Stony Brook University

Regulation and Target Specificity of Human Alternative Splicing Factors SF2/ASF and Fox-1/2

A Dissertation Presented

by

Shuying Sun

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Molecular and Cellular Biology

(Cellular and Developmental Biology)

Stony Brook University

May 2010

Stony Brook University

The Graduate School

Shuying Sun

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Adrian R. Krainer – Dissertation Advisor
Professor, Cold Spring Harbor Laboratory

Senthil Muthuswamy – Chairperson of Defense
Associate Professor, Cold Spring Harbor Laboratory

David L. Spector
Professor, Cold Spring Harbor Laboratory

Linda Van Aelst
Professor, Cold Spring Harbor Laboratory

Rui-Ming Xu
Professor, Institute of Biophysics, Chinese Academy of Sciences

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Regulation and Target Specificity of Human Alternative Splicing Factors SF2/ASF and Fox-1/2

by

Shuying Sun

Doctor of Philosophy

in

Molecular and Cellular Biology

(Cellular and Developmental Biology)

Stony Brook University

2010

Alternative splicing is a highly regulated process in eukaryotes. It greatly increases the diversity of proteins encoded by the genome, and its disruption can cause a number of genetic diseases. SF2/ASF is a prototypical serine/arginine-rich (SR) protein, with important roles in constitutive and alternative splicing and other aspects of mRNA metabolism. *SFRS1* (SF2/ASF) is a potent proto-oncogene with abnormal expression in many tumors. I found that SF2/ASF negatively autoregulates its expression to maintain homeostatic levels of the protein. I characterized six SF2/ASF alternatively spliced mRNA isoforms: the major isoform encodes full-length protein, whereas the others are either retained in the nucleus or degraded by NMD. Unproductive splicing accounts for only part of the autoregulation, which occurs primarily at the translational level. The effect is specific to SF2/ASF and requires RRM2, the second of two RNA-recognition motifs. The ultraconserved 3′UTR (untranslated region) is necessary and sufficient for downregulation. SF2/ASF overexpression shifts the distribution of target mRNA towards mono-ribosomes, and translational repression is partly

independent of Dicer and a 5′ cap. Thus, multiple post-transcriptional and translational mechanisms are involved in fine-tuning the expression of SF2/ASF.

Fox-1 and Fox-2 are brain- and muscle-specific alternative splicing factors. Their single RRM is conserved from worm to human, and specifically binds the RNA element UGCAUG. Fox-1/2 regulate alternative splicing positively or negatively in a position-dependent manner: they activate exon inclusion when binding to the downstream intron, and promote exon skipping when binding to the upstream intron. I explored the mechanisms of splicing activation and repression by Fox-1. I found that Fox-1 can enhance exon inclusion of a heterologous gene when tethered to the downstream intron by a phage MS2 hairpin/coat-protein interaction, and its C-terminal domain is sufficient for this activity. However, both the RRM and the C-terminal domain are required for exon repression when tethered to the upstream intron. I used immunoprecipitation and mass spectrometry to identify proteins that interact with the C-terminal domain of Fox-1. Characterization of several interacting candidates to elucidate their potential roles in alternative splicing regulation by Fox-1 is in progress.

We also applied Solexa high-throughput mRNA sequencing to assess global changes of alternative splicing controlled by Fox-2. We generated ~110 million paired-end reads to compare target-isoform expression levels in cells expressing Fox-2 versus cells treated by RNAi to reduce Fox-2 expression. We identified about 150 high-confidence alternative exons with Fox-dependent splicing, of which 95% could be experimentally validated.

Taken together, my studies provide insights about the regulatory mechanisms involving two kinds of human splicing factors, and have broad implications for post-transcriptional control of gene expression and its misregulation in disease.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my advisor, Adrian Krainer, who gave me tremendous guidance, encouragement and support through my graduate study. I learnt from him how well a good scientist should think and work, and what character and attitude a good scientist should have, a lot of which I still need to cultivate on myself. The training that I have received under his guidance has been invaluable in the path of becoming an independent scientist.

I would like to thank past and present members of the Krainer lab: Zuo Zhang, Rotem Karni, Eric Allemand, Michelle Hastings, Stephanie Shaw, Hazeem Okunola, Lisa Manche, Xavier Roca, Rahul Sinha, Yimin Hua, Oliver Fregoso, Mads Aaboe Jensen, Ying Hsiu Liu, Olga Anczukow-Camarda, Isabel Aznarez, Ruei-Ying Tzeng, Deblina Chatterjee, Shipra Das, Zhenxun Wang, Kentaro Sahashi, and Martin Akerman. Our lab provided a wonderful environment to work. I would like to express my special gratitude to Zuo Zhang, who gave me great help and suggestions in my research. Especially, he set up a model of a young scientist for me, and I was deeply impressed and influenced by his great passion, curiosity, diligence and persistence in scientific research. I also would like to especially thank Xavier Roca and Rotem Karni, who devoted their precious time and effort teaching me experiment skills at the beginning of my graduate study. I would like to thank Oliver Fregoso for his guidance and reagents on immunoprecipitation and mass spectrometry. Finally, I would like to thank Rahul Sinha, who as a senior graduate student, gave me detailed instructions on every step of the graduate life, and was always kind and patient to answer my various questions.

I would like to thank my thesis committee members: Senthil Muthuswamy, David Spector, Linda Van Aelst, Rui-Ming Xu, and Adrian Krainer, for their valuable suggestions on my project's progress, critical reading of my thesis proposal and dissertation, and precious advice for my future career path.

I would like to thank my friends and classmates at both Stony Brook and Cold Spring Harbor. Without them my life in the United States would have been much more difficult, especially during the first year.

Finally, I would like to express my great gratitude to my family. I would like to thank my mom Yan and my dad Guang for their support during my PhD study, even if this meant that I am thousands of miles away from them. I especially wish to thank my mom, who came to help me take care of my new-born daughter, so that I could focus on my research. I would like to thank my husband, Zhaozhu Qiu, who always stands by me, encourages me, shares everything with me no matter whether joy or pain, and has more confidence in me than I do myself. Last, I also want to thank my daughter, Megan, who makes my life more colorful and meaningful, and brings joy and surprise every day.

Publications and Ongoing Work

Sun S, Zhang Z, Sinha R, Karni R, Krainer AR. SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. **Nat Struct Mol Biol**. 2010; 17: 306-312.

Wu H, **Sun S**, Tu K, Gao G, Xie B, Krainer AR, Zhu J. A splicing independent function of SF2/ASF in microRNA processing. **Mol Cell**. 2010; 38: 67-77.

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

Chapter 2 describes the work on SF2/ASF autoregulation published in *Nature Structural Molecular Biology*. Zuo Zhang, Rahul Sinha and Rotem Karni provided valuable reagents and advice. Adrian Krainer and I wrote the manuscript.

Chapter 3 describes ongoing work addressing the mechanisms of activation and repression by the alternative splicing factor Fox-1. Zuo Zhang initially started the project, and I continued it when he moved to Merck. Oliver Fregoso provided a protocol and reagents for immunoprecipitation. Cristian Ruse performed mass spectrometry.

Chapter 4 describes the work using Solexa paired-end mRNA sequencing to globally assess alternative splicing regulation by Fox-2. This work involved extensive collaborations among several people in different labs. Zuo Zhang initially prepared the single-end sequencing cDNA library. I prepared the paired-end sequencing cDNA library and did all the experimental validation. Chaolin Zhang and Chenghai Xue in Michael Zhang's lab performed all the computational analysis. Dr. McCombie's lab ran the Solexa sequencing. Chaolin Zhang wrote the first draft of the manuscript, which will be submitted for publication with Chaolin Zhang, myself, Chenghai Xue, and Zuo Zhang as co-first authors, Richard McCombie as middle author, and Adrian Krainer and Michael Zhang as corresponding authors.

Chapter 1

Background

1.1 The splicing reaction and basic mechanism

The central dogma of molecular biology was first documented by Francis Crick in 1958, in which the process of gene expression was described: DNA passes the information to mRNA by transcription, and RNA encodes proteins by translation [1]. It was not until 1977 that split genes were discovered in adenovirus by two independent investigators, Richard Roberts and Phillip Sharp [2-3]. Since then, numerous studies revealed that mature mRNA is derived from discontinuous segments on DNA not only in viruses, but also in eukaryotes. Pre-mRNA splicing is a general feature and required step for most genes in high eukaryotes to produce correct mRNA for protein production. In this process, the intervening non-coding sequences (introns) are removed and expressed sequences (exons) are joined together to form the mature mRNA. It was reported that the human transcribed genome consists of approximately 230,000 exons and 210,000 introns, which means that there are nine exons and eight introns per gene, on average [4].

