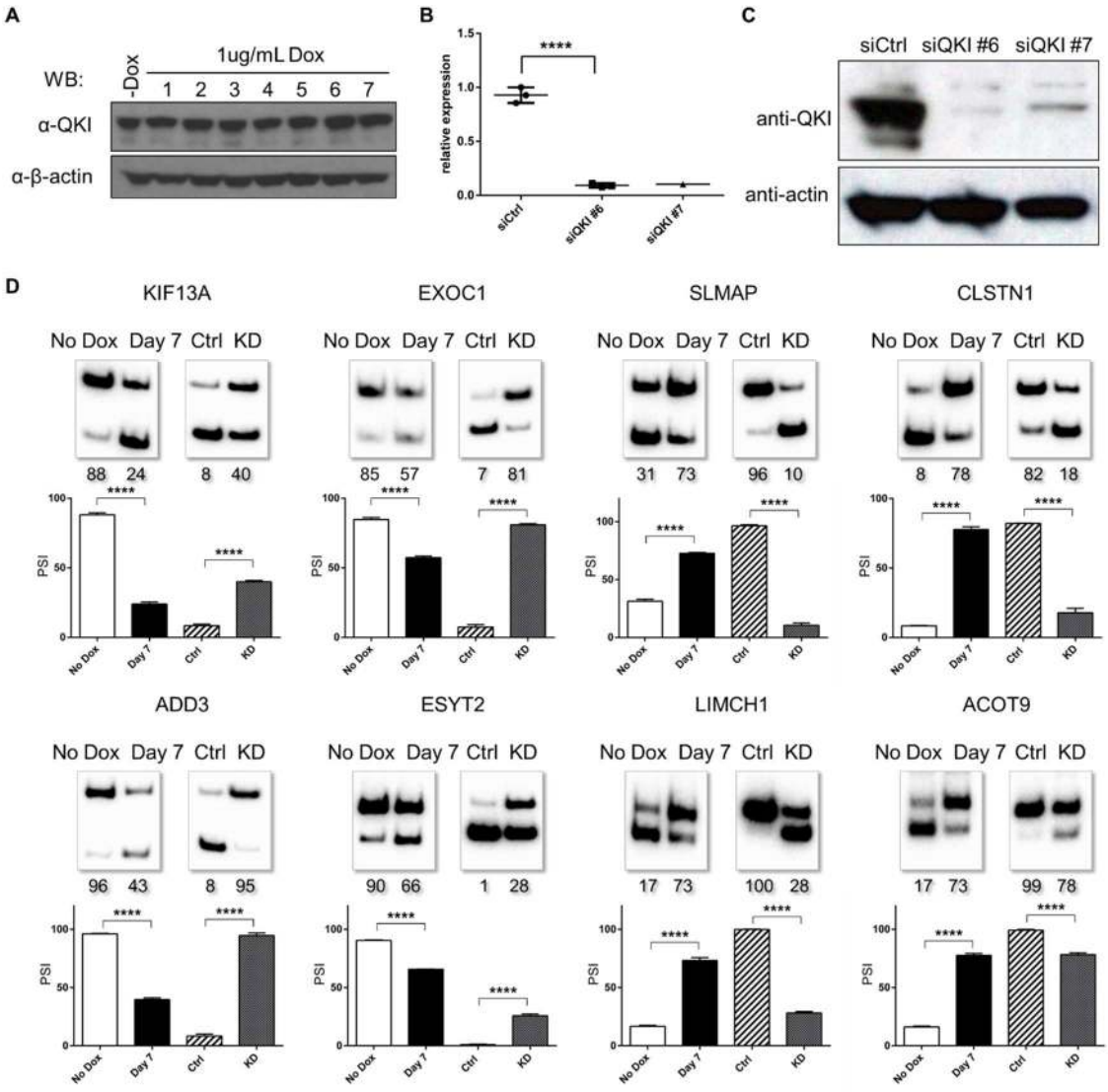


Figure 2-14 QKI promotes mesenchymal splicing patterns for AS events during EMT.

(A) Western blot of QKI doesn't show detectably change at protein level during EMT.

(B) qRT-PCR validation of *QKI* depletion at mRNA level in mesenchymal MDA-MB-231 cells using two different siRNAs (biological triplicate for #6 and one replicate for #7).

(C) Western blot validation of QKI knockdown at protein level.

(D) Validation of representative exons where QKI promotes the mesenchymal splicing patterns upon knockdown with siQKI #6.

2.3 Discussion and future directions

We comprehensively determined an AS program associated with EMT in a Zeb1 inducible model. We showed that while the ESRPs are major splicing regulators for EMT, two other RBPs, namely RBM47 and QKI, are also important for AS regulation during EMT. We further investigated the combinatorial regulation between the ESRPs and RBM47, by showing that they generally work cooperatively to promote epithelial splicing patterns, whereas QKI exclusively promotes mesenchymal splicing patterns. Based upon these findings we present a simplified model for combinatorial AS regulation during EMT involving the ESRPs, RBFOX2, QKI, and RBM47 (Fig. 2-15).

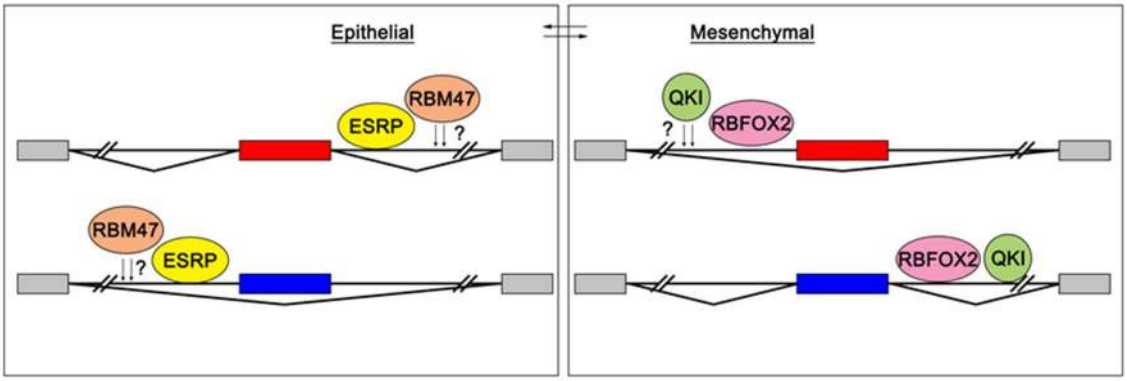


Figure 2-15 A model for AS regulation during EMT.

In epithelial cells, the ESRPs bind to downstream intron and promote epithelial specific exon inclusion, and bind to upstream intron and promote exon skipping. While it is not clear whether RBM47 binds directly to some overlapping targets, it also helps to maintain epithelial splicing for a subset of exons. In mesenchymal cells, expression of the ESRPs and RBM47 are decreased. In the absence of ESRP binding, RBFOX2 and QKI bind to downstream intron and promote mesenchymal specific exon inclusion; while RBFOX2 also binds to upstream intron and promote mesenchymal specific exon skipping. QKI also can promote exon skipping possibly also through binding in the upstream intron.

We showed that the ESRPs are the most down-regulated RBPs in our H358 EMT model, which was also noted in a previous study using another EMT model (Shapiro et al. 2011). This observation suggested that the loss of ESRP1 and ESRP2 is most likely to contribute to more splicing switches during EMT than any other specific splicing factor or paralog family. Consistent with this proposal, based on previous results and the RNA-Seq analysis from our EMT model, the ESRPs were shown to have more significant overlap with EMT-associated splicing switches than other proposed EMT splicing factors, including RBFOX2, PTB, HNRNPM, RBM47 and QKI, although the use of different cell lines may affect the comparison. Notably, ESRP1/2 depletion in H358 cells didn't affect the protein level for RBM47 or QKI (Fig. A2-3). While many EMT-associated splicing switches were not identified as ESRP targets based on rMATS analysis, validation of a subset of such target transcripts suggested that the majority are in fact ESRP-regulated, and that many such events are at least partially regulated by the ESRPs, despite not passing our stringent statistical thresholds. However, whereas ESRP depletion induced splicing changes in the same direction as those observed during EMT, the change in PSI was frequently less than that during EMT despite a similar reduction in ESRP expression levels. In addition, motif analysis confirmed that ESRP1 binding sites are enriched downstream of cassette exons with decreased inclusion after EMT, supporting a direct role of the ESRPs in promoting inclusion of epithelial specific exons. However, future studies using methods such as CLIP-Seq are needed to experimentally validate direct vs. indirect targets.

We identified RBM47 as another splicing factor that is down-regulated during EMT and showed that combined down-regulation of the ESRPs and RBM47 can account for the complete EMT splicing change for at least some of these examples. This combinatorial regulation by the ESRPs and RBM47 usually promotes an additive effect on splicing, although there were some examples of opposing effects on splicing that indicate the complexity of combinatorial regulation of splicing during EMT. Motif analysis led to our discovery of QKI which also plays a role in regulation of AS during EMT. QKI has been shown to bind to a ACUAAY motif both *in vitro* and *in vivo* and to both promote or inhibit splicing in a position dependent manner (Ryder and

Williamson 2004; Galarneau and Richard 2005; Hafner et al. 2010; Hall et al. 2013). Consistent with a direct function of QKI in promoting mesenchymal splicing, we found QKI binding motif enriched downstream of cassette exons with increased inclusion after EMT. Interesting, QKI is among the most down-regulated genes in cells with ectopically expressed epithelial specific miR-200, further suggesting QKI is associated with mesenchymal cell status (Park et al. 2008). In support of this, microarray data from the NCI60 panel of cell lines revealed that the expression of QKI is highly correlated with that of Vimentin, a mesenchymal marker (Park et al. 2008). These observations suggested an elevated expression of QKI in mesenchymal cells compared to epithelial cells. QKI showed a 40% increase in expression at the mRNA level in our model, but no appreciable change at protein level during EMT, yet it promoted mesenchymal splicing for about 30 cassette exons. It is possible that the modest change of QKI transcripts observed in our EMT model may be greater if extended over a longer time. Although we did not observe splicing isoform changes in QKI during EMT, we can't rule out the possibility that post-translation modifications of QKI may occur during EMT that alter its activity. Interestingly, a recent paper showed that QKI-5 promotes the formation of circRNAs that increase in abundance in EMT (Conn et al. 2015b). It will be of interest to determine the degree to which QKI regulated conventional AS events contribute to the EMT compared to circRNAs. Although we have shown that the ESRPs, RBM47 and QKI are responsible for a significant fraction of AS events during EMT, other splicing factors surely remain to be identified that also participate in combinatorial regulation of AS during this process.

A major challenge remains to define the functional consequences of AS switches that occur during EMT and how the change in protein isoforms confers differential activities that impact the cellular changes that accompany this developmental transition. The complete functional dissection of any specific AS event requires detailed investigations at the molecular and cell biological level in order to characterize these differences. For example, an in depth investigation into the AS of *EXOC7* (exocyst complex component 7) showed that the mesenchymal, but not epithelial, isoform promotes actin polymerization and cell invasion in

models of cancer metastasis (Lu et al. 2013). Another well-documented example is AS in *CD44* with a cluster of cassette exons in the pre-mRNA, while inclusion of one or more of these exons leads to *CD44* variant isoforms (*CD44v*, epithelial isoforms), exclusion of all cassette exons generates the *CD44* standard isoform that promotes EMT (*CD44s*, mesenchymal isoform). The transition from *CD44v* to *CD44s* was proposed to play a central role in EMT (Brown et al. 2011). In previous work, we also discussed several genes with known or predicted isoform specific functions such as *NUMB*, *EPB41L5* and *TCF7L2* that may contribute to processes that impact EMT (Warzecha et al. 2010; Dittmar et al. 2012). While there are limited examples of other EMT-associated splicing switches that have been functionally well characterized, there are a number of transcripts identified here with functions that are relevant to EMT, such as epithelial cell adhesion and polarity. It therefore merits further investigation as to how the epithelial vs. mesenchymal isoforms differentially affect these functions. For example, ADD3 (adducin 3), a subunit of the adducin family, is a membrane skeletal protein involved in assembly of the spectrin-actin network at the membrane–cytoskeleton interface and is localized at cell-cell contacts of epithelial cells (Gardner and Bennett 1987; Kaiser et al. 1989). It was shown to co-localize with E-cadherin and CTNNB1 (or β-catenin) at AJs in epithelial cells and knockdown of adducin significantly attenuated calcium-dependent AJ and TJ assembly and accelerated junctional disassembly, indicating that it is essential for stability of epithelial junctions (Naydenov and Ivanov 2010). ADD3 was also shown to be required for desmosomal cohesion in keratinocytes (Rotzer et al. 2014). It is notable that these roles are regulated by phosphorylation as is stability of the protein (Matsuoka et al. 1998). The cassette exon for which the inclusion decreased significantly after EMT, contains several known or predicted phosphorylation sites, suggesting the possibility that the different isoforms might have differential functions due to presence or absence of specific phosphorylation events (Huttlin et al. 2010; Lundby et al. 2012). Therefore, this AS event in *ADD3* during EMT may impact functions of the protein in maintenance of cell-cell junctions in epithelial cells relevant to EMT, but this requires further study.

Another potentially relevant example for which there have been some studies of isoform differences is *CEACAM1* or carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein), a member of the immunoglobulin superfamily specifically belonging to the carcinoembryonic antigen (CEA) family (Barnett et al. 1989). The highly conserved 53nt cassette exon residing in the C terminus, when included, generates a 71 amino-acid cytoplasmic tail (*CEACAM1* L isoform) while skipping of the exon generates a 9 amino-acid cytoplasmic tail (*CEACAM1* S isoform) (Dráberová et al. 2000). While both isoforms can mediate adhesive interaction, they have different adhesive properties, localizations and post-translational modifications on their unique cytoplasmic tails, suggesting distinct functions in mediating cell-cell interactions and signaling transduction for each isoform (Huber et al. 1999; Sundberg and Obrink 2002). During EMT, there is a transition from primarily *CEACAM1* S isoform to about a 1:1 ratio of *CEACAM1* S and *CEACAM* L isoforms. Similarly, it was shown that normal breast epithelial cells predominantly express the *CEACAM1* S isoform, while the ratio of S/L isoforms is reduced in breast cancer cells (Gaur et al. 2008). Since EMT is implicated in promoting tumorigenesis by changing cell-cell interactions and making individual cell more invasive, the anti-tumorigenesis property of CEACAM1 may partly rely on proper AS of *CEACAM1* (Yang and Weinberg 2008).