There are three basic cis-elements required for splicing: the 5′ splice site, the 3′ splice site and the branchpoint site (BPS). The BPS is usually located 18-40 nucleotides upstream of the 3′ splice site. The sequence stretch between the branch site and the 3′ splice site is enriched in pyrimidines, and therefore is called the polypyrimidine tract (PPT) (**Fig. 1.1**). These elements are recognized by the spliceosome, which then also catalyzes the splicing reactions. However, all these elements are highly degenerate. They are necessary but not sufficient for the precise definition of exon-intron boundaries. There are many other regulatory cis-elements as well as trans-acting factors that facilitate this process. These cis-elements are known as exonic or intronic splicing enhancers (ESE, ISE) and silencers (ESS, ISS). The enhancers are bound by splicing activators, which facilitate the assembly of the spliceosome at the splice sites, while the silencers are bound by splicing repressors, which inhibit the recognition of splice sites. Therefore, these elements play important roles in both constitutive splicing and alternative splicing.

Pre-mRNA splicing involves two transesterification reactions. First, the 2′OH of the adenosine at the BPS performs a nucleophilic attack on the first nucleotide of the intron at the 5′ splice site, forming a lariat intermediate. Second, the 3′OH of the released

5′ exon performs a nucleophilic attack on the last nucleotide of the intron at the 3′ splice site, joining the exons and releasing the intron lariat.

1.2 The spliceosome

The spliceosome is a large complex consisting of five snRNPs (small nuclear ribonucleoproteins) and more than 100 other spliceosomal proteins [5]. There are two kinds of spliceosomes: the major spliceosome (also called U2-type spliceosome) and the minor spliceosome (also called the U12-type spliceosome). The major spliceosome typically recognizes introns with GU-AG at the boundaries. The core components of the major spliceosome are U1, U2, U4/U6 and U5 snRNPs, which contain U1, U2, U4/U6 and U5 snRNAs (U-rich small nuclear RNAs), respectively. The minor spliceosome recognizes introns with AU-AC at the boundaries, which represent a small number of introns in humans; but it also recognizes introns with GU-AG boundaries, in which the splice sites share other characteristic sequence features with the AU-AC introns, especially at the 5′ splice site and BPS. The core components of the minor spliceosome are U11, U12, and U4atac/U6atac snRNPs, with U5 shared between both spliceosomes [6].

Spliceosome assembly preceding splicing catalysis is a highly dynamic process, with most proteins sequentially associating and releasing at one point or another along the path [7]. The first step involves ATP-independent binding of the U1 snRNP to the 5′ splice site of the intron through base-pairing with the 5′ end of the U1 snRNA. This RNA-RNA interaction is not stable and requires the assistance of proteins to stabilize it, such as proteins in the U1 snRNP and serine-arginine-rich (SR) proteins. In the meantime, SF1 binds to the BPS and interacts with the U2 auxiliary factor (U2AF) 65 kD subunit, which binds to the PPT. In addition, U2AF35 forms a heterodimer with U2AF65 and binds to the 3′ splice site AG dinucleotide. All these interactions form the pre-spliceosomal E complex and determine the initial recognition of the 5′ and 3′ splice sites.

In the next step, U2 snRNP displaces SF1 and binds to BPS by base-pairing, leading to the formation of the A complex. Again this interaction is stabilized by other proteins, such as the multi-subunit protein components SF3a and SF3b. This process requires ATP.

Subsequent to A-complex formation, a preassembled U4/U6.U5 tri-snRNP is recruited to form the B complex. U5 snRNP interacts with sequences at the 5′ and 3′ splice sites by weak base pairing between the uridine-rich loop of U5 snRNA and the borders of two exons, and the 3′ end of U6 snRNA base pairs with the 5′ end of U2 snRNA. At this point, all snRNPs are present, but the complex is catalytically inactive. Major conformational and compositional rearrangements including U1 and U4 destabilization and release take place to form an activated B* complex. The spliceosome then catalyzes the first transesterification reaction, giving rise to the C complex. Additional rearrangements occur before the second catalytic step. Finally, the spliceosome dissociates and releases the mRNA in the form of an mRNP complex. The snRNPs are recycled for a further round of splicing.

1.3 Alternative splicing

Alternative splicing is a process by which multiple mature mRNAs are generated from one single gene by using different splice sites. Alternative splicing is a widespread process: recent high-throughput RNA-sequencing analysis of tissue-specific splicing indicated that >90% of human genes express multiple spliced isoforms [8]. Different isoforms can be expressed simultaneously, but often they can be tightly regulated in different tissues, cell types, development stages, in response to cell signaling, etc.

Alternative splicing is a critical mechanism both for regulating gene expression and for generating proteomic diversity. An mRNA with a stop codon >50 nt before the last exon, which is called a premature stop codon (PTC), is typically degraded by nonsense-mediated mRNA decay (NMD) and does not produce proteins [9]. Alternative splicing can downregulate gene expression post-transcriptionally by generating non-

productive isoforms with PTCs. It can also diversify the proteome by generating different protein isoforms with similar, distinct, or even antagonistic functions, depending on their sequences and structures.

1.3.1 Mechanisms of alternative splicing regulation

a. Trans-acting factors

Alternative splicing relies on recognition of the appropriate exon/intron boundaries by the basal splicing machinery, and is also influenced by additional intronic and exonic *cis*-acting elements (ESE, ESS, ISE, and ISS) and their cognate *trans*-acting factors. There are three major groups of splicing factors.

i. SR and SR-related proteins:

The classic SR proteins have similar sequences and structures, with one or two copies of the RNA-recognition motif (RRM) and a C-terminal Arg/Ser-rich (RS) domain. SR-related proteins have RS domains, but may have a different domain organization or completely lack RRMs. Most SR proteins are localized in the nucleus, though some of them shuttle between the nucleus and cytoplasm [10]. They are ubiquitously expressed, although their relative abundances vary in different tissues and developmental stages.

SR proteins are involved in both constitutive splicing and alternative splicing. They typically enhance splicing by recruiting spliceosome components. Most of them can be exchanged for one another in complementation of splicing-deficient HeLa S100 extracts, with at least one being required for constitutive splicing. However, their functions in alternative splicing, which requires enhancer elements, are not fully redundant, as they have different binding sites [10].

Their distinct functions are also represented by the requirements in cell and organism viability. Knockout of SF2/ASF, SRp20, and SC35 in mice are all early embryonic lethal, with different proteins affecting different tissues and developmental stages [11-14]. SF2/ASF is also essential for cell viability in chicken DT40 cells, and the lethality phenotype cannot be rescued by other SR

proteins [15]. B52/SRp55 has been shown to be crucial for Drosophila development [16-17].

ii. hnRNPs:

Heterogeneous nuclear ribonuclear proteins (hnRNPs) are proteins that bind to newly synthesized pre-mRNAs. They influence the structure of RNA and the binding of other RNA processing proteins. There are over 30 hnRNPs (A to U) found in metazoans. Many of them have one or more RNA-binding domains at the N-terminus and an unstructured auxiliary domain at the C-terminus. The most common RNA-binding domains are the RRM, arginine-glycine-glycine repeats called the RGG domain, and the hnRNP K homology (KH) domain [18]. The most frequent auxiliary domain is the glycine-rich domain, which is important for protein-protein interactions. Similar to SR proteins, hnRNPs are also predominantly located in the nucleus, with some members shuttling between nucleus and cytoplasm. And their expression is also ubiquitous among tissues in general.

A number of the hnRNP proteins generally inhibit splicing by binding to silencer elements, but some of them can both activate and repress alternative splicing. The repression mechanisms are also variable for different hnRNPs in different circumstances. hnRNPs can multimerize and block the binding of splicing activators, such as SR proteins, and spliceosome assembly [19-20]. Alternatively, they may bind to both sides of an exon, and the proteins interact with each other to loop out the exon and cause the splice sites to be invisible to the spliceosome [21-22].

There are not many in vivo knockout studies of this protein group so far. One possible reason is many proteins have a few paralogues and they have redundant functions. For example, the mammalian hnRNP A/B family includes three genes: hnRNP A1/A1B, hnRNP A2/B1 and hnRNP A3. They tend to have the same binding sites and have similar effects on splicing repression in *in vitro* assays [23]. A genome-wide analysis of four Drosophila hnRNP A/B family members showed

that individual proteins have different binding motifs *in vitro* and they have overlapping but distinct endogenous pre-mRNA targets *in vivo* [24]. It is not known whether the mammalian proteins also have different endogenous targets. Genomic studies using high-throughput methodologies may provide the answers in the near future.

iii. Other RNA-binding proteins:

Besides the above two large groups of splicing factors, there are several other proteins with distinct structures and activities. They were discovered more recently to be key splicing regulators and many of them have tissue-specific expression patterns.

Fox-1 and Fox-2 are specifically or highly expressed in brain, heart and skeletal muscle. They have one RRM in the middle, and bind to UGCAUG uniquely [25-26]. Nova-1 and Nova-2 are specifically expressed in brain. They have three KH domains, and bind to YCAY clusters [27-28]. Both of these classes of proteins either activate or repress alternative splicing in a position-dependent manner [29-30].

MBNL1 (muscleblind-like 1) and CUGBP1 (CUG-binding protein 1) are two RNA-binding proteins which are believed to partly account for the pathology of myotonic dystrophy (DM) [31]. There are two types of DM caused by different expansions in two loci. DM1 (DM type 1) is caused by (CTG) repeats in the 3′UTR of DMPK (dystrophia myotonica protein kinase), and DM2 (DM type 2) is caused by (CCTG) expansion in the first intron of ZNF9 (zinc finger 9). CUGBP1 belongs to the CELF/BRUNOL family and is expressed ubiquitously. MBNL1 is highly expressed in cardiac, skeletal muscle and during myoblast differentiation. Both of them bind to CUG repeats and positively or negatively regulate alternative splicing of several target genes. In DM1, MBNL1 is sequestered by abnormally expanded (CUG)n tracts, resulting in loss-of-function, while CUGBP1 is up-regulated through a signaling pathway [31]. These two proteins antagonize each other in regulating alternative splicing of several targets, such as

cTAT and IR alternative exons, explaining some of the pathological features of the disease [32].

Other examples are ESRP1 and ESRP2 (Epithelial Splicing Regulatory Proteins 1 and 2), recently discovered epithelial-specific RNAbinding proteins [33-34]; Quaking [35]; Hu proteins [36]; TIA1 and TIAL1 [37]; Sam68 (KHDRBS1) [38-40]; etc.

It is clear that a limited number of proteins regulate a large number of complicated alternative splicing decisions, more than 100,000 in human cells [41]. But on the other hand, each splice site is influenced by multiple factors. Splice-site usage is determined by the combinatorial effects of both positive acting sites and negative acting sites. The balance between the splicing activators and repressors affects the alternative splicing patterns in different tissues and developmental stages. Moreover, nearly all known splicing activators can also function as repressors in some circumstances, and nearly all splicing repressors can occasionally function as activators. Their context-dependent function increases the complexity of alternative splicing regulation.

b. Other mechanisms

The cis-elements and trans-acting factors mentioned above do not compose the whole picture. Another layer of complexity is added when splicing is viewed as a highly dynamic process. It has been shown that splicing is coupled to transcription, and spliceosome components are recruited co-transcriptionally [42]. The rate of transcription can affect alternative splicing [43]. When RNA polymerase II quickly transcribes a portion of a gene, that region of the nascent transcript is quickly folded and the splice sites are less exposed to the spliceosome; therefore, the exons are likely to be skipped during splicing. When the transcription speed is low, the splice sites are more accessible to the spliceosome, and the exons will be more included in the mature mRNA. Of course, weak splice sites are more susceptible to this influence, as strong splice sites will tend to be recognized efficiently under any conditions. In this scenario, any factors that change transcription rates would affect alternative splicing decisions.

An emerging research area is epigenetic regulation of splicing, including chromatin remodeling and histone modification. Short interfering RNA (siRNA) has been shown to induce heterochromatin formation and reduce transcription elongation, thereby changing alternative splicing patterns [44]. Several studies showed a strong correlation between exons with histone H3 trimethylated at lysine 36 (H3K36me3) and nucleosomes [45-47]. It is possible that the exon positions are already pre-marked in the chromatin. But it is not clear how this happens, which is the driving force, and how splice-site recognition is influenced. A recent study showed that a subset of PTB-dependent exons is regulated by H3K36me3 through interactions of H3K36me3, MRG and PTB [48]. However, much remains to be learned about the mechanisms behind the correlating landscapes of chromatin structure and exon position, and the potential effects on alternative splicing.

1.3.2 Signal-induced regulation of alternative splicing

Cell signaling pathways link extracellular signals to gene-expression regulation, and changes in gene expression feed back on cell signaling to modulate cell physiology and function. The regulation of gene transcription by signal transduction pathways has been extensively characterized. Increasing evidence shows that a similar concept also applies to alternative splicing and signaling, but the pathways are much less studied and the mechanistic details are far from complete.

a. Signal-induced modification of splicing factors

Splicing factors usually have multiple modifications, which could be affected by signaling, and thereby influence their activities. SR proteins are subject to extensive phosphorylation on serine residues in their RS domain, and the phosphorylation status regulates splicing activity [49-50]. One concept is that phosphorylation affects protein-protein interactions between SR proteins and components of the spliceosome. It has also been demonstrated that phosphorylation modulates the RNA-binding properties and potentially modifies the target specificity [49]. The subnuclear localization of SR proteins is modulated by phosphorylation too. SR proteins are found to concentrate in nuclear

speckles, also called interchromatin granule clusters (IGCs). During active transcription, the splicing factors are recruited from the IGCs to the active transcription site [51], and phosphorylation is required for this process [52]. Many kinases have been identified to phosphorylate SR proteins both in vitro and in vivo, such as Clk/Sty1, 2, 3 and 4 [53-54], SRPK1 and 2 [55-56], topoisomerase I (in a non-traditional role) [57], as well as the fission yeast kinase DSK1 [58]. However, specific signaling pathways that function through these kinases have not been elucidated yet. The serine/threonine phosphatases PP1 and PP2A have been shown to dephosphorylate SR proteins, and PP1 interacts directly with the first RRM of several SR proteins [59-60].

SRp38 is an unusual member of the SR-protein family. It has the typical domain structure of SR proteins. Phosphorylated SRp38 functions as a sequence-specific splicing activator and this activity requires other unidentified cofactors [61]. It is dephosphorylated during the cell cycle M phase or under heat shock, and becomes a potent repressor of the general splicing machinery [62-63]. The inhibitory effect was proposed to be brought about by the interaction of dephosphorylated SRp38 with U1 70K, which may interfere with the association with other SR proteins that are necessary for U1 snRNP function [62]. Two PP1 isoforms, PP1β and PP1γ, were shown to be responsible for SRp38 dephosphorylation during heat shock [64].

hnRNP A1 has been shown to be phosphorylated upon different stress stimuli, including osmotic shock or ultraviolet-C (UVC) irradiation, which changes its localization from the nucleus to cytoplasmic stress granules (SGs). This phosphorylation is catalyzed by Mnk1/2 kinases activated through the mitogen-activated protein (MAP) kinase 3/6-p38 (MKK3/6) stress-signaling pathway [65-66]. Accumulation of hnRNP A1 in the cytoplasm causes an altered ratio of SR protein and hnRNP A/B protein antagonistic alternative splicing factors in the nucleus and consequently affects alternative splicing.

The protein phosphorylation status also plays important roles in core spliceosome assembly. PP1 and some PP2A phosphatases have essential but redundant effects on

splicing. They dephosphorylate components of U2 and U5 snRNPs and are crucial for the second step of the splicing reaction [67].

Although the examples mentioned above involve changes in protein phosphorylation, other modifications might play significant roles in splicing too, such as methylation. Several components of the core spliceosomehave been shown to have Arg dimethylation, which is essential for snRNP assembly. For example, Sm proteins D1 and D3 are asymmetrically dimethylated, Lsm4 has symmetrical dimethylation, and Sm B/B′ has both symmetrical and asymmetrical dimethylation [68-69]. Many hnRNPs contain extensive methylation of RGG motifs in the C-terminal Gly-rich domain [70], and methylation has been shown to promote hnRNP A2 nuclear localization in mammals [71]. Similarly, three methylated Arg residues in the inter-RRM linker region of SF2/ASF also control its subcellular localization [72]. However, it is not known how signaling pathways modulate these modifications.

b. Signal-induced splicing of specific transcripts

Besides the modification changes of splicing factors, other pieces of information come from the signal-induced alternative splicing of particular minigenes. Expression of the cell-surface molecule CD44 is a good example. The most common isoform is the smallest CD44 in most tissues, whereas proliferating cells and many tumors express various other isoforms, among which variant exon 5 (v5) is extensively studied. Inclusion of exon v5 is induced by the Ras-ERK pathway [73]. SAM68 was proposed to be the link between signaling activation and exon inclusion. SAM68 is phosphorylated by ERK in response to Ras signaling, and enhances exon v5 inclusion [74].