Our findings lay the groundwork for studies of specific AS events that have isoform specific functions in EMT and related processes, such as metastasis, strengthening our understanding of development and tumorigenesis. Future studies are required to understand further how alternative splicing can affect cellular transitions and cell behavior. However, the study of even a single alternative event regulated during EMT or upon ESRP depletion is technically challenging and time-consuming. An alternative strategy to study the biological consequences of alternative splicing program during developmental transition is to look at differences in the protein-protein interaction (PPI) networks of different isoforms. Recent studies have shown that alternatively spliced isoforms have distinct protein-protein interaction profiles, similar to that of different genes, indicating alternative splicing induces extensive changes at

cellular level and information from the PPI can be used as a guide to study their potential functions (Yang et al. 2016).

While the identification of the ESRPs and two novel EMT splicing factors, RBM47 and QKI, helps to extend our understanding of the AS regulatory network during EMT, there are surely more splicing factors that contribute to this regulatory network that are yet to be identified. Future experiments are required to determine the degree to which these splicing factors directly regulate these splicing events. The example of the combinatorial regulation between the ESRPs, RBM47, RBFOX, and QKI are just the tip of the iceberg, underscoring an extensive network among different splicing factors in splicing programs during cellular transitions and determination of cell- and tissue-specific splicing. Lastly, future studies are required to identify the mechanisms by which the ESRPs regulate splicing.

2.4 Material and methods

2.4.1 Plasmids

pGIPZ-shESRP1 #3, pGIBZ-shESRP2 #7 as well as the corresponding non-targeting control vectors, psPAX2 and pCMV-VSV-G were described previously (Warzecha et al. 2010). We PCR amplified the cDNA sequence for mCherry and used it to replace the coding sequence for EGFP in pGIBZ to derive pRIBZ-shESRP2 and corresponding control vector pRIBZ-shCtrl (sequences and cloning strategies are available at request). Note that these shRNA vectors facilitated visualization of transduction (green or red fluorescence) and drug selection (puromycin or blasticidin). The pGIPZ-shRBM47 expression vector was purchased from GE Dharmacon (V3LSH_393928). The pTet-On advanced vector was purchased from Clontech (631069). We cloned the rtTA-Advanced cassette into the pIBX vector described previously using EcoRI and BamHI sites to make the pIBX-Tet-On vector (Warzecha et al. 2009a). The pTRE-tight vector was purchased from Clontech (631059). To enable selection, we cloned the TRE-CMVmin element into the pcDNA3 vector to derive pcDNA3-TRE-CMVmin. A subsequent construct, pcDNA-TRE-CMVmini-C-FF(B)-mCherry was derived by inserting a sequence encoding a 2X FLAG tag followed by the coding sequence for mCherry. The coding sequence for *ZEB1* was subsequently inserted upstream of the sequences for the FLAG tag and mCherry to derive pcDNA3-TRE-CMVmin-C-FF(B)-mCherry-ZEB1 that encoded ZEB1 with C-terminal FLAG and mCherry (sequences and cloning strategies are available at request).

2.4.2 Cell culture and transfection

Human non-small cell lung cancer cell line H358 (obtained from the American Type Culture Collection) and the H358 Zeb1 clone were maintained in RPMI1640 with 10% FBS (SH30071.03, GE). 293T, and MDA-MB-231 cells were maintained in DMEM with 10% FBS. To make a Tet-on inducible Zeb1 H358 clone: First, H358 cells were transfected with the pIBX-Tet-On plasmid using Lipofectamine® 2000 (11668027, Life technologies) according the

manufacturers' protocols, selected in 10 ug/ml blasticidin for two weeks, and used serial dilution in 96 well plates to obtain single cell derived clones. Second, pcDNA3-TRE-CMVmin-C-FF(B)-mCherry-Zeb1 was transfected into the H358 Tet-On clone using Lipofectamine® 2000, selected in 500 ug/ml G418 (10131035, Life technologies) and single cell clones were derived by serial dilution. For the EMT time course experiment, the Tet-on inducible ZEB1 H358 cells were seeded in 6 cm dishes in biological triplicate. Then we treated cells with 1 ug/ml Doxycycline every other day for 7 days in total to induce and maintain ZEB1 expression. During the time course, we harvested RNA and protein from each day for downstream analyses. Note that cells reached confluence at Day 2 and Day 4, therefore we trypsinized and re-plated following the isolation of RNA at the Day 2 and 4 time points.

2.4.3 Viral packaging and transduction

Lentiviral production and transduction were performed as described previously (Warzecha et al. 2010). Briefly, 293T cells were transfected in 6 cm dishes with 3 ug of the shRNA expression vector, 2.7 ug of psPAX2, and 300 ng of pCMV-VSV-G using TransIT®-293 (MIR2700, Mirus). After 16–20 hours, the media was replaced with fresh DMEM with 10% FBS, and virus was harvested after an additional 24 hours. Target cells were transduced with a 50/50 mix of viral supernatant and growth media. Selection was carried out using 2 ug/ml puromycin and/or 10 ug/ml blasticidin for 48-96 hours. RNA and protein were harvested 7-8 days post infection. ESRP and RBM47 knockdown using shRNAs followed by RNA-Seq experiments were both done in biological triplicate.

2.4.4 RNA interference using siRNA

siRNA for RBM47 (SI04356884, target sequences are CACGGTGGCTCCAAACGTTCA) and QKI (#6 SI04218221, target sequences are CCCGAAGCTGGTTTAATCTAT; #7 SI04367342, target sequences are CAGAGTACGGAAAGACATGTA) were purchased from Qiagen. siRNAs for ESRP1 and ESRP2 were previously described (Warzecha et al. 2009a). The

non-specific Allstar siRNA (SI03650318) was purchased from Qiagen as a negative control. Briefly, cells were seeded in 6 well plates and transfected with siRNAs twice over a period of 48 hours using Lipofectamine® RNAiMAX (11778075, Life technologies). A total of 30 pmole siRNA was used for each transfection. RNA was extracted 24-48 hours after the second siRNA transfection (72-96 hours following the first transfection). Knockdown experiment of ESRP and/or RBM47 using siRNAs to test the combinatorial effect was done in biological triplicate. Knockdown experiment of QKI in MDA-MB-231 cells was also done in biological triplicate.

2.4.5 RT-PCR, and qRT-PCR

Total RNA was extracted using TRIzol (15596018, Life technologies). Reverse transcription was performed as described previously. Primers used for AS targets validation, exon sizes and PCR products sizes are summarized in Table A2-12. qRT-PCR analysis was performed and analyzed as described previously. 18S was used as an endogenous control for normalization. Each qRT-PCR contained average data from technical triplicate. A one-tailed unpaired t-test was used to calculate the p-values for all RT-PCR and qRT-PCR from biological triplicate. Graphs are shown as mean with error bars representing standard deviation. Taqman assay probes for human ESRP1 (Hs00214472_m1), ESRP2 (Hs00227840_m1), RBM47 (Hs01001785_m1), QKI (Hs00287641_m1) and 18S (Hs03003631_g1) were purchased from Life technologies.

2.4.6 Antibodies and Western blotting

Total cell extracts were harvested in RIPA buffer with protease inhibitor cocktail, PMSF and sodium orthovanadate (sc-24948, Santa Cruz). Immunoblotting was performed as described (Warzecha et al. 2009a). Antibodies used are as below: ESRP1/2 (23A7, mouse, 1:200) (Warzecha et al. 2010); RBM47 (SAB2104562, Sigma, rabbit, 1:1000); QKI (HPA019123, Sigma, rabbit, 1:200); vimentin (MS-129-P1, Thermo Scientific, mouse, 1:500); CDH1 (24E10) (#3195, Cell signaling, rabbit, 1:500); ZEB1 (sc-25388, Santa Cruz, rabbit, 1:1000); FLAG (F1804, Sigma,

mouse, 1:5000); β-actin (A2228, Sigma, mouse, 1:5000). Secondary antibodies were purchased from GE Healthcare Life Sciences (sheep anti-Mouse IgG NA931; donkey anti-Rabbit IgG NA934V from 1:1000; to 1:10,000).

2.4.7 cDNA libraries and RNA-Seq

1 ug total RNA was used to make each cDNA library using the TruSeq Stranded mRNA LT Sample Prep Kit (RS-122-2102, Illumina) for EMT time course and ESRP1/2 knockdown experiments and the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina (E7420S, NEB) for RMB47 knockdown experiment. For cDNA library preparation using the NEB kit, poly-A selection from total RNA was done using the NEBNext® Poly(A) mRNA Magnetic Isolation Module (E7490S, NEB). 100bp paired-end RNA-Seq using the Illumina Hiseq 2000 or Hiseq 2500 was done by Penn Genome Frontiers Institute (PGFI) or Next-Generation Sequencing Core (NGSC) facilities at Penn.

2.4.8 RNA-Seq analysis

We mapped RNA-Seq reads to the human genome (hg19) and transcriptome (Ensembl, release 72) using the software TopHat (v1.4.1) allowing up to 3 bp mismatches per read and up to 2 bp mismatches per 25 bp seed. We computed RNA-Seq based gene expression levels (FPKM metric: fragments per kilobase of exon per million fragments mapped) using Cuffdiff (v2.2.0), and identified differential gene expression between the two sample groups at FDR<5%, >2 fold difference in gene expression based on average FPKM, and minFPKM>0.1 (Trapnell et al. 2010). We identified differential AS events between the two sample groups using rMATS v3.0.8 (http://rnaseq-mats.sourceforge.net) that detects five major types of AS events from RNA-Seq data with replicates (Shen et al. 2014). In each rMATS run, the first group was compared to the second group to identify differentially spliced events with an associated change in Percent Spliced In (ΔPSI or $\Delta\psi$) of these events. We ran rMATS using -c 0.0001 parameter to

compute p-values and FDRs of splicing events with $|\Delta\psi|>0.01\%$ cutoff then collected the splicing events with an FDR<5% and $|\Delta\psi|\geq 5\%$.

2.4.9 Motif Enrichment Analysis

We performed motif enrichment analysis as described previously to identify binding sites of splicing factors and other RNA binding proteins (RBPs) that were significantly enriched in differential exon skipping events (Bebee et al. 2015). We used 115 known binding sites (motifs) of RBPs from the literature (Ule et al. 2003; Anderson et al. 2012; Dittmar et al. 2012; Ray et al. 2013a; Vanharanta et al. 2014) which includes well-known splicing factors. We examined the enrichment of the RBPs in exon body, 250 bp of upstream and downstream intron separately.

2.4.10 RNA Map Analysis

To identify the RNA binding map of the ESRPs for the differential SE events between two groups as compared to control alternative exons, we examined ESRP binding sites using the top twelve GU-rich binding motifs previously identified by the SELEX-Seq (Dittmar et al. 2012). If an alternative exon didn't show splicing changes (rMATS FDR>50%, maxPSI>15%, minPSI<85%) and it was from highly expressed genes (average FPKM>5.0 in at least one group), we classified it as control alternative exon. As described previously, we examined the exon body, 250 bp of upstream and downstream intron, flanking exons, and 250 bp intronic regions of flanking exons to assign motif scores (Bebee et al. 2015). Motif scores were assigned as the overall percentage of nucleotides covered by any of twelve ESRP motifs within a 50-bp sliding window. We slid the window by 1 bp in each region.

2.4.11 Data Availability

The RNA-Seq data from this publication have been submitted to the NCBI Gene Expression Omnibus repository (http://www.ncbi.nlm.nih.gov/geo/) under the accession number

GSE75492, with the link below:

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=qlsfiwaybpihbmf&acc=GSE75492.

3) Chapter 3: Alternative splicing mediated subcellular localization of Esrp1

3.1 Background

Esrp1 is a highly conserved RBP, with orthologues in all vertebrates as well as D. Melanogaster (fusilli) and C. elegans (sym-2) that are also regulators of alternative splicing (Barberan-Soler and Zahler 2008; Warzecha et al. 2009). In addition to the three highly conserved RRMs, there are no other clearly defined protein domains. Several alternative splicing events are present in the C-terminus of Esrp1, downstream of all three RRM domains, including an alternative 5' splice site (5'ss) at the end of exon 12 and two consecutive cassette exons (exon 14 and 15) that can be included individually, together in tandem, or skipped. In this study, we demonstrated that the differential subcellular localization of Esrp1 results from the alternative 5' splice site at the end of exon 12. We identified the minimal peptide sequence that is necessary and sufficient for nuclear localization of Esrp1 nuclear isoforms, which is different from previously characterized NLS consensus. We also determined the key residues in the Esrp1 NLS. We further showed that the production of both nuclear and cytoplasmic isoforms through alternative splicing is maintained by the fly orthologue fusilli. These findings strongly suggest that Esrp1 and its orthologues have conserved functions in yet-to-defined post-transcriptional regulatory roles in the cytoplasm beyond splicing regulation in the nucleus

3.2 Results

3.2.1 Choice of different alternative 5’ splice sites downstream of exon 12 gives rise to Esrp1 isoforms with differential subcellular localization.