Another example is CD45 pre-mRNA, which encodes a transmembrane protein phosphatase in T cells. Exon 4 skipping is induced upon extracellular signal activation. Activation of the Ras or protein kinase C pathway is important for the splicing change [75]. Several factors have been shown to contribute to the silencing of this exon, including hnRNP L, PSF, and hnRNP LL [76-77]. PSF and hnRNP LL bind to an ESS element preferentially upon activation, and the expression level of hnRNP LL also

increases. However, it is not known why the RNA affinity changes and how signaling activation induces these changes. It is possible that additional factors are involved in the regulation and remain to be discovered.

Alternative splicing is most prevalent in the nervous system. Many external stimuli, such as chronic depolarization or other treatments that stimulate excitatory activity, dynamically regulate alternative splicing, which feeds back to modulate neuronal activity. The STREX (stress axis) exon of the BK channel gene is repressed after depolarization, through the calcium/calmodulin-dependent protein-kinase (CaMK)-mediated pathway. A CaMKIV-responsive element (CaRRE) was identified near the 3′ splice site of the exon [78]. This element is also found near exon 5 and exon 21 of NMDAR1 [79-80]. The two exons individually encode N1 and C1 peptide cassettes, which affect the binding and trafficking of the receptor. The CaRRE can repress a heterologous exon upon expression of activated CaMKIV, but the factors that bind to it and are regulated by cell excitation and calcium signaling are mostly unknown [81].

The regulation of splicing by signal transduction pathways is an important aspect of gene expression control. However, we still have a long way to go before fully understanding the complexity. Details of the mechanisms linking cell signaling to splicing control need to be explored. The physiological effect of each splicing change should also be evaluated.

1.3.3 Global insights on alternative splicing

Traditional research on alternative splicing regulation usually focuses on specific minigenes, either *in vitro* or *in vivo*. The number of known targets for each splicing factor remained very limited for a long time. In recent years, high-throughput techniques have allowed the analysis of thousands of alternative splicing events, which begins to give a global view of the whole transcriptome, and also provides important insights about the regulation mechanisms. These tools allow large-scale quantitative characterization of splice variants under different conditions.

Initially, microarrays were used in various research areas. Splicing-sensitive microarrays can be divided into four categories, depending on the probe locations: (1) tiling arrays, with overlapping probes across a known genomic sequence; (2) exon-body arrays, with probes located within exons; (3) splice-junction arrays, with probes crossing spliced junctions; (4) exon-junction arrays, with probes both within exons and across exon junctions [82]. Although microarrays have been applied successfully over the years, they have some technical limitations, such as cross-hybridization problems and background noise. In particular, probe design is limited to known isoforms and organisms with sequenced genomes; therefore, microarrays are not an ideal method for discovery of new alternative splicing events.

High-throughput sequencing technologies developed rapidly in recent years. RNA-seq can generate millions of short sequence reads of cDNAs derived from polyA-enriched mRNA [83]. Reads are then mapped to the unique locations in the genome and transcriptome (for splice-junction reads). Differences in read densities across genes or exons for RNA obtained under different conditions represent the quantitative variation of gene expression and alternative isoforms. RNA-seq directly provides sequence information, which makes it a powerful tool for identification of new alternative splicing isoforms. Furthermore, read counts give a more accurate estimation of relative expression levels and do not suffer from the noise that is caused by different hybridization affinities [8].

More and more groups are applying these techniques to address different aspects of alternative splicing, including tissue-specific changes, developmentally regulated splicing, signal-activated alterations, disease-associated splicing, and targets of particular splicing factors by depletion or overexpression, etc. These techniques can also be combined with biochemical tools, such as CLIP, to identify the endogenous binding targets of splicing factors. In this assay, RNA is cross-linked to bound proteins bound by UV irradiation *in vivo*, then fragmented and isolated by immunoprecipitation of the bound protein, reverse transcribed and sequenced [84]. This method gives a global view of the landscape of individual proteins. When combined with the functional changes on

alternative splicing, this can reveal the so-called 'RNA code', such as the position-dependent activities of Nova and Fox [29-30, 85]. The endogenous functions of proteins with similar structures and activities *in vitro* can also be analyzed and compared. It is also possible to discover and dissect novel functions besides splicing, as many RNA-binding proteins have multiple functions in RNA metabolism. Finally, all this information can provide valuable insights for further mechanistic dissection at the molecular level.

1.4 The impact of splicing on other aspects of RNA metabolism

Splicing is not an isolated process. Instead, it is highly coordinated with other RNA processing events, both in the nucleus and in the cytoplasm. Splicing is coupled to transcription, as well as to 5′- and 3′-end processing. The splicing history of mRNAs also influences the downstream steps in RNA metabolism. This is largely mediated by mRNP proteins which dynamically interact with a large number of factors, thus exerting their widespread regulatory effects on various aspects of RNA metabolism.

1.4.1 Splicing and transcription

Splicing is usually coupled with transcription. Besides the effect on splicing by transcription rate, the splicing machinery can also influence the polymerase speed. The presence of a promoter-proximal splice site increases transcription, partly at the level of enhanced polymerase II initiation [86]. Pol II elongation is promoted by the elongation factor TAT-SF1, which interacts with U2 snRNP [87]. A member of the SR protein family, SC35, has been shown to promote transcription elongation by interacting with CDK9, which phosphorylates serine-2 of the Pol II C-terminal domain [88].

1.4.2 Splicing and 3′-end processing

3′-end processing can also be influenced by splicing, through a coupled network of proteins involved in both processes [89]. SRm160 (Serine/arginine repeat-related nuclear matrix protein of 160 kDa), a splicing coactivator, associates with the cleavage polyadenylation specificity factor and promotes 3′-end cleavage of splicing-active pre-

mRNAs more efficiently than splicing-inactive ones [90]. PTB has been shown to either repress or activate 3′-end formation in different transcripts [91-92]. Nova, a brain splicing factor, also binds close to the poly (A) sites and regulates alternative polyadenylation through unknown mechanisms [93].

1.4.3 Splicing and mRNA export

Splicing can enhance the nuclear export efficiency by different mechanisms. The THO/TREX complex associates with mRNA in a 5′-cap- and splicing-dependent manner [94]. It binds preferentially at a single location just downstream of the cap, as mediated by an interaction between one of its components, REF/Aly, and the nuclear cap-binding protein CBP80 [95]. THO/TREX interacts with nuclear export factors and helps anchor mRNPs to the NPC (Nuclear Pore Complex) [96]. The recruitment of the complex increases the efficiency of mRNA export, and is believed to allow the 5′ end to emerge first into the cytoplasm [97].

Unlike the THO/TREX complex, the EJC (Exon Junction Complex) is deposited approximately 20 nucleotides upstream of every exon-exon junction, and marks the original positions of the exon-intron boundaries in the mature mRNA. EJCs are exported into the cytoplasm with the mRNA and are removed by the translation machinery during the first round of translation [98]. Experiments in *Xenopus* oocytes indicated that the presence of EJCs can enhance the efficiency of mRNA export [99-100]. Several components of the EJC complex have been reported to facilitate export, including UAP56, REF/Aly, and Pinin [98].

SR and SR-like proteins can also function as mRNA export adaptors. The shuttling SR proteins SRp20, 9G8, and SF2/ASF interact with the general export receptor NXF1/TAP in their hypophosphorylated state. It was reported that SR proteins initially regulate splicing in their hyperphosphorylated state, and become partially dephosphorylated by the end of splicing. This presumably marks the mRNA as ready for export. After export, the RS domains are rephosphorylated and the SR proteins release from the mRNA and are reimported into the nucleus [96, 101].

1.4.4 Splicing and translation

It is well established that spliced mRNAs have higher translational efficiency than identical cDNA-expressed transcripts. Spliced mRNAs are more efficiently incorporated into the translation machinery and therefore more mRNA molecules are associated with polyribosomes, leading to a higher translation rate [102].

The EJC is a key factor for enhancement of translation initiation [103]. One mediator between EJC-bound mRNAs and ribosomes is PYM. It interacts with the EJC proteins Y14-Magoh, and also binds, via a separate domain, to the small (40S) ribosomal subunit and the 48S preinitiation complex [104]. Another mechanism is through EJC-bound SKAR, which recruits activated S6K1 (S6 kinase 1) to the newly synthesized mRNA and promotes the pioneer round of translation [105]. S6K1 is a key player in the mTOR signaling pathway and a critical regulator of protein translation related to cell growth [106].