The alternative 5’ splice sites in Esrp1 exon 12 consists of two identical 5' splice site consensus sequences that are separated by 12 nucleotides (5-TGAAGTTACCAT-3), usage of which generates protein isoforms that differ by four amino acids "Cys-Lys-Leu-Pro" (CKLP) (Fig. 3-1A). Cassette exons 14 and 15, which are 151 nt and 72 nt in length respectively, can be included or skipped individually or together. While translation terminates in exon 16 when both exons are skipped (NA), inclusion of both exons (2A) introduces a stop codon in exon 15 (Fig. 3-2A, B). In our initial studies using cDNAs encoding several different splice isoforms of Esrp1, we noted these Esrp1 isoforms showed differential subcellular localization to the nucleus or cytoplasm, suggesting some Esrp1 splice isoforms contain a nuclear localization signal (NLS). To test whether the NLS is encoded by the differentially spliced region around the alternative 5’ splice site or the exons 14 and/or 15, we generated constructs expressing four Esrp1 isoforms (2A+CKLP, 2A-CKLP, NA+CKLP and NA-CKLP) fused with a fluorescence tag (mCherry) and transfected them into HeLa cells (Fig. 3-1B). As shown in Fig. 3-1C, isoforms containing the peptide CKLP (derived from the distal 5’ss) were predominantly nuclear while isoforms lacking the peptide (derived from the proximal 5’ss) were predominantly localized to the cytoplasm. However, the inclusion or skipping of exon 14 and 15 together or individually did not influence subcellular localization (Fig. 3-1C and data not shown). These observations suggested that CKLP is part of an NLS required for the nuclear localization of Esrp1. Using semi-quantitative RT-PCR with primers flanking both 5’ splice sites, we confirmed that alternative splicing of these 5’ splice sites across a panel of human and mouse cell lines yielded both splice variants, with some variation in ratios, but generally averaging close to 1:1 ratio of the nuclear and cytoplasmic isoforms (Fig. 3-1D). Interesting, we observed an increase in the nuclear isoforms of ESRP1 upon ESRP depletion in the human epithelial cell line H358 using shRNAs, indicating this

alternative splicing event is regulated by the ESRPs (Fig. 3-1E). This autoregulation is also conserved in mouse as a similar increase in nuclear Esrp1 was observed in the epidermis tissue from *Esrp1*/*Esrp2* double knockout mice compared to the wild type (Fig. 3-1E) (Bebee et al. 2015)

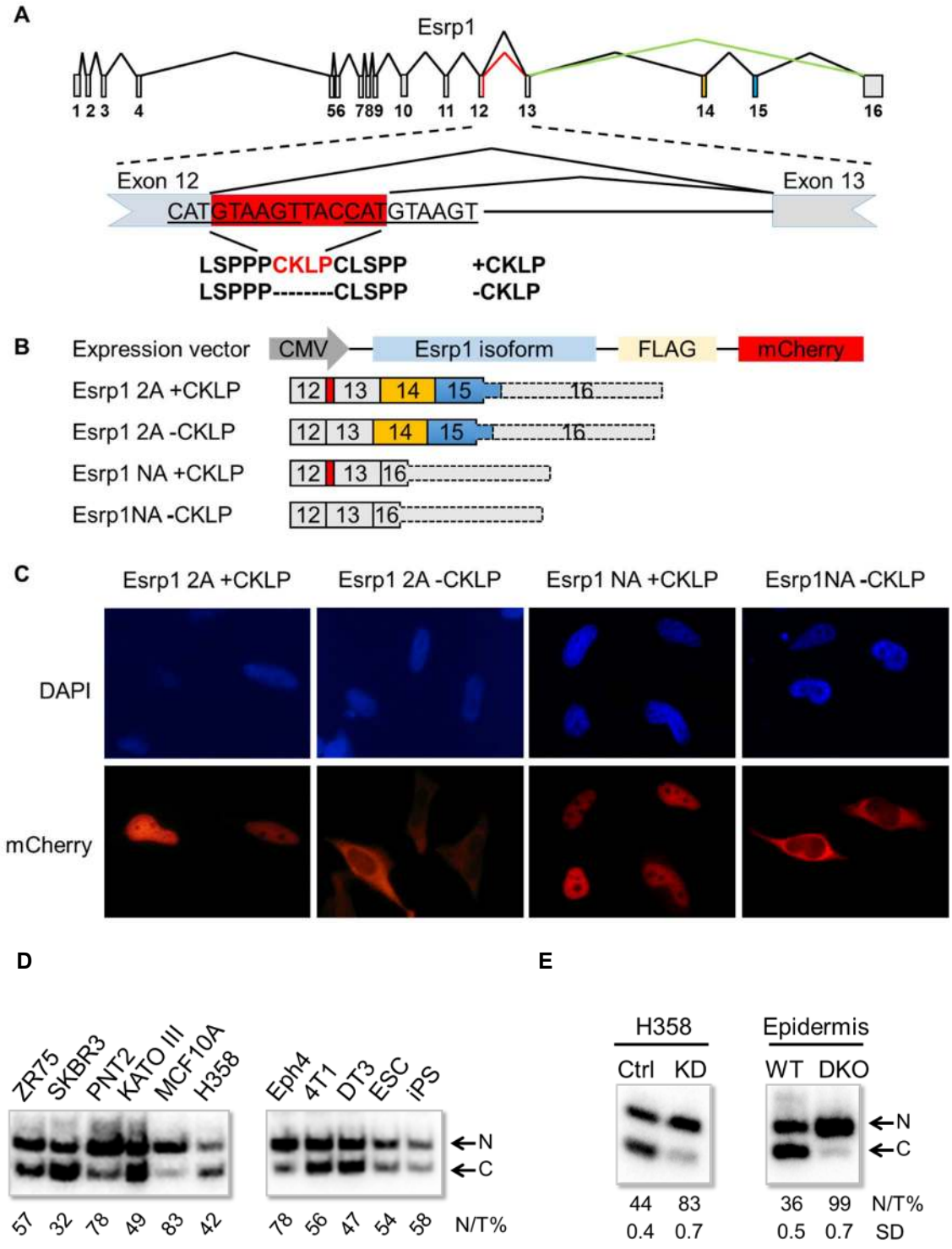


Figure 3-1 Usage of two 5' splice sites at the end of exon 12 gives rise to Esrp1 isoforms with differential subcellular localization.

(A) Schematic of genomic organization and alternative splicing of Esrp1. Gray boxes indicate constitutive exons. Exons 14 (yellow box) and 15 (blue box) are cassette exons, and exon 12 has an alternative 5' splice site (highlighted in red) with the specific sequences shown below.

(B) Schematic of vectors expressing different Esrp1 isoforms as mCherry fusion for transfection and microscopy in (C).

(C) Representative images of HeLa cells (20x) transfected with expression vectors of different Esrp1 isoforms depicted in (B), from which we concluded that CLKP is required for nuclear localization. mCherry represents the localization of Esrp1 expression, and DAPI represents the nucleus.

(D) RT-PCR using primers flanking both 5' splice sites across a panel of human and mouse cell lines confirmed the expression of both splice variants. N represents nuclear isoforms, C represents cytoplasmic isoforms. The percentage of nuclear isoforms over total transcripts (N/T) is calculated below the gel.

(E) ESRP1 (or Esrp1) nuclear isoforms are increased upon depletion of ESRPs (or Esrps). Representative gels are shown with average N/T percentage and standard deviation calculated from biological triplicates below.

3.2.2 Determination of the Esrp1 nuclear localization signal.

To further characterize the peptide sequence that is necessary and sufficient for nuclear localization, we aligned Esrp1 sequences from different species and focused on peptide sequences around the CKLP region to identify conserved amino acids (Fig. 3-2A). Importantly, this analysis also revealed that the expression of isoforms that do or do not contain CKLP is highly conserved across vertebrate evolution. Based on sequence conservation in the alignment, we selected three peptides around the CKLP region, which are 46, 27 or 15 amino acids (aa) in length respectively, and predicted these peptides might contain the Esrp1 NLS. To test this hypothesis, we used a well-defined cDNA reporter that has previously been used to identify and characterize NLSs (Kalderon et al. 1984; Siomi and Dreyfuss 1995). This vector encodes chicken pyruvate kinase (18-443 aa), a cytoplasmic protein and we further modified this reporter by adding a fluorescent tag (mCherry) to the C-terminus. We then inserted sequences encoding the three peptides from Esrp1 that included CKLP into the reporter and tested whether they could translocate chicken pyruvate kinase into the nucleus of HeLa cells (Fig. 3-2A, B). The SV40 NLS was used as a positive control. While all three peptides lacking “CKLP” showed a consistent cytoplasmic localization, all three peptides with “CKLP” were sufficient for nuclear localization including the shortest one, the fifteen-amino-acid stretch “GLSPPPCKLPCLSPP” (Fig. 3-2C). We therefore defined this sequence as the NLS for Esrp1.

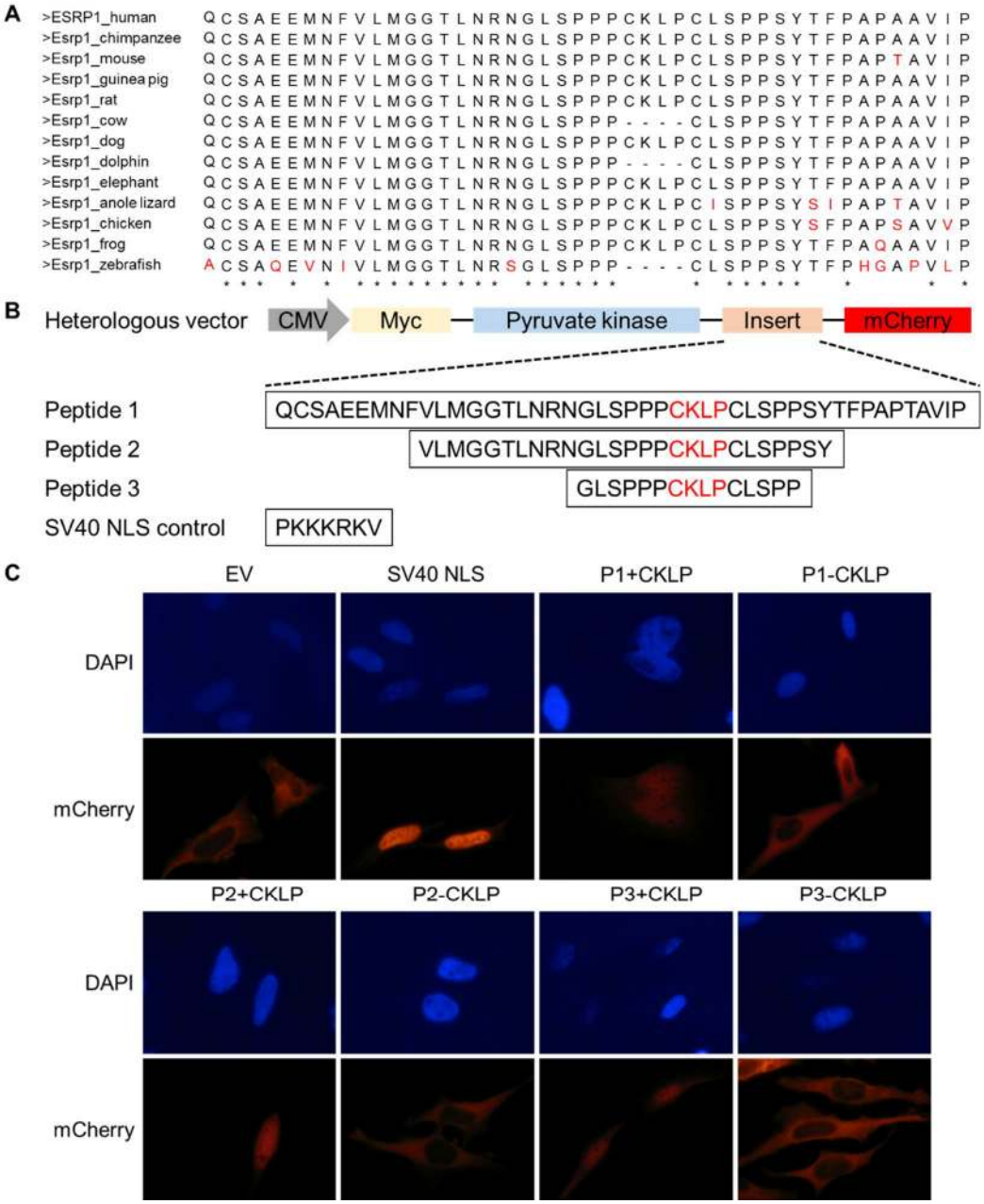


Figure 3-2 Determination of the Esrp1 nuclear localization signal.

(A) Alignment of sequences around the CKLP region of Esrp1 from different species indicates high conservation.

(B) Schematic of the reporter vector and the peptide sequences used for transfection and microscopy.

(C) Representative images of HeLa cells (20x) transfected with different reporter vectors depicted in (B), from which we concluded that the shortest peptide (P1) is sufficient for nuclear localization.

3.2.3 Determination of key residues in the Esrp1 NLS that are required for the nuclear localization.

After identification of a 15-amino acid peptide sequence sufficient for nuclear localization, we sought to further define the key amino acids within the element that are required for nuclear import. We therefore introduced single amino acid mutations at each position in the Esrp1 NLS and examined their effects on subcellular localization using the chicken pyruvate kinase vector. Mutation of CKLP to AAAA completely abolished the nuclear localization (Fig. 3-3A). Additionally, we identified seven amino acids for which single alanine substitution led to nearly complete cytoplasmic localization, five amino acids that displayed a mixed cytoplasmic and nuclear localization after alanine substitution, and three that remained nuclear after mutation (Fig. 3-3A). Since Leucine and Valine are both nonpolar amino acids and share a similar structure, we also mutated L to V at 2nd and 9th position. Interestingly, L2A and L2V both showed a mixed phenotype, in contrast, while L9A completely abolished nuclear localization, L9V remained nuclear, suggesting L9V preserves NLS function (Fig. 3-3B). It has been reported that phosphorylation within NLSs can both positively or negatively affect nuclear import (Kaffman and O'Shea 1999; Jans et al. 2000). The Serine residue at the third position in the Esrp1 NLS is a known phosphorylation site that is also conserved in human (Van Hoof et al. 2009; Rigbolt et al. 2011; Klammer et al. 2012; Mertins et al. 2014). Therefore, we generated a mutation which mimics the phosphorylated status of serine residue. While S3A exhibited a mixed cytoplasmic and nuclear localization, S3D restored nearly complete nuclear import, suggesting that phosphorylation of Serine at this position may play a role in facilitating the nuclear import of Esrp1 (Fig. 3-3B). It will be of interest in future studies to determine conditions under which phosphorylation at this position might regulate Esrp1 localization. Since “Glycine” at the first position is not required for the nuclear localization, the minimal Esrp1 NLS is “LSPPPCKLPCLSPP” (Fig. 3-3C). All mutants are summarized in Table 3-1.

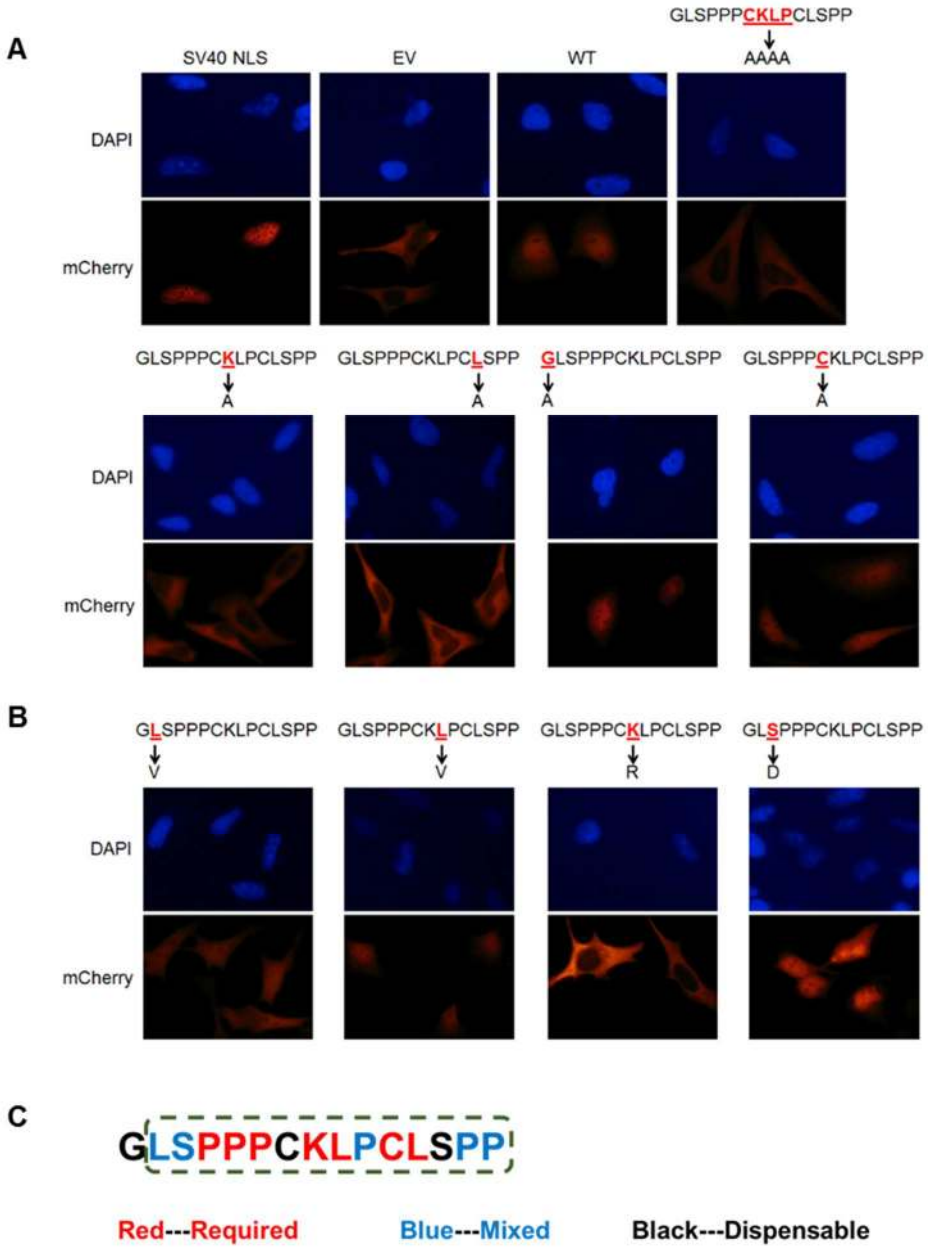


Figure 3-3 Determination of key residues in the Esrp1 NLS that are required for nuclear localization.

(A) Examples of amino acids where alanine substitution abolished nuclear localization (K8A and L12A) and didn't affect nuclear localization (G1A and C7A).

(B) Differential effect of various point mutation on nuclear localization.

(C) Summary of key residues in Esrp1 NLS based on alanine substitution. Since the first Glycine is not required for nuclear localization, "LSPPPCKLPCLSPP" is the Esrp1 NLS.

Table 3-1 Summary of the effect of all tested mutants on nuclear localization.

Mutation	Localization	Mutation	Localization
G1A	Nuclear	K8R	Cytoplasmic
L2A	Mixed	L9A	Cytoplasmic
L2V	Mixed	L9V	Nuclear
S3A	Mixed	P10A	Mixed
S3D	Nuclear	C11A	Cytoplasmic
P4A	Cytoplasmic	L12A	Cytoplasmic
P5A	Cytoplasmic	S13A	Nuclear
P6A	Cytoplasmic	S13D	Nuclear
C7A	Nuclear	P14A	Mixed
K8A	Cytoplasmic	P15A	Mixed

3.2.4 Fusilli, the D. Melanogaster orthologue of Esrp1 also expresses both nuclear and cytoplasmic isoforms that result from alternative splicing.

The conservation of the alternative splicing event leading to both nuclear and cytoplasmic isoforms among Esrp1 orthologs in vertebrates suggested functional roles for both protein isoforms in post-transcriptional regulation. We sought to further investigate the degree of conservation for the expression of nuclear and cytoplasmic Esrp1 orthologs through alternative splicing by studying the subcellular localization of different isoforms of fusilli, the Esrp1 ortholog in D. Melanogaster. We previously showed that fusilli can regulate alternative splicing of some Esrp1 targets when ectopically expressed in mammalian cells (Warzecha et al. 2009a). We also identified numerous endogenous AS events switched upon ectopic expression of fusilli in a D. Melanogaster cell line (data not shown), suggesting a conserved role for fusilli in regulation of splicing. We noted that the fusilli gene has several alternative promoters and splicing events that generate multiple protein isoforms (Fig 3-4A). Three splicing mRNA variants (A, B and H) result from a distal promoter and lack the majority of the first highly conserved RRM and therefore were not studied further. All other transcript variants encoding longer isoforms that contain all three RRMs are derived from alternative splicing, including a 93 nt retained intron within exon 9 and an alternative 5' splice site coupled to alternative polyadenylation (APA5) that leads to distinct C-

termini. We were able to generate cDNAs encoding isoform D that contains the retained intron as well as isoform G that utilizes the alternative 5' splice site (Fig. 3-4A). Transfection of cDNAs encoding both isoforms as mCherry fusion proteins in HeLa cells showed that isoform D was nuclear while isoform G was predominantly cytoplasmic (Fig. 3-4B). To determine whether a presumed NLS in isoform D was present within the region encoded by the retained intron or the C-terminus we made a truncated form of isoform D lacking the region encoded by the distinct C-terminus and determined that it was cytoplasmic, similar to isoform G. We therefore concluded that an NLS resides in the C-terminus of isoform D, not the retained intron. (Fig. 3-4B). To further pinpoint the NLS, we generated multiple truncations from the C-terminus of isoform D and tested their subcellular localization (Fig. 3-5A). While truncation 1 preserved the nuclear localization, all subsequent truncations generated isoforms that were cytoplasmic. Therefore, we narrowed down the NLS to the region between the end of truncation 2 and truncation 1, which is a 37 amino acid fragment (peptide 1). To guide our determination of the minimal peptide sequence that is sufficient for the nuclear localization, we aligned this peptide sequences in 12 different fly species as well as other insects and noted the last 14 amino acids (peptide 3) are highly conserved compared to the first 23 amino acids (peptide 2) (Fig. 3-5B). Using the previously described chicken pyruvate kinase reporter transfected into HeLa cells, we determined that peptide 3 "QSMKRSYENAFQQE" is sufficient for the nuclear localization, thereby accounting for nuclear localization of isoforms that contain this NLS in fusilli (Fig.3-5C). These observations thus indicate that while the sequences used to derive both nuclear and cytoplasmic isoforms of Fusilli differ from that in vertebrate Esrp1, there has been functional conservation of the mechanism to maintain both isoforms via alternative splicing, suggesting the biological significance over a long period of evolutionary history.

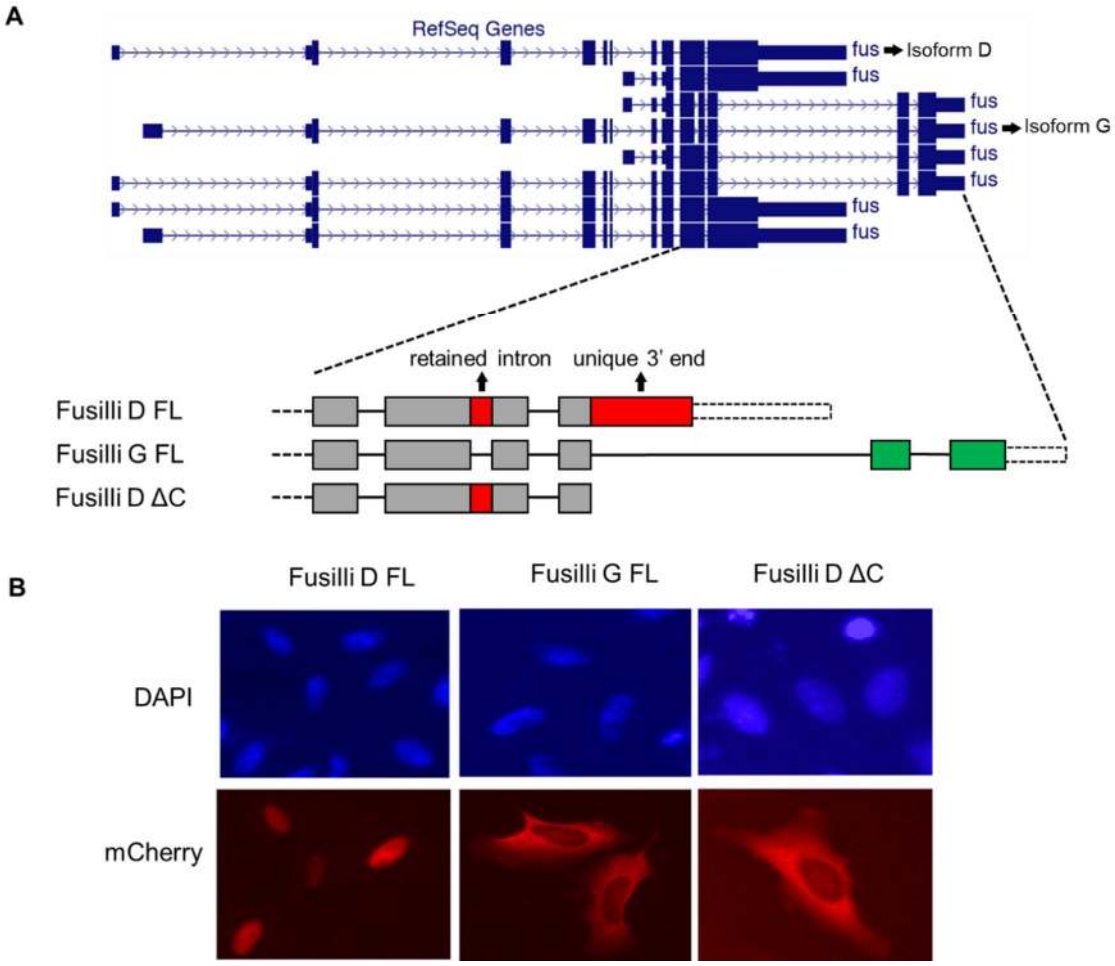


Figure 3-4 Fusilli expresses both nuclear and cytoplasmic isoforms as a result of alternative splicing.

(A) A screenshot of the genome browser view with all fusilli isoforms and schematic of isoform D and G. Fusilli D ΔC represents the C-terminal truncation of isoform D.

(B) Representative images of HeLa cells (20x) transfected with expression vectors of different fusilli isoforms as mCherry fusion depicted in (A), from which we concluded that the C-terminus is required for nuclear localization.