SF2/ASF has also been shown to promote translation in an enhancer-dependent manner [107]. It promotes phosphorylation of 4E-BP1 and thereby releases its inhibition on translation initiation. It was also proposed to repress the activity of PP2A, which dephosphorylates S6K1 [108]. Phosphorylated S6K1 is active in promoting translation.

1.4.5 Splicing and nonsense-mediated mRNA decay

Nonsense-mediated mRNA decay (NMD) is a translation-dependent decay pathway that specifically targets mRNAs containing premature termination codons (PTCs). In mammals, the natural stop codons are generally located in the last exon, or within 50 nucleotides of the last exon-exon junction [109]. Otherwise, the EJC complex is loaded downstream of the stop codon, and this will trigger mRNA degradation after the pioneer round of translation [110]. The EJC plays an important role by recruiting the NMD-specific factors Upf2 and Upf3. The central factor in NMD is the RNA helicase Upf1, which associates with its kinase Smg1 and two release factors, eRF1 and eRF3, forming the SURF complex. Contact between SURF and Upf2 bound to a downstream EJC leads to phosphorylation of Upf1 by Smg1 [111]. Then, phosphorylated Upf1

recruits the RNA decay machinery by an unknown mechanism and triggers mRNA degradation.

The SR protein SF2/ASF has been shown to promote NMD of PTC-containing transcripts [112]. However, the detailed mechanism is not known. It may be related to its function in translation enhancement, but more experiments are needed to evaluate this hypothesis.

1.5 Concluding remarks

RNA splicing is a critical post-transcriptional step in gene expression, and it is subject to regulation. Alternative splicing greatly increases the proteome diversity and organism complexity in higher eukaryotes. It provides cells with another layer of capacity to fine-tune their functions. Although great progress has been made in understanding splicing mechanisms, numerous issues remain unresolved. For example, a lot of splicing isoforms, even for extensively characterized genes, are still unknown, not to mention the functional relevance of these isoforms. Furthermore, the splice-codes of most splicing factors have not been fully characterized. How do they coordinate with each other? Do their functions change under different conditions, such as with certain external stimuli or in the presence or absence of other splicing factors? Recent technological advances may provide new opportunities to understand the complexity of splicing regulatory networks and the diversity of molecular mechanisms.

1.6 Figures and Figure Legends

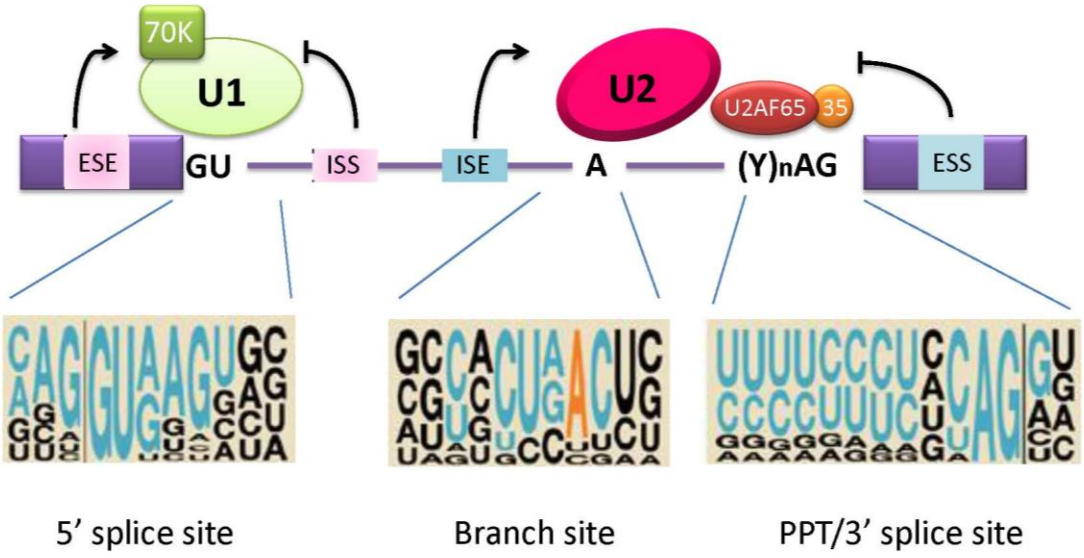

Figure 1.1 Critical cis-elements for splicing regulation

The consensus motifs of the 5′ splice site, branch site, polypyrimidine tract and 3′ splice site are described below the diagram. The height of the letters is proportional to the frequencies of the nucleotides at the corresponding positions. The positions of ESE, ESS, ISE, and ISS are representative. U1 snRNP binds to the 5′ splice site. U2 snRNP recognizes the branch point sequence. U2AF65 and 35 interact with the polypyrimidine tract and 3′ splice site as a heterodimer. ESE: exonic splicing enhancer; ESS: exonic splicing silencer; ISE: intronic splicing enhancer; ISS: intronic splicing silencer.

Chapter 2

SF2/ASF Autoregulation Involves Multiple Layers of Post-transcriptional and Translational Control

2.1 Introduction

Alternative splicing is widespread: recent high-throughput RNA-sequencing analysis of tissue-specific splicing indicated that >90% of human genes express multiple spliced isoforms [1]. SF2/ASF is a prototypical SR protein that participates in both constitutive and alternative splicing [2]. Additional functions of SF2/ASF extend to other aspects of mRNA metabolism, such as NMD (nonsense-mediated mRNA decay) [3], mRNA export [4-5], and translation [6].

Although SF2/ASF levels vary widely among cell types [7], tight control of its expression is important for normal cell and organismal physiology. Knockdown of SF2/ASF results in genomic instability, cell-cycle arrest, and apoptosis [8-9]. Knockout of SF2/ASF in cardiomyocytes results in defective postnatal heart remodeling in mice, due to incorrect *CAMK2D* splicing [10]. Moderate (2-3 fold) overexpression of SF2/ASF is sufficient to transform immortal rodent fibroblasts, which then rapidly form sarcomas in nude mice [11]. SF2/ASF also regulates alternative splicing of the *MST1R* (RON) proto-oncogene, inducing cell motility and invasion [12]. SF2/ASF shows abnormal expression in many tumors [11], but little is known about how its expression is regulated, or why it is up-regulated in cancer, though gene amplification was found in some breast tumors [11].

Besides transcription, gene expression can be regulated at both post-transcriptional and translational levels. Alternative splicing can regulate gene expression by generating non-productive isoforms, such as mRNAs that are retained in the nucleus or are subject to NMD, or by encoding proteins with different functions [2, 13]. mRNA turnover and translation are also key control points for gene-expression regulation, frequently mediated by 3′UTR elements. For example, AU-rich elements (AREs) and associated proteins affect mRNA stability and translational efficiency [14]. In addition, microRNAs (miRNAs) and small interfering RNAs (siRNAs) are important regulators of translation and mRNA decay [15].

Many splicing factors are regulated post-transcriptionally. In *C. elegans*, two SR proteins, SRp20 and SRp30b, have premature termination codon (PTC)-containing splicing isoforms, whose degradation depends on the smg genes [16]. Likewise, the mammalian SR protein SC35, and the polypyrimidine-tract-binding protein (PTB) autoregulate by promoting the expression of unstable alternatively spliced mRNA isoforms that undergo NMD [17-18]. SRp20, another SR protein, promotes expression from its own gene of a splicing isoform encoding a truncated protein, and SF2/ASF antagonizes this regulation [19]. Recent reports described ultraconserved (UCR) elements in every member of the SR protein family, as well as in PTB [20-23]. UCRs are present in regions that undergo alternatively splicing events that introduce PTCs, such that some of the resulting mRNAs are NMD targets. Thus, unproductive splicing can regulate SR protein expression [20-21].

Here we report that SF2/ASF negatively regulates its own expression, and we investigate the underlying mechanisms. We demonstrate that multiple layers of control, including alternative splicing and translational regulation, are involved in this homeostatic process.

2.2 Results

2.2.1 SF2/ASF autoregulation by negative feedback

We placed an SF2/ASF cDNA under the control of the TRE-CMV promoter and transduced HeLa cells stably expressing the tetracycline trans-activator protein tTA (tet-off) [24]. In medium without tetracycline, SF2/ASF expression was turned on (**Fig. 2.1a**). An N-terminal T7 tag allowed separation of ectopic and endogenous SF2/ASF by SDS-PAGE. Western blotting revealed that expression of endogenous SF2/ASF was reduced by ~70% in T7-SF2/ASF overexpressing cells, compared to uninduced cells, whereas a β-catenin loading control was unaffected (**Fig. 2.1b**). These data confirm that SF2/ASF autoregulates its expression, as reported for stable retroviral transduction of human, mouse, and rat cells [11].