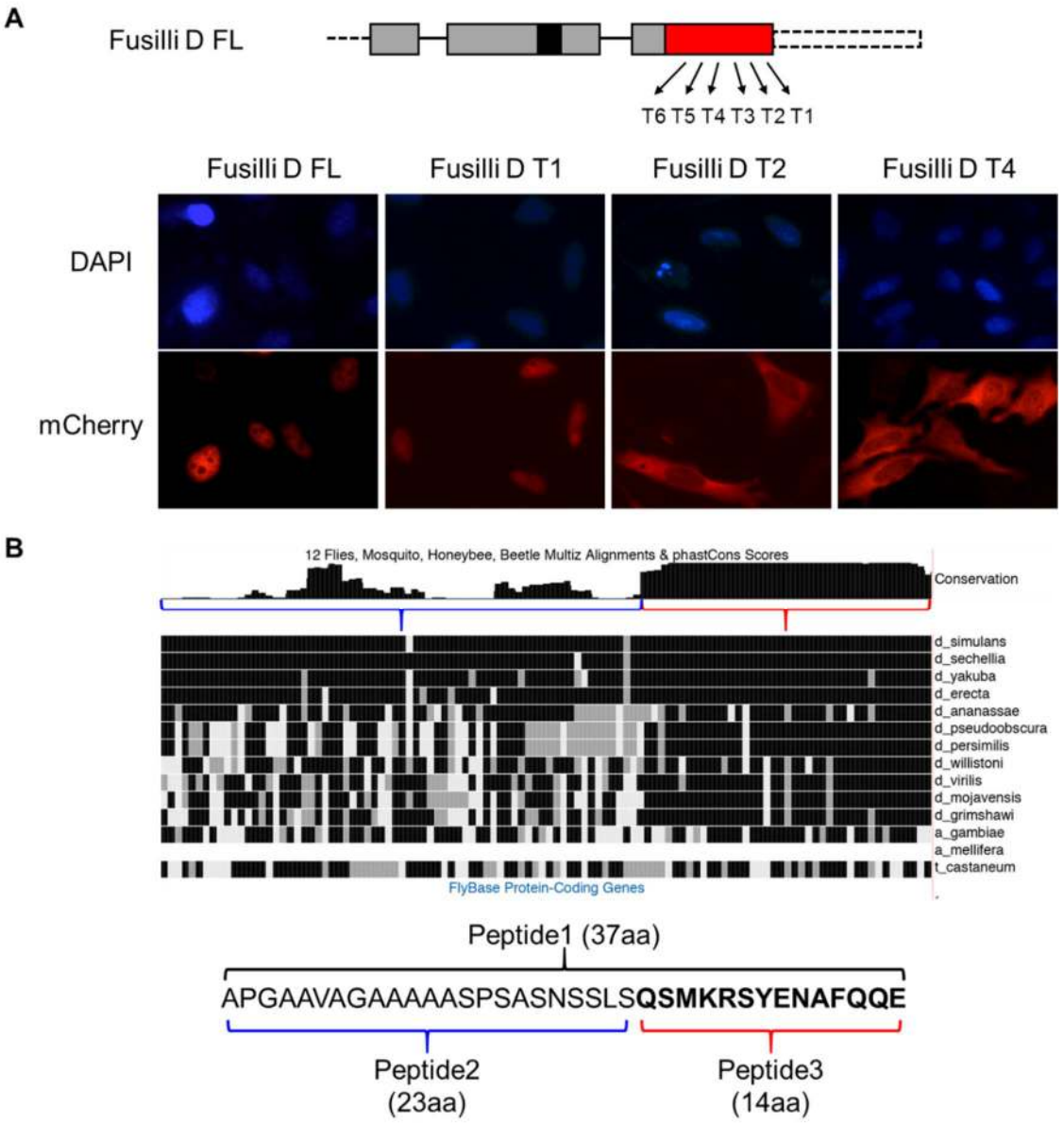

A
Fusilli D FL
T6 T5 T4 T3 T2 T1
Fusilli D FL
Fusilli D T1
Fusilli D T2
Fusilli D T4
DAPI
mCherry
B
12 Flies, Mosquito, Honeybee, Beetle Multiz Alignments & phastCons Scores
Conservation
d_simulans
d_sechellia
d_yakuba
d_erecta
d_ananassae
d_pseudoobscura
d_persimilis
d_willistoni
d_virilis
d_mojavensis
d_grimshawi
a_gambiae
a_mellifera
t_castaneum
FlyBase Protein-Coding Genes
Peptide1 (37aa)
APGAAVAGAAAAASPSASNSSLSQSMKRSYENAFQQE
Peptide2
(23aa)
Peptide3
(14aa)
C
EV
Peptide1
Peptide2
Peptide3
DAPI
mCherry

Figure 3-5 Mapping of fusilli NLS.

(A) Representative images of HeLa cells (20x) transfected with expression vectors of multiple truncations from the C-terminus of isoform D, from which we concluded that fusilli NLS resides in the region between T1 and T2.

(B) Alignment of sequences between T1 and T2 of fusilli from different species reveal a highly conserved peptide sequence (peptide 3).

(C) Representative images of HeLa cells (20x) transfected with different reporter vectors confirmed that peptide 3 is sufficient for nuclear localization.

3.3 Discussion and future directions

Post-transcriptional regulation is a complex process where a group of functionally related mRNAs are coordinately regulated at many steps after transcription, including 5' capping, splicing, editing, polyadenylation, mRNA transport, localization, mRNA stability and translation, by RBPs and/or microRNAs, two important classes of trans-acting factors in mRNA processing (Keene 2007). While microRNAs primarily regulate gene expression via mRNA degradation and translation inhibition, it has become increasingly appreciated that many RBPs are multifunctional, regulating more than one aspect of mRNA processing (Keene 2007). Recently, it has become increasingly appreciated that many RBPs are multifunctional, regulating more than one aspect in mRNA processing and some of them are able to do so by generating different isoforms in nucleus and cytoplasm through alternative splicing.

Our study adds Esrp1 and its orthologs to the list of RBPs that are known to expand their potential post-transcriptional regulatory functions through AS to generate both nuclear and cytoplasmic isoforms. The conservation in the fruit fly ortholog fusilli strongly suggests that there is an important biological function for both isoforms. While the alternative 5' splice site only leads to a difference by four amino acids (CKLP) among isoforms, the functional consequences can be substantial. A well-known example, the Wilm's Tumor (WT1) gene produces two protein isoforms that differ by only three amino acids (lysine-threonine-serine, or KTS), WT1 +KTS and WT1 -KTS, which are encoded by transcript variants that are alternatively spliced through the usage of two adjacent 5' splice sites in the pre-mRNA (Haber et al. 1991). While the WT1 -KTS is a transcription factor that binds to DNA and act as either transcriptional activator or repressor, the WT1 +KTS isoforms has been suggested to bind to RNA and regulate splicing or other post-transcriptional processes (Larsson et al. 1995; Hammes et al. 2001; Markus et al. 2006; Huff 2011). Since the 5' splice sites in Esrp1 that lead to +CKLP and –CKLP isoforms share the same 5' splice site consensus sequence (CAT|GTAATG) and are separated by only 3 nucleotides, we initially surmised that the choice of either site during splicing was stochastic. However, we

observed an increase in the ratio of mRNAs encoding nuclear to cytoplasmic isoforms when ESRPs are depleted from the human H358 cell line and ablated in the mouse epidermis, which suggests the alternative splicing event is under regulation. We hypothesize that this represents autoregulation wherein high levels of Esrp1 in the nucleus promote an increase in expression of the cytoplasmic isoform. Regulation of such alternative splicing event can be fundamental for development. For example, there is a decrease in the inclusion of exon 5 in MBNL1 during cardiac morphogenesis (Terenzi and Ladd 2010). Exon 5 is functionally important in regulating the subcellular localization as well as function of MNBL1. Specifically, isoforms containing exon 5 are predominantly nuclear while the loss of exon 5 leads to a cytoplasmic localization. Therefore, the nuclear localization of MBNL1 in the early stage of heart development is required for alternative splicing regulation and the loss of exon 5 is essential for the fetal-to-adult transitions of alternative splicing. Moreover, the alternative splicing of exon 19 in Rbfox1 is also regulated by nuclear Rbfox1 upon neuronal depolarization which leads to a switch from a cytoplasmic to nuclear isoform (Lee et al. 2009).

While we have extensively characterized the role of Esrp1 in alternative splicing, the function of cytoplasmic Esrp1 remains unexplored. A recent study identified conserved Esrp binding motifs enriched in 3'UTRs and proposed a role for Esrp1 in the regulation of mRNA stability based on the correlation between the expression levels of Esrp1 and mRNAs with Esrp1 binding motifs in their 3' UTRs among different tissues, suggesting a role of Esrp1 in addition to splicing regulation (Ray et al. 2013b). Similar to RBFOX1 where the nuclear and cytoplasmic isoforms regulate a different set of genes through distinct mechanisms yet still coordinately contribute to neuronal development, we expect the cytoplasmic Esrp1 to also play a role in EMT regulation and maintenance of epithelial cell functions. However, direct experimental evidence to support this and other possible regulatory roles for Esrp1 in cytoplasm is required. To identify the direct mRNA targets and binding sites in the cytoplasm for Esrp1 *in vivo*, experiments such as RIP-Seq and/or CLIP-Seq in the cytoplasmic fraction will provide valuable information. To study the potential function of cytoplasmic Esrp1 in regulating mRNA stability, experiments such as

traditional inhibition of transcription by actinomycin D followed by RNA-Seq or metabolic labeling of nascent mRNA by pulsed nucleotide analogs (such as 4-thiouridine) followed by RNA-Seq are required. For studies to determine whether Esrp1 binding can regulate translation experiments such as ribosome profiling or polysome profiling could be conducted (Rabani et al. 2011; Ingolia et al. 2012).

A nuclear localization signal (NLS) is a short stretch of amino acids that mediates the transport of proteins into the nucleus (Lange et al. 2007). The best characterized "classical NLS" (cNLS) consists of either one (monopartite) or two (bipartite) stretches of basic amino acids, for examples the well-known SV40 large antigen NLS and the nucleoplasmin NLS (Kalderon et al. 1984; Robbins et al. 1991). Proteins containing a cNLS are translocated by the importin-α/β heterodimer. Importin-α (Impα; also known as Karyopherin-α) is a protein adaptor that directly binds to the cNLS and importin-β (Impβ; also known as Karyopherin-β1), which is associated with nuclear pore complex (NPC) to facilitate the nuclear import (Lange et al. 2007). The other well-characterized NLS is the "PY-NLS". While the sequences for PY-NLSs show limited similarity, they usually consist of a loose N-terminal hydrophobic or basic motif and a C-terminal $RX_{2-5}PY$ motif (Lee et al. 2006; Süel et al. 2008; Xu et al. 2010). Proteins containing a PY-NLS, for example HnRNP A1, are imported by Karyopherin-β2 (Kapβ2) (Pollard et al. 1996; Bonifaci et al. 1997). However, many NLSs have been identified with highly diverse sequences that do not belong to either class (Soniat and Chook 2015). In addition, only a limited number of transport cargos are known for other importins (Xu et al. 2010). The NLS sequences identified in this study for both Esrp1 and fusilli don't conform to any of the well-characterized NLS consensus sequences and potentially represent a novel class of NLS. It will also be of interest to determine whether the Esrp1 NLS represents a larger class of NLS motifs and to define the import pathway that it uses to translocate to the nucleus to regulate splicing.

3.4 Material and methods

3.4.1 Plasmids and cloning

The pIBX-C-FF(B) was described previously (Warzecha et al. 2009a). We PCR amplified the cDNA sequence for mCherry and cloned into the vector using NheI and NsiI to make the pIBX-C-FF(B)-mCherry vector. We then PCR amplified the cDNA sequences for different isoforms of Esrp1 and cloned into the using EcoRV and NotI sites to make the expression vectors. pCMV-myc-PK vector was described previously (Siomi and Dreyfuss 1995). PCR amplified coding sequence for EGFP was inserted into NheI and NsiI digested pCMV-myc-PK to drive pCMV-myc-PK-EGFP. We then PCR amplified the coding sequence for 2x FLAG tag followed by SV40NLS and cloned it downstream of pyruvate kinase coding sequence and upstream of EGFP coding sequence to drive pCMV-myc-PK-FF-NLS-EGFP using Not I and NheI sites. We replaced the coding sequence for EGFP with that for mCherry to generate pCMV-myc-PK-FF-NLS-mCherry. We digested the vector with NotI and NheI in order to clone in other sequences. To determine the minimum peptide sequences sufficient for Esrp1 nuclear localization, we PCR amplified coding sequence for peptide 1 and 2 from Esrp1 +CKLP and Esrp1 -CKLP isoforms and used them to swap the coding sequence for 2XFLAG tag and SV40NLS. Since peptide 3 is relatively short, for the WT peptide 3 and all tested mutants, the corresponding sense and antisense oligoes were annealed and ligated into the cut reporter vector. pIBX-C-FF(B)-fusilli D has been described before (Warzecha et al. 2009a). We cloned in mCherry coding sequence into NheI and NsiI digested pIBX-C-FF(B)-fusilli D to drive pIBX-C-FF(B)-fusilli D-mCherry. A series of reverse primer in the C-terminus of fusilli D and the universal forward T7 primer were used to make all the C-terminal truncation proteins. Fusilli G was made by gene synthesis. To determine the minimum peptide sequences sufficient for fusilli nuclear localization, for peptide 1, 2 and 3, the corresponding sense and antisense oligoes were annealed and ligated into the cut reporter vector. Primers used for cloning are summarized in Table A3-1.

3.4.2 RNA extraction and RT-PCR

Total RNA was extracted using TRIzol (15596018, Life technologies). Reverse transcription and semi-quantitative PCR was performed as described previously (Warzecha et al. 2009a). Primers used to detect the nuclear and cytoplasmic isoforms are in Table A3-1. Mouse epidermis tissue is from our previous work (Bebee et al. 2015).

3.4.3 Cell culture and transfection

HeLa cells were maintained in DMEM medium with 10% FBS (SH30071.03, GE). For transfection, HeLa cells were seeded on coverslips in 6-well plates at 200,000 cells per well and incubated overnight. Cells were transfected using Lipofectamine® 2000 (11668027, Life technologies) according the manufacturers' protocols. 2ug of plasmids were used for each transfection.

3.4.4 Microscopy

24 hours post transfection, the coverslip containing transfected cells was washed once with cold 1xPBS, the cells were then fixed with acetone for three minutes and washed with cold 1XPBS for four times. The coverslips were mounted to a clear glass slide using mounting media with DAPI staining (P36931, Thermo Scientific). For each transfection, at least three randomly chosen fields were imaged.

APPENDIX

Figure A2-1 Validation of exons that are RBM47-regulated using a siRNA that targets different sequences from those of the shRNA used for RNA-Seq experiment.

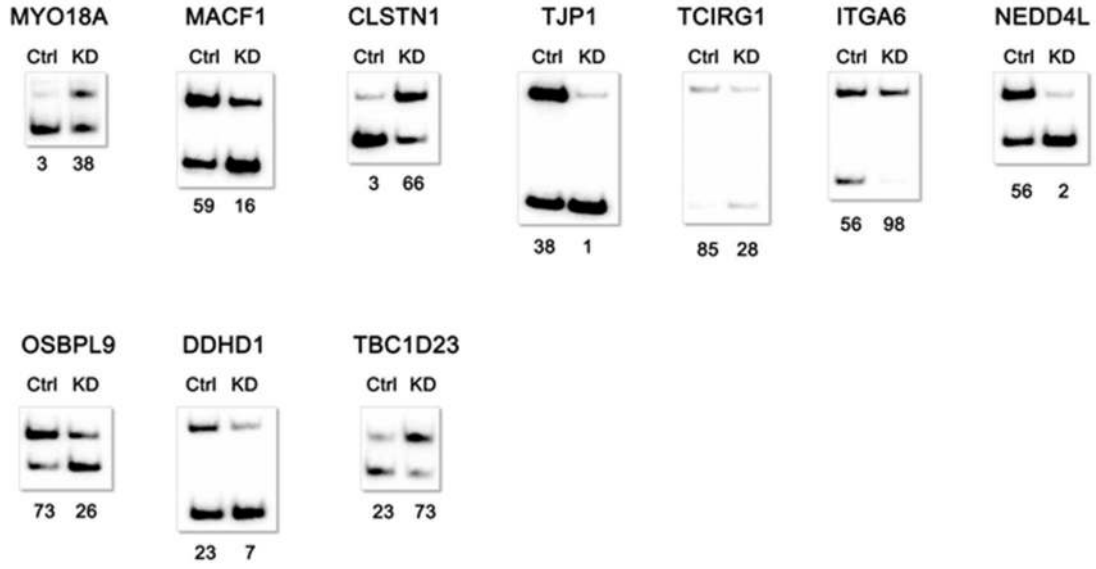


Figure A2-2 Validation of representative exons where QKI promotes the mesenchymal splicing patterns upon knockdown with siQKI #7.