2.2.2 Alternative splicing contributes to autoregulation

We first examined whether SF2/ASF autoregulation occurs via alternative splicing, as previously proposed [20-21]. To identify all the isoforms expressed in HeLa cells, we amplified them by RT-PCR from total RNA using primers positioned at the ends of the first and last exon of the canonical isoform [25-26] (**Fig. 2.2a**). We detected six isoforms, with the canonical one being by far the most abundant. Other human cell lines, such as HEK293 and IMR90, showed similar patterns of SF2/ASF mRNA isoforms (not shown).

Cloning and sequencing revealed that isoforms III-VI undergo excision of one or two introns in their 3′UTR, resulting in PTCs that should trigger NMD [27]. However, when we inhibited NMD with cycloheximide, only isoforms V and VI increased substantially (**Fig. 2.2b**).

To determine the subcellular localization of the various mRNA isoforms, we performed cell fractionation, and we extracted RNAs from nuclear and cytoplasmic fractions for RT-PCR analysis (**Fig. 2.2c**). Surprisingly, isoforms II, III, and IV were all retained in the nucleus, which could explain why isoforms III and IV escape NMD, as this pathway requires a round of cytoplasmic translation [27].

To determine how SF2/ASF overexpression affects each isoform, we performed RT-PCR of endogenous SF2/ASF mRNAs using the same samples as in Figure 2.1. The reverse primer corresponds to the end of the 3′UTR, which is absent in the ectopic SF2/ASF cDNA. After induction of T7-SF2/ASF, isoforms III and VI increased markedly (**Fig. 2.2d**). The protein-coding isoform I decreased by ~30% (**Fig. 2.2e**), considerably less than the ~70% reduction at the protein level (**Fig. 2.1b**).

These data show that SF2/ASF modulates alternative splicing of its own transcript, and downregulates itself in part by decreasing the production of the protein-coding isoform and increasing the isoforms that are retained in the nucleus or degraded by NMD. However, this switch in alternative splicing does not fully account for the downregulation

at the protein level, suggesting that additional mechanisms are involved in SF2/ASF autoregulation.

2.2.3 Autoregulation is specific to SF2/ASF and requires RRM2

To better understand the mechanisms underlying SF2/ASF homeostasis, we amplified the genomic segment of the transcribed region of *SFRS1* from human DNA by PCR, and subcloned it into pcDNA3.1+. To detect the proteins expressed from the transfected genomic construct, we added a V5 tag before the start codon, and omitted the natural 5′UTR (**Fig. 2.3a**). Except where indicated, we used V5-SF2/ASF as a reporter and co-expression of T7-SF2/ASF cDNA (including the coding exons but not the UTRs) to study SF2/ASF autoregulation. By co-expressing V5-tagged genomic SF2/ASF and T7-tagged SF2/ASF cDNA, we sought to recapitulate the autoregulation. We transiently co-transfected HeLa cells with a constant amount of genomic V5-SF2/ASF plasmid and increasing amounts of T7-SF2/ASF cDNA plasmid (**Fig. 2.3b**). Western blotting using V5 and T7 antibodies showed that overexpression of SF2/ASF cDNA strongly repressed the protein expressed from genomic SF2/ASF in a dose-dependent manner.

By co-transfecting HeLa cells with equal amounts of V5 genomic SF2/ASF plasmid and T7-tagged cDNAs of SF2/ASF mutants or other SR proteins, we confirmed that downregulation is not via promoter competition, and established that it is SF2/ASF-specific and requires RRM2 (**Fig. 2.3c**). Co-expression of SC35 or two other SR proteins, SRp55 and SRp75, did not affect the expression level of SF2/ASF from the genomic construct (**Fig. 2.3c**, lane 13, and data not shown). SRp30c—the closest paralog of SF2/ASF—had lower expression than most of the other proteins, even when we transfected three times more plasmid; even after normalizing to the expression level, its effect was slight (see histogram below the gel). Considering that SF2/ASF-ΔRS was also weakly expressed, yet it strongly decreased V5-SF2/ASF expression, the effect of SRp30c, if any, is much less pronounced than that of SF2/ASF. Most of the SF2/ASF mutants, including RS-domain deletion (ΔRS), RRM1 deletion (ΔRRM1), and nuclear-retained SF2/ASF with the NRS signal from SC35 (NRS-SC35), retained the

downregulation activity of SF2/ASF. Only the RRM2-deletion mutant (ΔRRM2) was defective in downregulation.

Using a forward primer corresponding to the V5 tag and a reverse primer in SF2/ASF exon 3 for radioactive RT-PCR, we specifically amplified the total mRNA expressed from the transfected genomic construct. Interestingly, the change in mRNA level was not always consistent with the downregulation of SF2/ASF protein expression. Overexpression of T7-SF2/ASF led to accumulation of unspliced pre-mRNA and a decrease in mature mRNA—a decrease in splicing efficiency previously observed with other splicing reporters [3]. As the T7-SF2/ASF protein increased, the V5-SF2/ASF protein level decreased steeply, whereas the spliced mRNA decreased much more gradually (**Fig. 2.3b and 2.3d**, lanes 1-6). ΔRS did not cause a decrease in mRNA level, but still caused strong downregulation of SF2/ASF protein expression, whereas ΔRRM2 resulted in decreased mRNA, but no change at the protein level (**Fig. 2.3c and 2.3d**, lanes 9 and 11). Furthermore, two other SR proteins, SC35 and SRp30c, also markedly inhibited splicing and decreased the mature mRNA level, but did not markedly repress protein expression (**Fig. 2.3c and 2.3d**, lanes 13 and 14). Therefore, the changes in steady-state mRNA levels do not consistently account for the observed decrease at the protein level.

2.2.4 The 3′UTR is necessary and sufficient for autoregulation

To identify regions important for SF2/ASF autoregulation, we constructed a genomic version of SF2/ASF with all three coding-region introns precisely deleted (**Fig. 2.4a**). We co-transfected HeLa cells with wild-type or Δintron123 V5-SF2/ASF and T7-SF2/ASF cDNA. Western blotting showed that SF2/ASF still downregulated protein expression from this intronless construct (**Fig. 2.4b**, lanes 1-2). Therefore, splicing out the first three introns is not required for autoregulation. To eliminate further splicing within the 3′UTR, without changing its length, we also inactivated the two pairs of alternative splice sites in this region by mutating G to C at the +1 position of the 5′ splice sites, and mutating G to T at the -1 position of the 3′ splice sites. SF2/ASF still showed autoregulation with this construct (**Fig. 2.4b**, lanes 3-4). However, when we replaced the

3′UTR with bacterial sequences, but kept the length constant, SF2/ASF no longer downregulated the protein expression (**Fig. 2.4b**, lanes 5-6).

Another construct, ΔUTR, replaces the entire 1.9-Kb 3′UTR of SF2/ASF with ~100 bp of vector sequence (**Fig. 2.4a**). This construct also gave very different results than the wild-type construct. First, the basal level of protein greatly increased (**Fig. 2.4c**). To obtain comparable expression, we transfected cells with only 1/5 as much DNA for this construct, and loaded half as much protein for Western analysis **(Fig. 2.4b**, lanes 7-8). Second, the protein expressed from this construct was not downregulated by overexpression of T7-SF2/ASF cDNA (**Fig. 2.4b**, lanes 7-8; **2.4c**, lanes 7-9). When we deleted both the 3′UTR and the first three introns, we obtained similar results as with ΔUTR (**Fig. 2.4c**, lanes 10-12). The lower expression level of SF2/ASF with its natural 3′UTR may reflect further inhibition by endogenous SF2/ASF, and may also be a non-specific effect of 3′UTR length, as a bacterial-sequence 3′UTR of the same length gave comparable basal-level expression as the natural 3′UTR (**Fig. 2.4b**, lanes 5-6). In general, very long 3′UTRs tend to repress translation [28]. Finally, in all cases, despite very large differences at the protein level, there was relatively little variation at the mRNA level (**Fig. 2.4b**).