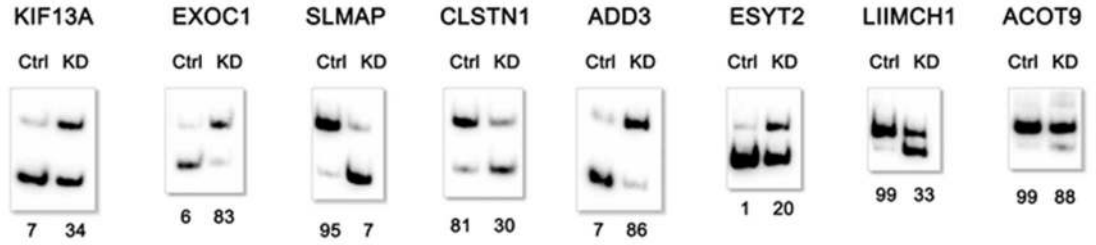
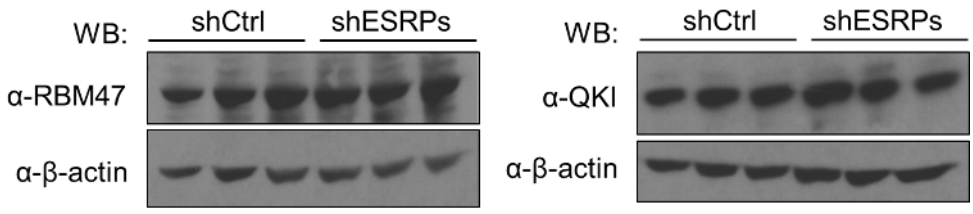


Figure A2-3 The protein levels of RBM47 (Left) and QKI (Right) upon ESRP1/2 depletion.



BIBLIOGRAPHY

Abdelmohsen K, Gorospe M. 2010. Posttranscriptional regulation of cancer traits by HuR. *Wiley Interdisciplinary Reviews - RNA* **1**: 214-229.

Akaike Y, Masuda K, Kuwano Y, Nishida K, Kajita K, Kurokawa K, Satake Y, Shoda K, Imoto I, Rokutan K. 2014. HuR Regulates Alternative Splicing of the TRA2β Gene in Human Colon Cancer Cells under Oxidative Stress. *Molecular and Cellular Biology* **34**: 2857-2873.

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.

Anderson ES, Lin CH, Xiao X, Stoilov P, Burge CB, Black DL. 2012. The cardiotonic steroid digitoxin regulates alternative splicing through depletion of the splicing factors SRSF3 and TRA2B. *Rna* **18**: 1041-1049.

Ast G. 2004. How did alternative splicing evolve? *Nat Rev Genet* **5**: 773-782.

Barnett TR, Kretschmer A, Austen DA, Goebel SJ, Hart JT, Elting JJ, Kamarck ME. 1989. Carcinoembryonic Antigens - Alternative Splicing Accounts for the Multiple Messenger-Rnas That Code for Novel Members of the Carcinoembryonic Antigen Family. *Journal of Cell Biology* **108**: 267-276.

Bebee TW, Park JW, Sheridan KI, Warzecha CC, Cieply BW, Rohacek AM, Xing Y, Carstens RP. 2015. The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. *Elife* **4**.

Berget SM. 1995. Exon Recognition in Vertebrate Splicing. *Journal of Biological Chemistry* **270**: 2411-2414.

Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 3171-3175.

Beyer AL, Osheim YN. 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes & Development* **2**: 754-765.

Bhalla K, Phillips HA, Crawford J, McKenzie OLD, Mulley JC, Eyre H, Gardner AE, Kremmidiotis G, Callen DF. 2004. The de novo chromosome 16 translocations of two patients with abnormal phenotypes (mental retardation and epilepsy) disrupt the A2BP1 gene. *J Hum Genet* **49**: 308-311.

Bill BR, Lowe JK, DyBuncio CT, Fogel BL. 2013. Chapter Eight - Orchestration of Neurodevelopmental Programs by RBFOX1: Implications for Autism Spectrum Disorder. In *International Review of Neurobiology*, Vol Volume 113 (ed. K Genevieve), pp. 251-267. Academic Press.

Blencowe BJ. 2000. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends in Biochemical Sciences* **25**: 106-110.

Bonifaci N, Moroianu J, Radu A, Blobel G. 1997. Karyopherin β2 mediates nuclear import of a mRNA binding protein. *Proceedings of the National Academy of Sciences* **94**: 5055-5060.

Bonnal S, Martínez C, Förch P, Bachi A, Wilm M, Valcárcel J. 2008. RBM5/Luca-15/H37 Regulates Fas Alternative Splice Site Pairing after Exon Definition. *Molecular Cell* **32**: 81-95.

Boutet A, De Frutos CA, Maxwell PH, Mayol MJ, Romero J, Nieto MA. 2006. Snail activation disrupts tissue homeostasis and induces fibrosis in the adult kidney. *The EMBO Journal* **25**: 5603-5613.

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

Brennan CM, Steitz JA. 2001. HuR and mRNA stability. *Cell Mol Life Sci* **58**: 266-277.

Brown RL, Reinke LM, Damerow MS, Perez D, Chodosh LA, Yang J, Cheng CH. 2011. CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *J Clin Invest* **121**: 1064-1074.

Carrillo Oesterreich F, Herzel L, Straube K, Hujer K, Howard J, Neugebauer Karla M. Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell* **165**: 372-381.

Cech TR. 1990. Self-splicing of group I introns. *Annu Rev Biochem* **59**: 543-568.

Chang S-H, Elemento O, Zhang J, Zhuang ZW, Simons M, Hla T. 2014. ELAVL1 regulates alternative splicing of eIF4E transporter to promote postnatal angiogenesis. *Proceedings of the National Academy of Sciences* **111**: 18309-18314.

Chaudhury A, Chander P, Howe PH. 2010. Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles. *RNA* **16**: 1449-1462.

Chen M, Manley JL. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature reviews Molecular cell biology* **10**: 741-754.

Cho S, Hoang A, Sinha R, Zhong X-Y, Fu X-D, Krainer AR, Ghosh G. 2011. Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proceedings of the National Academy of Sciences* **108**: 8233-8238.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**: 1-8.

Cieply B, Carstens RP. 2015. Functional roles of alternative splicing factors in human disease. *Wiley Interdiscip Rev RNA* **6**: 311-326.

Conn Simon J, Pillman Katherine A, Toubia J, Conn Vanessa M, Salmanidis M, Phillips Caroline A, Roslan S, Schreiber Andreas W, Gregory Philip A, Goodall Gregory J. 2015a. The RNA Binding Protein Quaking Regulates Formation of circRNAs. *Cell* **160**: 1125-1134.

Conn SJ, Pillman KA, Toubia J, Conn VM, Salmanidis M, Phillips CA, Roslan S, Schreiber AW, Gregory PA, Goodall GJ. 2015b. The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**: 1125-1134.

Cooper TA, Wan L, Dreyfuss G. 2009. RNA and Disease. *Cell* **136**: 777-793.

Crawford JB, Patton JG. 2006. Activation of α-Tropomyosin Exon 2 Is Regulated by the SR Protein 9G8 and Heterogeneous Nuclear Ribonucleoproteins H and F. *Molecular and Cellular Biology* **26**: 8791-8802.

Culbertson MR, Winey M. 1989. Split tRNA genes and their products: a paradigm for the study of cell function and evolution. *Yeast* **5**: 405-427.

Das R, Dufu K, Romney B, Feldt M, Elenko M, Reed R. 2006. Functional coupling of RNAP II transcription to spliceosome assembly. *Genes & Development* **20**: 1100-1109.

Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, Gygi SP, Reed R. 2007. SR Proteins Function in Coupling RNAP II Transcription to Pre-mRNA Splicing. *Molecular Cell* **26**: 867-881.

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 & Development* **25**: 972-983.

David CJ, Chen M, Assanah M, Canoll P, Manley JL. 2010. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* **463**: 364-368.

De Conti L, Baralle M, Buratti E. 2013. Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip Rev RNA* **4**: 49-60.

de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein Ma, Pelisch F, Cramer P, Bentley D, Kornblihtt AR. 2003. A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. *Molecular Cell* **12**: 525-532.

Dittmar KA, Jiang P, Park JW, Amirikian K, Wan J, Shen S, Xing Y, Carstens RP. 2012. Genome-wide determination of a broad ESRP-regulated posttranscriptional network by high-throughput sequencing. *Mol Cell Biol* **32**: 1468-1482.

Doukhanine E, Gavino C, Haines JD, Almazan G, Richard S. 2010. The QKI-6 RNA Binding Protein Regulates Actin-interacting Protein-1 mRNA Stability during Oligodendrocyte Differentiation. *Molecular Biology of the Cell* **21**: 3029-3040.

Dráberová L, Černá H, Brodská H, Boubelík M, Watt SM, Stanners CP, Dráber P. 2000. Soluble isoforms of CEACAM1 containing the A2 domain: increased serum levels in patients with

obstructive jaundice and differences in 3-fucosyl-N-acetyl-lactosamine moiety. *Immunology* **101**: 279-287.

Duband JL, Thiery JP. 1982. Appearance and distribution of fibronectin during chick embryo gastrulation and neurulation. *Developmental Biology* **94**: 337-350.

Eperon IC, Makarova OV, Mayeda A, Munroe SH, Cáceres JF, Hayward DG, Krainer AR. 2000. Selection of Alternative 5′ Splice Sites: Role of U1 snRNP and Models for the Antagonistic Effects of SF2/ASF and hnRNP A1. *Molecular and Cellular Biology* **20**: 8303-8318.

Expert-Bezançon A, Sureau A, Durosay P, Salesse R, Groeneveld H, Lecaer JP, Marie J. 2004. hnRNP A1 and the SR Proteins ASF/SF2 and SC35 Have Antagonistic Functions in Splicing of β-Tropomyosin Exon 6B. *Journal of Biological Chemistry* **279**: 38249-38259.

Fan XC, Steitz JA. 1998. HNS, a nuclear-cytoplasmic shuttling sequence in HuR. *Proceedings of the National Academy of Sciences* **95**: 15293-15298.

Fischer KR, Durrans A, Lee S, Sheng J, Li F, Wong STC, Choi H, El Rayes T, Ryu S, Troeger J et al. 2015. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**: 472-476.

Fogel BL, Wexler E, Wahnich A, Friedrich T, Vijayendran C, Gao F, Parikshak N, Konopka G, Geschwind DH. 2012. RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Human Molecular Genetics* **21**: 4171-4186.

Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saldi T, Diener K, Jones K, Fu X-D, Bentley DL. 2014. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes & Development* **28**: 2663-2676.

Fossat N, Tourle K, Radziewic T, Barratt K, Liebhold D, Studdert JB, Power M, Jones V, Loebel DA, Tam PP. 2014. C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. *EMBO Rep* **15**: 903-910.

Galarneau A, Richard S. 2005. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat Struct Mol Biol* **12**: 691-698.

Gardner K, Bennett V. 1987. Modulation of spectrin-actin assembly by erythrocyte adducin. *Nature* **328**: 359-362.

Gaur S, Shively J, Yen Y, Gaur R. 2008. Altered splicing of CEACAM1 in breast cancer: Identification of regulatory sequences that control splicing of CEACAM1 into long or short cytoplasmic domain isoforms. *Molecular Cancer* **7**: 46.

Gehman LT, Meera P, Stoilov P, Shiue L, O'Brien JE, Meisler MH, Ares M, Otis TS, Black DL. 2012. The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. *Genes & Development* **26**: 445-460.

Geuens T, Bouhy D, Timmerman V. 2016. The hnRNP family: insights into their role in health and disease. *Human Genetics* **135**: 851-867.

Görnemann J, Kotovic KM, Hujer K, Neugebauer KM. 2005. Cotranscriptional Spliceosome Assembly Occurs in a Stepwise Fashion and Requires the Cap Binding Complex. *Molecular Cell* **19**: 53-63.

Grande MT, Sanchez-Laorden B, Lopez-Blau C, De Frutos CA, Boutet A, Arevalo M, Rowe RG, Weiss SJ, Lopez-Novoa JM, Nieto MA. 2015. Snail1-induced partial epithelial-to-mesenchymal transition drives renal fibrosis in mice and can be targeted to reverse established disease. *Nat Med* **21**: 989-997.

Graveley BR. 2000. Sorting out the complexity of SR protein functions. *RNA* **6**: 1197-1211.

Graveley BR. 2005. Mutually Exclusive Splicing of the Insect Dscam Pre-mRNA Directed by Competing Intronic RNA Secondary Structures. *Cell* **123**: 65-73.

Greer CL, Abelson J. 1984. RNA splicing: rearrangement of RNA sequences in the expression of split genes. *Trends in Biochemical Sciences* **9**: 139-141.

Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y, Goodall GJ. 2008. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* **10**: 593-601.

Gu B, Eick D, Bensaude O. 2013. CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Research* **41**: 1591-1603.

Guo M, Lo PC, Mount SM. 1993. Species-specific signals for the splicing of a short Drosophila intron in vitro. *Molecular and Cellular Biology* **13**: 1104-1118.

Haber DA, Sohn RL, Buckler AJ, Pelletier J, Call KM, Housman DE. 1991. Alternative splicing and genomic structure of the Wilms tumor gene WT1. *Proc Natl Acad Sci U S A* **88**: 9618-9622.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**: 129-141.

Hall MP, Nagel RJ, Fagg WS, Shiue L, Cline MS, Perriman RJ, Donohue JP, Ares M, Jr. 2013. Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *RNA* **19**: 627-638.

Hamid Fursham M, Makeyev Eugene V. 2014. Emerging functions of alternative splicing coupled with nonsense-mediated decay. *Biochemical Society Transactions* **42**: 1168-1173.

Hammes A, Guo J-K, Lutsch G, Leheste J-R, Landrock D, Ziegler U, Gubler M-C, Schedl A. 2001. Two Splice Variants of the Wilms' Tumor 1 Gene Have Distinct Functions during Sex Determination and Nephron Formation. *Cell* **106**: 319-329.

Hardy RJ, Loushin CL, Friedrich VL, Chen Q, Ebersole TA, Lazzarini RA, Artzt K. 1996. Neural cell type-specific expression of QKI proteins is altered in quakingviable mutant mice. *J Neurosci* **16**: 7941-7949.

Hastings ML, Krainer AR. 2001. Pre-mRNA splicing in the new millennium. *Current Opinion in Cell Biology* **13**: 302-309.

Horiguchi K, Sakamoto K, Koinuma D, Semba K, Inoue A, Inoue S, Fujii H, Yamaguchi A, Miyazawa K, Miyazono K et al. 2012. TGF-beta drives epithelial-mesenchymal transition through deltaEF1-mediated downregulation of ESRP. *Oncogene* **31**: 3190-3201.

House AE, Lynch KW. 2006. An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nat Struct Mol Biol* **13**: 937-944.

Hsin J-P, Manley JL. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & Development* **26**: 2119-2137.

Huber M, Izzi L, Grondin P, Houde C, Kunath T, Veillette A, Beauchemin N. 1999. The carboxyl-terminal region of biliary glycoprotein controls its tyrosine phosphorylation and association with protein-tyrosine phosphatases SHP-1 and SHP-2 in epithelial cells. *Journal of Biological Chemistry* **274**: 335-344.

Huff V. 2011. Wilms' tumours: about tumour suppressor genes, an oncogene and a chameleon gene. *Nat Rev Cancer* **11**: 111-121.

Humphreys BD, Lin S-L, Kobayashi A, Hudson TE, Nowlin BT, Bonventre JV, Valerius MT, McMahon AP, Duffield JS. 2010. Fate Tracing Reveals the Pericyte and Not Epithelial Origin of Myofibroblasts in Kidney Fibrosis. *The American Journal of Pathology* **176**: 85-97.

Huttlin EL, Jedrychowski MP, Elias JE, Goswami T, Rad R, Beausoleil SA, Villén J, Haas W, Sowa ME, Gygi SP. 2010. A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. *Cell* **143**: 1174-1189.

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature protocols* **7**: 1534-1550.

Iwano M, Plieth D, Danoff TM, Xue C, Okada H, Neilson EG. 2002. Evidence that fibroblasts derive from epithelium during tissue fibrosis. *The Journal of Clinical Investigation* **110**: 341-350.

Izquierdo JM. 2008. Hu Antigen R (HuR) Functions as an Alternative Pre-mRNA Splicing Regulator of Fas Apoptosis-promoting Receptor on Exon Definition. *Journal of Biological Chemistry* **283**: 19077-19084.

Izquierdo JM, Majós N, Bonnal S, Martínez C, Castelo R, Guigó R, Bilbao D, Valcárcel J. 2005. Regulation of Fas Alternative Splicing by Antagonistic Effects of TIA-1 and PTB on Exon Definition. *Molecular Cell* **19**: 475-484.

Jangi M, Boutz PL, Paul P, Sharp PA. 2014. Rbfox2 controls autoregulation in RNA-binding protein networks. *Genes Dev* **28**: 637-651.

Jans DA, Xiao C-Y, Lam MHC. 2000. Nuclear targeting signal recognition: a key control point in nuclear transport? *BioEssays* **22**: 532-544.

Jensen KB, Dredge BK, Stefani G, Zhong R, Buckanovich RJ, Okano HJ, Yang YYL, Darnell RB. 2000. Nova-1 Regulates Neuron-Specific Alternative Splicing and Is Essential for Neuronal Viability. *Neuron* **25**: 359-371.

Kaffman A, O'Shea EK. 1999. Regulation of nuclear localization: a key to a door. *Annu Rev Cell Dev Biol* **15**: 291-339.

Kaiser HW, O'Keefe E, Bennett V. 1989. Adducin: Ca++-dependent association with sites of cell-cell contact. *J Cell Biol* **109**: 557-569.

Kalderon D, Roberts BL, Richardson WD, Smith AE. 1984. A short amino acid sequence able to specify nuclear location. *Cell* **39**: 499-509.

Kalluri R, Weinberg RA. 2009. The basics of epithelial-mesenchymal transition. *The Journal of Clinical Investigation* **119**: 1420-1428.

Keene JD. 2007. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**: 533-543.

Keren H, Lev-Maor G, Ast G. 2010. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* **11**: 345-355.

Klammer M, Kaminski M, Zedler A, Oppermann F, Blencke S, Marx S, Müller S, Tebbe A, Godl K, Schaab C. 2012. Phosphosignature Predicts Dasatinib Response in Non-small Cell Lung Cancer. *Molecular & Cellular Proteomics* **11**: 651-668.

Knecht AK, Bronner-Fraser M. 2002. Induction of the neural crest: a multigene process. *Nat Rev Genet* **3**: 453-461.

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. *Molecular and Cellular Biology* **29**: 1072-1082.

Kriz W, Kaissling B, Le Hir M. 2011. Epithelial-mesenchymal transition (EMT) in kidney fibrosis: fact or fantasy? *The Journal of Clinical Investigation* **121**: 468-474.

Lacadie SA, Rosbash M. 2005. Cotranscriptional Spliceosome Assembly Dynamics and the Role of U1 snRNA:5′ss Base Pairing in Yeast. *Molecular Cell* **19**: 65-75.

Lakiza O, Frater L, Yoo Y, Villavicencio E, Walterhouse D, Goodwin EB, Iannaccone P. 2005. STAR proteins quaking-6 and GLD-1 regulate translation of the homologues GLI1 and tra-1 through a conserved RNA 3′UTR-based mechanism. *Developmental Biology* **287**: 98-110.

Lal D, Trucks H, Møller RS, Hjalgrim H, Koeleman BPC, de Kovel CGF, Visscher F, Weber YG, Lerche H, Becker F et al. 2013. Rare exonic deletions of the RBFOX1 gene increase risk of idiopathic generalized epilepsy. *Epilepsia* **54**: 265-271.

Lallena MaJ, Chalmers KJ, Llamazares S, Lamond AI, Valcárcel J. 2002. Splicing Regulation at the Second Catalytic Step by Sex-lethal Involves 3′ Splice Site Recognition by SPF45. *Cell* **109**: 285-296.

Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH. 2007. Classical Nuclear Localization Signals: Definition, Function, and Interaction with Importin α. *Journal of Biological Chemistry* **282**: 5101-5105.

Lapuk A, Marr H, Jakkula L, Pedro H, Bhattacharya S, Purdom E, Hu Z, Simpson K, Pachter L, Durinck S et al. 2010. Exon-level microarray analyses identify alternative splicing programs in breast cancer. *Mol Cancer Res* **8**: 961-974.

Larsson SH, Charlieu J-P, Miyagawa K, Engelkamp D, Rassoulzadegan M, Ross A, Cuzin F, Heyningen Vv, Hastie ND. 1995. Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. *Cell* **81**: 391-401.

Lebedeva S, Jens M, Theil K, Schwanhausser B, Selbach M, Landthaler M, Rajewsky N. 2011. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol Cell* **43**: 340-352.

LeBleu VS, Taduri G, O'Connell J, Teng Y, Cooke VG, Woda C, Sugimoto H, Kalluri R. 2013. Origin and function of myofibroblasts in kidney fibrosis. *Nat Med* **19**: 1047-1053.

Lee BJ, Cansizoglu AE, Süel KE, Louis TH, Zhang Z, Chook YM. 2006. Rules for Nuclear Localization Sequence Recognition by Karyopherinβ2. *Cell* **126**: 543-558.

Lee JA, Damianov A, Lin CH, Fontes M, Parikshak NN, Anderson ES, Geschwind DH, Black DL, Martin KC. 2016. Cytoplasmic Rbfox1 Regulates the Expression of Synaptic and Autism-Related Genes. *Neuron* **89**: 113-128.

Lee JA, Tang ZZ, Black DL. 2009. An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. *Genes Dev* **23**: 2284-2293.

Lee KY, Li M, Manchanda M, Batra R, Charizanis K, Mohan A, Warren SA, Chamberlain CM, Finn D, Hong H et al. 2013. Compound loss of muscleblind-like function in myotonic dystrophy. *EMBO Mol Med* **5**: 1887-1900.

Lee Y, Rio DC. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* **84**: 291-323.

Li B, Yen TSB. 2002. Characterization of the Nuclear Export Signal of Polypyrimidine Tract-binding Protein. *Journal of Biological Chemistry* **277**: 10306-10314.

Li L, Zepeda-Orozco D, Black R, Lin F. 2010. Autophagy Is a Component of Epithelial Cell Fate in Obstructive Uropathy. *The American Journal of Pathology* **176**: 1767-1778.

Li Q, Zheng S, Han A, Lin C-H, Stoilov P, Fu X-D, Black DL. 2014. *The splicing regulator PTBP2 controls a program of embryonic splicing required for neuronal maturation*.

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**: 464-469.

Licatalosi DD, Yano M, Fak JJ, Mele A, Grabinski SE, Zhang C, Darnell RB. 2012. Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. *Genes Dev* **26**: 1626-1642.

Lim J, Thiery JP. 2012. Epithelial-mesenchymal transitions: insights from development. *Development* **139**: 3471-3486.

Lim SR, Hertel KJ. 2004. Commitment to Splice Site Pairing Coincides with A Complex Formation. *Molecular Cell* **15**: 477-483.

Lin CH, Patton JG. 1995. Regulation of alternative 3' splice site selection by constitutive splicing factors. *RNA* **1**: 234-245.

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.

Liu P, Wakamiya M, Shea MJ, Albrecht U, Behringer RR, Bradley A. 1999. Requirement for Wnt3 in vertebrate axis formation. *Nat Genet* **22**: 361-365.

Lovci MT, Ghanem D, Marr H, Arnold J, Gee S, Parra M, Liang TY, Stark TJ, Gehman LT, Hoon S et al. 2013. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat Struct Mol Biol* **20**: 1434-1442.

Lu HZ, Liu JL, Liu SJ, Zeng JW, Ding DQ, Carstens RP, Cong YS, Xu XW, Guo W. 2013. Exo70 Isoform Switching upon Epithelial-Mesenchymal Transition Mediates Cancer Cell Invasion. *Dev Cell* **27**: 560-573.

Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. 2011. Epigenetics in Alternative Pre-mRNA Splicing. *Cell* **144**: 16-26.

Lundby A, Secher A, Lage K, Nordsborg NB, Dmytriyev A, Lundby C, Olsen JV. 2012. Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. *Nat Commun* **3**: 876.

Mangus DA, Evans MC, Jacobson A. 2003. Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol* **4**: 223.

Mani SA, Guo W, Liao M-J, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M et al. 2008. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**: 704-715.

Markus MA, Heinrich B, Raitskin O, Adams DJ, Mangs H, Goy C, Ladomery M, Sperling R, Stamm S, Morris BJ. 2006. WT1 interacts with the splicing protein RBM4 and regulates its ability to modulate alternative splicing in vivo. *Experimental Cell Research* **312**: 3379-3388.

Martin CL, Duvall JA, Ilkin Y, Simon JS, Arreaza MG, Wilkes K, Alvarez-Retuerto A, Whichello A, Powell CM, Rao K et al. 2007. Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **144B**: 869-876.

Matsuoka Y, Li X, Bennett V. 1998. Adducin Is an In Vivo Substrate for Protein Kinase C: Phosphorylation in the MARCKS-related Domain Inhibits Activity in Promoting Spectrin–Actin Complexes and Occurs in Many Cells, Including Dendritic Spines of Neurons. *The Journal of Cell Biology* **142**: 485-497.

May GE, Olson S, McManus CJ, Graveley BR. 2011. Competing RNA secondary structures are required for mutually exclusive splicing of the Dscam exon 6 cluster. *RNA* **17**: 222-229.

Mayeda A, Krainer AR. 1992. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**: 365-375.

Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J. 1999. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3[prime] splice site AG. *Nature* **402**: 838-841.

Mertins P, Yang F, Liu T, Mani DR, Petyuk VA, Gillette MA, Clauser KR, Qiao JW, Gritsenko MA, Moore RJ et al. 2014. Ischemia in Tumors Induces Early and Sustained Phosphorylation Changes in Stress Kinase Pathways but Does Not Affect Global Protein Levels. *Molecular & Cellular Proteomics* **13**: 1690-1704.

Moroy T, Heyd F. 2007. The impact of alternative splicing in vivo: mouse models show the way. *RNA* **13**: 1155-1171.

Muh SJ, Hovhannisyan RH, Carstens RP. 2002. A Non-sequence-specific Double-stranded RNA Structural Element Regulates Splicing of Two Mutually Exclusive Exons of Fibroblast Growth Factor Receptor 2 (FGFR2). *Journal of Biological Chemistry* **277**: 50143-50154.

Mukherjee N, Corcoran DL, Nusbaum JD, Reid DW, Georgiev S, Hafner M, Ascano M, Jr., Tuschl T, Ohler U, Keene JD. 2011. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol Cell* **43**: 327-339.

Mukherjee N, Lager PJ, Friedersdorf MB, Thompson MA, Keene JD. 2009. Coordinated posttranscriptional mRNA population dynamics during T-cell activation. *Mol Syst Biol* **5**: 288.

Naydenov NG, Ivanov AI. 2010. Adducins regulate remodeling of apical junctions in human epithelial cells. *Mol Biol Cell* **21**: 3506-3517.

Nieto MA. 2002. The snail superfamily of zinc-finger transcription factors. *Nat Rev Mol Cell Biol* **3**: 155-166.

Nieto MA, Cano A. 2012. The epithelial-mesenchymal transition under control: Global programs to regulate epithelial plasticity. *Semin Cancer Biol* **22**: 361-368.

Nieto MA, Huang Ruby Y-J, Jackson Rebecca A, Thiery Jean P. 2016. EMT: 2016. *Cell* **166**: 21-45.