To examine whether the 3′UTR of SF2/ASF is sufficient to repress expression in response to SF2/ASF overexpression, we subcloned the 3′UTR after the coding sequence of a Renilla luciferase reporter. We co-transfected reporter constructs with or without the SF2/ASF 3′UTR with T7-SF2/ASF cDNA or control vector into HeLa cells. We measured luciferase activity and performed radioactive RT-PCR of luciferase mRNA as a normalization control (**Fig. 2.4d**). The basal expression of luciferase with SF2/ASF's 3′UTR was approximately 60% of the control. Overexpression of SF2/ASF downregulated the luciferase reporter with the SF2/ASF 3′UTR to ~20%, but had no repressive effect with the control gene. Therefore, the SF2/ASF 3′UTR is both necessary and sufficient to mediate downregulation of gene expression by SF2/ASF overexpression.

2.2.5 The 3′UTR of SF2/ASF does not inhibit mRNA export

To address whether SF2/ASF inhibits nuclear export of its own mRNA, we performed subcellular fractionation after co-transfecting HeLa cells with either wild-type or Δintron123 V5-SF2/ASF and T7-SF2/ASF cDNA or control vector. Radioactive RT-PCR of GAPDH pre-mRNA, which is retained in the nucleus, confirmed the clean separation of nucleus and cytoplasm (**Fig. 2.5**). The proportion of V5-SF2/ASF mRNA present in the cytoplasm was very similar with and without T7-SF2/ASF co-expression. Thus, SF2/ASF mRNA export is not inhibited by SF2/ASF overexpression, and is not the mechanism of SF2/ASF autoregulation.

2.2.6 SF2/ASF downregulates itself at the level of translation

Translation is a highly regulated process, and initiation is usually the rate-limiting step [29]. To determine whether SF2/ASF autoregulation involves decreased translational efficiency, we performed *in vitro* translation in HeLa cell extract [30]. We *in vitro* transcribed luciferase-reporter mRNAs with or without the SF2/ASF 3′UTR, and in some cases added a poly(A) tail (**Fig. 2.6a**). We incubated equal amounts of mRNAs in the extract, and measured luciferase activity (**Fig. 2.6b**). Translation of the 3′UTR-containing mRNA was less efficient, consistently with the above transfection result (**Fig. 2.4d**). However, there was little if any change when we added purified recombinant SF2/ASF—expressed in bacteria or in mammalian cells (**Fig. 2.6b**). Therefore, we could not recapitulate the autoregulation of SF2/ASF *in vitro*. However, the same translation extract did respond to SF2/ASF addition when we assayed for ESE-dependent stimulation (not shown), as previously reported[6], suggesting that different mechanisms underlie positive and negative control of translation by SF2/ASF.

This negative result *in vitro* does not rule out translation inhibition as the mechanism of SF2/ASF autoregulation. Therefore, we next used sucrose gradients to directly analyze the distribution of reporter mRNAs on polyribosomes *in vivo*. We co-transfected HeLa cells with V5-SF2/ASF Δintron123 and either T7-SF2/ASF cDNA or control vector. We also co-transfected as an internal control a Renilla-luciferase reporter

with a bacterial-sequence 3′UTR of the same length. After 48 h, we fractionated cytoplasmic extracts on 10%-50% sucrose gradients, and detected V5-SF2/ASF mRNA in each fraction by radioactive RT-PCR (**Fig. 2.7**). In contrast to the control endogenous GAPDH mRNA, which peaked in the heavy polyribosome fractions, the main peak of V5-SF2/ASF mRNA or Rluc-pucUTR control-reporter mRNA was in the monoribosome fractions (**Fig. 2.7a,b**, left panels). This distribution is consistent with the repressive effect of the long 3′UTRs. An additional, broad peak of V5-SF2/ASF mRNA sedimented with polyribosomes. Co-expression of T7-SF2/ASF shifted this broad peak towards the monoribosome fractions, indicating that SF2/ASF reduced the translation efficiency of V5-SF2/ASF with the natural 3′UTR (**Fig. 2.7a,b**, right panels). The difference between the two distribution profiles is significant (p=0.028, Kolmogorov-Smirnov test). In contrast, the distribution of the Rluc-pucUTR control mRNA was not changed by SF2/ASF overexpression, consistent with our finding that SF2/ASF did not reduce the luciferase activity in the presence of the bacterial-sequence 3′UTR (see below, **Fig. 2.9b**). Treatment of cells with puromycin confirmed that sedimentation of the mRNAs in denser fractions indeed reflected polysome association (**Fig. 2.7c**).

2.2.7 Potential contribution of miRNAs to autoregulation

miRNAs regulate gene expression by controlling the translation or stability of target mRNAs. TargetScan predicts multiple putative miRNA targets in the 3′UTR of SF2/ASF (not shown). Dicer is an enzyme required for miRNA maturation [31]. To examine the role of miRNAs in SF2/ASF autoregulation, we used Dicer-disrupted or -knockout cell lines. We first used DicerEx5/Ex5 RKO cells, in which exon 5 of human *Dicer* is disrupted, interrupting the helicase domain [32]. We co-transfected V5-SF2/ASF with T7-SF2/ASF cDNA or control vector into wild-type or DicerEx5/Ex5 RKO cells. Western blotting showed that SF2/ASF downregulated itself in both wild-type and Dicer-disrupted cells (**Fig. 2.8a**). However, because the biogenesis of some miRNAs is not disrupted in these cells [32], the potential involvement of some miRNA(s) in SF2/ASF autoregulation could not be ruled out. We therefore used Dicer-null mouse ES cells, which have compromised proliferation but are viable [33]. The *Dicer* gene is completely

knocked out in these cells, and the biogenesis of all miRNAs is thought to be fully disrupted. We performed similar co-transfection experiments as above with Dicer$^{-/-}$ and control Dicer$^{+/-}$ ES cells, and observed downregulation in both cases, although there was less repression in Dicer-null cells, perhaps due to their reduced proliferation (**Fig. 2.8b**). This experiment suggests that miRNA-mediated gene repression may contribute to SF2/ASF autoregulation, though not as the main mechanism.

2.2.8 Effect of cap-dependent versus IRES-dependent translation

We next examined whether 3′UTR-mediated translational repression of SF2/ASF can occur in the context of internal ribosome entry site (IRES)-dependent translation initiation. Translation driven by different viral IRES elements requires distinct subsets of the initiation factors necessary for cap-dependent translation [34]. The encephalomyocarditis virus (EMCV) IRES requires most initiation factors, except for the cap-binding protein eIF4E. The hepatitis-C virus (HCV) IRES only requires eIF3 and eIF2. Finally, the cricket-paralysis virus (CrPV) IRES bypasses the requirement for all the initiation factors. We placed these IRES sequences 5′ of a Renilla luciferase reporter with or without the SF2/ASF 3′UTR (**Fig. 2.9a**). We inserted a hairpin structure upstream of each IRES to block ribosomes initiating at the 5′ cap and ensure IRES-dependent initiation [35]. We co-transfected the various reporter constructs into HeLa cells together with control pCGT vector or T7-SF2/ASF cDNA. 40 h later, we measured luciferase activity and carried out radioactive RT-PCR of luciferase mRNA as a normalization control (**Fig. 2.9b**). As with cap-dependent translation, with the EMCV or the HCV IRES, SF2/ASF repressed translation in a manner that depended on the natural 3′UTR of SF2/ASF. In contrast, CrPV-IRES-dependent translation was not repressed by SF2/ASF overexpression (**Fig. 2.9b**). We conclude that SF2/ASF autoregulation takes place at the translation-initiation step, and that eIF3 and/or eIF2 may be involved in this effect.

2.3 Discussion

Negative autoregulation is an effective mechanism for homeostatic control of gene expression. SF2/ASF is an abundant and highly conserved RNA-binding protein with multiple functions and oncogenic potential, whose expression level needs to be precisely controlled for normal cell physiology. Post-transcriptional regulation of splicing factors can be complex, involving multiple layers of control. For example, PTB antagonizes the expression of its paralog, nPTB, by promoting an NMD-targeted alternative splicing isoform, and possibly also by inhibiting translation of correctly spliced mRNA through an unkown mechanism [36-37]. During neuronal differentiation, PTB expression is repressed by the neuron-specific miR-124, resulting in increased nPTB protein [37]. nPTB expression is also repressed during myoblast differentiation by the muscle-specific miR-133 [38]. Our study shows that multiple levels of post-transcriptional and translational control are likewise involved in fine-tuning SF2/ASF expression.