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

Nistico P, Bissell MJ, Radisky DC. 2012. Epithelial-mesenchymal transition: general principles and pathological relevance with special emphasis on the role of matrix metalloproteinases. *Cold Spring Harb Perspect Biol* **4**.

Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmady B, Pitsch S, Black DL et al. 2005. Structure of PTB Bound to RNA: Specific Binding and Implications for Splicing Regulation. *Science* **309**: 2054-2057.

Onder TT, Gupta PB, Mani SA, Yang J, Lander ES, Weinberg RA. 2008. Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. *Cancer Res* **68**: 3645-3654.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**: 1413-1415.

Pandit S, Zhou Y, Shiue L, Coutinho-Mansfield G, Li H, Qiu J, Huang J, Yeo Gene W, Ares M, Jr., Fu X-D. 2013. Genome-wide Analysis Reveals SR Protein Cooperation and Competition in Regulated Splicing. *Molecular Cell* **50**: 223-235.

Pandya-Jones A, Black DL. 2009. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15**: 1896-1908.

Park SM, Gaur AB, Lengyel E, Peter ME. 2008. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* **22**: 894-907.

Pérez I, McAfee JG, Patton JG. 1997. Multiple RRMs Contribute to RNA Binding Specificity and Affinity for Polypyrimidine Tract Binding Protein. *Biochemistry* **36**: 11881-11890.

Phizicky EM, Greer CL. 1993. Pre-tRNA splicing: variation on a theme or exception to the rule? *Trends in Biochemical Sciences* **18**: IX.

Pollard VW, Michael WM, Nakielny S, Siomi MC, Wang F, Dreyfuss G. 1996. A Novel Receptor-Mediated Nuclear Protein Import Pathway. *Cell* **86**: 985-994.

Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N et al. 2011. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotech* **29**: 436-442.

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A et al. 2013a. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**: 172-177.

Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A et al. 2013b. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**: 172-177.

Rigbolt KTG, Prokhorova TA, Akimov V, Henningsen J, Johansen PT, Kratchmarova I, Kassem M, Mann M, Olsen JV, Blagoev B. 2011. System-Wide Temporal Characterization of the Proteome and Phosphoproteome of Human Embryonic Stem Cell Differentiation. *Science Signaling* **4**: rs3-rs3.

Robberson BL, Cote GJ, Berget SM. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Molecular and Cellular Biology* **10**: 84-94.

Robbins J, Dilwortht SM, Laskey RA, Dingwall C. 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence. *Cell* **64**: 615-623.

Rotzer V, Breit A, Waschke J, Spindler V. 2014. Adducin Is Required for Desmosomal Cohesion in Keratinocytes. *Journal of Biological Chemistry* **289**: 14925-14940.

Ruby SW, Abelson J. 1991. Pre-mRNA splicing in yeast. *Trends in Genetics* **7**: 79-85.

Ryder SP, Williamson JR. 2004. Specificity of the STAR/GSG domain protein Qk1: implications for the regulation of myelination. *RNA* **10**: 1449-1458.

Saccomanno L, Loushin C, Jan E, Punkay E, Artzt K, Goodwin EB. 1999. The STAR protein QKI-6 is a translational repressor. *Proceedings of the National Academy of Sciences* **96**: 12605-12610.

Saldi T, Cortazar MA, Sheridan RM, Bentley DL. 2016. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J Mol Biol* **428**: 2623-2635.

Sauka-Spengler T, Bronner-Fraser M. 2008. A gene regulatory network orchestrates neural crest formation. *Nat Rev Mol Cell Biol* **9**: 557-568.

Sawicka K, Bushell M, Spriggs KA, Willis AE. 2008. Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. *Biochem Soc Trans* **36**: 641-647.

Scheel C, Weinberg RA. 2012. Cancer stem cells and epithelial-mesenchymal transition: concepts and molecular links. *Semin Cancer Biol* **22**: 396-403.

Shapiro IM, Cheng AW, Flytzanis NC, Balsamo M, Condeelis JS, Oktay MH, Burge CB, Gertler FB. 2011. An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet* **7**: e1002218.

Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL. 2008. Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat Struct Mol Biol* **15**: 183-191.

Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A* **111**: E5593-5601.

Shook D, Keller R. 2003. Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development. *Mech Dev* **120**: 1351-1383.

Simone LE, Keene JD. 2013. Mechanisms coordinating ELAV/Hu mRNA regulons. *Current Opinion in Genetics & Development* **23**: 35-43.

Singh A, Settleman J. 2010. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene* **29**: 4741-4751.

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

Siomi H, Dreyfuss G. 1995. A nuclear localization domain in the hnRNP A1 protein. *J Cell Biol* **129**: 551-560.

Skromne I, Stern CD. 2001. Interactions between Wnt and Vg1 signalling pathways initiate primitive streak formation in the chick embryo. *Development* **128**: 2915-2927.

Soniat M, Chook Yuh M. 2015. Nuclear localization signals for four distinct karyopherin-β nuclear import systems. *Biochemical Journal* **468**: 353-362.

Spellman R, Smith CW. 2006. Novel modes of splicing repression by PTB. *Trends Biochem Sci* **31**: 73-76.

Srikantan S, Gorospe M. 2012. HuR function in disease. *Front Biosci (Landmark Ed)* **17**: 189-205.

Stamm S, Zhang MQ, Marr TG, Helfman DM. 1994. A sequence compilation and comparison of exons that are alternatively spliced in neurons. *Nucleic Acids Research* **22**: 1515-1526.

Stevens SW, Ryan DE, Ge HY, Moore RE, Young MK, Lee TD, Abelson J. 2002. Composition and Functional Characterization of the Yeast Spliceosomal Penta-snRNP. *Molecular Cell* **9**: 31-44.

Süel KE, Gu H, Chook YM. 2008. Modular Organization and Combinatorial Energetics of Proline?Tyrosine Nuclear Localization Signals. *PLoS Biol* **6**: e137.

Sugnet CW, Kent WJ, Ares M, Jr., Haussler D. 2004. Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac Symp Biocomput*: 66-77.

Sundberg U, Obrink B. 2002. CEACAM1 isoforms with different cytoplasmic domains show different localization, organization and adhesive properties in polarized epithelial cells. *J Cell Sci* **115**: 1273-1284.

Tacke R, Chen Y, Manley JL. 1997. Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: Creation of an SRp40-specific splicing enhancer. *Proceedings of the National Academy of Sciences* **94**: 1148-1153.

Talerico M, Berget SM. 1994. Intron definition in splicing of small Drosophila introns. *Molecular and Cellular Biology* **14**: 3434-3445.

Tam PPL, Behringer RR. 1997. Mouse gastrulation: the formation of a mammalian body plan. *Mechanisms of Development* **68**: 3-25.

Taube JH, Herschkowitz JI, Komurov K, Zhou AY, Gupta S, Yang J, Hartwell K, Onder TT, Gupta PB, Evans KW et al. 2010. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc Natl Acad Sci U S A* **107**: 15449-15454.

Terenzi F, Ladd AN. 2010. Conserved developmental alternative splicing of muscleblind-like (MBNL) transcripts regulates MBNL localization and activity. *RNA Biol* **7**: 43-55.

Thiery JP, Acloque H, Huang RYJ, Nieto MA. 2009. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell* **139**: 871-890.

Thomson S, Petti F, Sujka-Kwok I, Mercado P, Bean J, Monaghan M, Seymour SL, Argast GM, Epstein DM, Haley JD. 2011. A systems view of epithelial-mesenchymal transition signaling states. *Clin Exp Metastasis* **28**: 137-155.

Tiwari N, Gheldof A, Tatari M, Christofori G. 2012. EMT as the ultimate survival mechanism of cancer cells. *Semin Cancer Biol* **22**: 194-207.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511-515.

Turunen JJ, Niemela EH, Verma B, Frilander MJ. 2013. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA* **4**: 61-76.

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**: 1212-1215.

Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**: 580-586.

Ule J, Ule A, Spencer J, Williams A, Hu J-S, Cline M, Wang H, Clark T, Fraser C, Ruggiu M et al. 2005. Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* **37**: 844-852.

Van Hoof D, Muñoz J, Braam SR, Pinkse MWH, Linding R, Heck AJR, Mummery CL, Krijgsveld J. 2009. Phosphorylation Dynamics during Early Differentiation of Human Embryonic Stem Cells. *Cell Stem Cell* **5**: 214-226.

Vanharanta S, Marney CB, Shu W, Valiente M, Zou Y, Mele A, Darnell RB, Massague J. 2014. Loss of the multifunctional RNA-binding protein RBM47 as a source of selectable metastatic traits in breast cancer. *Elife* **3**.

Venables JP, Brosseau J-P, Gadea G, Klinck R, Prinos P, Beaulieu J-F, Lapointe E, Durand M, Thibault P, Tremblay K et al. 2013. RBFOX2 Is an Important Regulator of Mesenchymal Tissue-Specific Splicing in both Normal and Cancer Tissues. *Molecular and Cellular Biology* **33**: 396-405.

Wagner EJ, Garcia-Blanco MA. 2001. Polypyrimidine Tract Binding Protein Antagonizes Exon Definition. *Molecular and Cellular Biology* **21**: 3281-3288.

Wang Eric T, Cody Neal AL, Jog S, Biancolella M, Wang Thomas T, Treacy Daniel J, Luo S, Schroth Gary P, Housman David E, Reddy S et al. 2012. Transcriptome-wide Regulation of Pre-mRNA Splicing and mRNA Localization by Muscleblind Proteins. *Cell* **150**: 710-724.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470-476.

Wang Y, Vogel G, Yu Z, Richard S. 2013. The QKI-5 and QKI-6 RNA Binding Proteins Regulate the Expression of MicroRNA 7 in Glial Cells. *Molecular and Cellular Biology* **33**: 1233-1243.

Wang YL, Lacroix G, Haines J, Doukhanine E, Almazan G, Richard S. 2010. The QKI-6 RNA Binding Protein Localizes with the MBP mRNAs in Stress Granules of Glial Cells. *Plos One* **5**.

Warzecha CC, Jiang P, Amirikian K, Dittmar KA, Lu H, Shen S, Guo W, Xing Y, Carstens RP. 2010. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J* **29**: 3286-3300.

Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. 2009a. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol Cell* **33**: 591-601.

Warzecha CC, Shen S, Xing Y, Carstens RP. 2009b. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. *RNA Biol* **6**: 546-562.

Weyn-Vanhentenryck Sebastien M, Mele A, Yan Q, Sun S, Farny N, Zhang Z, Xue C, Herre M, Silver Pamela A, Zhang Michael Q et al. 2014. HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism. *Cell Reports* **6**: 1139-1152.

Witten JT, Ule J. 2011. Understanding splicing regulation through RNA splicing maps. *Trends in Genetics* **27**: 89-97.

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

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

Xiao SH, Manley JL. 1997. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes & Development* **11**: 334-344.

Xiao SH, Manley JL. 1998. Phosphorylation–dephosphorylation differentially affects activities of splicing factor ASF/SF2. *The EMBO Journal* **17**: 6359-6367.

Xie J, Lee J-A, Kress TL, Mowry KL, Black DL. 2003. Protein kinase A phosphorylation modulates transport of the polypyrimidine tract-binding protein. *Proceedings of the National Academy of Sciences* **100**: 8776-8781.

Xu D, Farmer A, Chook YM. 2010. Recognition of nuclear targeting signals by Karyopherin-β proteins. *Current Opinion in Structural Biology* **20**: 782-790.

Xu Y, Gao XD, Lee JH, Huang H, Tan H, Ahn J, Reinke LM, Peter ME, Feng Y, Gius D et al. 2014. Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes Dev* **28**: 1191-1203.

Yamagishi R, Tsusaka T, Mitsunaga H, Maehata T, Hoshino S-i. 2016. The STAR protein QKI-7 recruits PAPD4 to regulate post-transcriptional polyadenylation of target mRNAs. *Nucleic Acids Research* doi:10.1093/nar/gkw118.

Yang J, Weinberg RA. 2008. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell* **14**: 818-829.

Yang X, Coulombe-Huntington J, Kang S, Sheynkman Gloria M, Hao T, Richardson A, Sun S, Yang F, Shen Yun A, Murray Ryan R et al. 2016. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* **164**: 805-817.

Ye X, Weinberg RA. 2015. Epithelial–Mesenchymal Plasticity: A Central Regulator of Cancer Progression. *Trends in Cell Biology* **25**: 675-686.

Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, Isakoff SJ, Ciciliano JC, Wells MN, Shah AM et al. 2013. Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition. *Science* **339**: 580-584.

Zhang C, Frias MA, Mele A, Ruggiu M, Eom T, Marney CB, Wang H, Licatalosi DD, Fak JJ, Darnell RB. 2010. Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science (New York, NY)* **329**: 439-443.

Zhang C, Zhang Z, Castle J, Sun S, Johnson J, Krainer AR, Zhang MQ. 2008. Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev* **22**: 2550-2563.

Zhao W, Zhao J, Hou M, Wang Y, Zhang Y, Zhao X, Zhang C, Guo D. 2014. HuR and TIA1/TIAL1 are involved in regulation of alternative splicing of SIRT1 pre-mRNA. *Int J Mol Sci* **15**: 2946-2958.

Zheng X, Carstens JL, Kim J, Scheible M, Kaye J, Sugimoto H, Wu C-C, LeBleu VS, Kalluri R. 2015. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**: 525-530.

Zhou H-L, Lou H. 2008. Repression of Prespliceosome Complex Formation at Two Distinct Steps by Fox-1/Fox-2 Proteins. *Molecular and Cellular Biology* **28**: 5507-5516.

Zhu J, Krainer AR. 2000. Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes & Development* **14**: 3166-3178.

Zong FY, Fu X, Wei WJ, Luo YG, Heiner M, Cao LJ, Fang Z, Fang R, Lu D, Ji H et al. 2014. The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS Genet* **10**: e1004289.

Zorio DAR, Blumenthal T. 1999. Both subunits of U2AF recognize the 3[prime] splice site in Caenorhabditis elegans. *Nature* **402**: 835-838.

Zuo P, Maniatis T. 1996. The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes & Development* **10**: 1356-1368.