We identified and characterized six alternatively spliced mRNA isoforms of SF2/ASF in HeLa cells, of which isoforms IV and VI are not shown in the UCSC or ENSEMBL browsers. The major isoform, I, encodes full-length protein, and has a long 3′UTR [25-26]. Isoform II, which retains the third intron, was previously reported [25, 39]. A third isoform was also described in these studies, involving an alternative 3′ splice site in the third intron. We used a specific primer to amplify that isoform, but did not detect it in the cell lines we tested. Isoforms II and III retain the third intron, which changes the reading frame and results in a stop codon shortly after exon 3; this would result in a truncated protein without the C-terminal RS domain. However, we found that these two isoforms are retained in the nucleus, and are therefore not translated. This explains why our SF2/ASF antibody, which recognizes an epitope near the N-terminus, fails to detect any smaller protein isoforms by Western blotting [7].

In general, intron-containing pre-mRNAs are retained in the nucleus, and only mature mRNAs are exported to the cytoplasm, preventing translation of incompletely processed messages [40]. Interestingly, isoform IV retains one intron, compared to

isoform V, and it remains nuclear; however, the major isoform I retains that plus one additional intron, but somehow is compatible with efficient nuclear export, which might involve potential RNA cis-acting elements that are recognized as export signals. Many retroviruses and some cellular mRNAs, such as Tap, employ this mechanism [40-41].

Isoforms III, IV, V, and VI are generated by splicing that removes one or two introns in the 3′UTR. Among these, isoforms V and VI are exported to the cytoplasm and accumulate after cycloheximide treatment, suggesting that they are NMD targets. Isoform V encodes the same full-length protein as isoform I, whereas isoform VI encodes a truncated protein lacking the RS domain. SF2/ASF overexpression upregulates the unproductive isoforms III and VI, and decreases the protein-encoding major isoform I, but only modestly. Quantitation of the changes at the mRNA and protein levels indicates that alternative splicing associated with NMD or nuclear retention only partly explains the autoregulation of SF2/ASF.

By co-transfecting a V5-tagged genomic SF2/ASF construct with a T7-tagged SF2/ASF cDNA, we recapitulated the autoregulation seen with endogenous SF2/ASF. Co-transfection experiments with different mutants showed that RRM2 is required, and the 3′UTR is the only critical cis-element for the regulation. The length of the 3′UTR affects basal expression, but is not responsible for autoregulation.

Post-transcriptional regulation is frequently mediated by RNA-protein interactions in the UTRs [42], and this is also where the two UCRs are located in SF2/ASF (**Fig. 2.2a**) [20-23]. We tried to map cis-element(s) required for downregulation, but were unable to narrow them down to well-defined sequences. First, when the 3′UTR was divided into four fragments, three still showed downregulation by SF2/ASF overexpression (**Fig. 2.10**). Second, when each of the functional fragments was further subdivided, each subfragment gave much less or no repression (not shown). It appears that multiple elements in the 3′UTR are involved in SF2/ASF autoregulation, and the signals are dispersed and partially redundant. The roles of the two UCRs remain unclear, especially considering that the entire 3′UTR of SF2/ASF is ~95% conserved between human and mouse.

A recent quantitative-proteomics study showed that each miRNA has hundreds of target genes, but individual genes are only modestly repressed by a single miRNA [43]. Therefore, several miRNAs might target multiple regions in this 3′UTR, with their combined action resulting in downregulation. However, the experiments with Dicer-disrupted or -knockout cell lines suggest that miRNA-mediated repression is not the main mechanism of SF2/ASF autoregulation, although it may contribute to some extent. Indeed, miR7 was recently found to reduce SF2/ASF levels through a single binding site in the 3′UTR.

Using a sucrose-gradient assay, we found that SF2/ASF overexpression reduces the translational efficiency of an SF2/ASF-3′UTR-containing mRNA reporter. However, we could not recapitulate the translation inhibition by adding purified SF2/ASF protein to a cell-free translation system. Possible reasons for this include: i) a component(s) required for translation inhibition might be lost during extract preparation; ii) SF2/ASF does not repress translation directly, but could instead affect alternative splicing of a translational regulator; iii) the substrate for translational regulation might be a 3′UTR in the form of mRNP generated by a defined pathway, involving transcription, processing, and export.

Translation is a cytoplasmic event, but surprisingly, a nuclear-retained version of SF2/ASF was still able to autoregulate (**Fig. 2.3c**). Perhaps nuclear SF2/ASF affects the mRNP composition of its own transcript, which in turn affects how efficiently it is translated in the cytoplasm. Nuclear events often determine the downstream cytoplasmic fate of mRNAs [44]. It is also possible that SF2/ASF regulates translational control indirectly through its nuclear functions, such as splicing of a putative translational regulator's pre-mRNA. Finally, nuclear retention of the SF2/ASF-NRS variant might be slightly leaky. However, SF2/ASF can enhance translation of reporter mRNAs in a binding-site-dependent manner, which can be recapitulated in the cell-free system [6]; this effect, which is reproducible in our hands (not shown), requires the shuttling activity of SF2/ASF, and the nuclear-retained mutant is no longer active [6].

Our experiments with viral IRES elements suggest that SF2/ASF translational autoregulation is cap-independent, and that eIF2 and/or eIF3 are important, although the exact mechanism remains unknown. On the other hand, SF2/ASF enhances cap-dependent translation by repressing the activity of 4E-BP, an inhibitor of eIF4E, and no enhancement was observed for IRES-dependent translation [45]. Therefore, we believe that these two opposite effects of SF2/ASF in translation involve distinct mechanisms, and are not contradictory.

A recent study showed that SF2/ASF binds to its own transcript within the second UCR in the cytoplam, and enhances polysome association [46]. Although we observed neither translational repression nor activation by *in vitro* translation of a reporter with the SF2/ASF 3′UTR, it is possible that the long 3′UTR mediates complex positive as well as negative regulation, and that different mechanisms are dominant depending on the context.

SF2/ASF autoregulation is a complex process involving multiple mechanisms operating at different levels. We found that both alternative splicing and translation have contributing roles, and SF2/ASF translation itself may be negatively regulated at different steps by different factors. Multi-level regulation presumably serves to control SF2/ASF homeostasis more precisely. The relative contribution of each control mechanism might vary in different tissues or physiological states. Conversely, particular control mechanisms may be disrupted in different tumors associated with SF2/ASF upregulation [11].

2.4 Methods

2.4.1 Plasmids

The T7-tagged SF2/ASF, SRp30c, and SC35 constructs are in the pCGT vector; SRp30c, SC35, SF2/ASF wild type, and the NRS variant have been described [47-48]. We used quick-change mutagenesis to construct the SF2/ASF ΔRRM1, ΔRRM2, and ΔRS mutants.

We subcloned V5-tagged genomic SF2/ASF and V5-SF2/ASF ΔUTR into the EcoRI and XhoI sites of pcDNA3.1+ (Invitrogen). We used two-step cloning to construct V5-SF2/ASF Δintron123, pucUTR and UTR fragments A/B/C/D. First we cloned the V5-SF2/ASF cDNA coding region into pcDNA3.1+ via NheI and BamHI sites. Then we cloned the different 3′UTRs after the cDNA via BamHI/BglII and XhoI sites. We used a similar strategy to construct Rluc-SF2 UTR and Rluc-pucUTR. To mutate the splice sites in SF2/ASF 3′UTR we used site-directed mutagenesis. To construct the hp-IRES Renilla luciferase reporters, we used three-steps cloning. First, we inserted the hairpin sequence GCCUAGGCCGGAGCGCCCAGAUCUGGGCGCUCCGGCCUAGGC [35] into pcDNA3.1+ via NheI and BamHI sites. We amplified the EMCV IRES by PCR from the pWZL vector (gift from Dr. Scott Lowe's lab). We amplified the HCV and CrPV IRES from pAR233 HCV la IRES and pAR237 CrPV IRES, respectively, which were generously provided by Dr. Vincent Racaniello (Columbia University). We cloned the IRES fragments after the hairpin using the BamHI and EcoRI sites. Finally, we inserted Renilla luciferase, with or without the SF2/ASF 3′UTR, after the IRES using the EcoRI and XhoI sites. We also subcloned T7-SF2/ASF cDNA into the XhoI and EcoRI sites of the inducible vector STP [24].

2.4.2 Cell culture and transfection

We cultured HeLa and RKO cells in DMEM supplemented with 10% (v/v) FBS, 100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. To transfect plasmids, we used Fugene 6 (Roche). We grew HeLa tet-off cells in the same medium, with 2 μg ml^{-1} tetracycline. To generate stable cell lines, we infected Hela tet-off cells with STP retroviral vectors with an SF2/ASF cDNA, and selected stable transductants with puromycin (2 μg ml^{-1}) for 72 h. To induce SF2/ASF, we placed the cells in medium lacking tetracycline. To grow ES cells, we used DMEM knockout medium containing 15% (v/v) FBS, 100 μM β-mercaptoethanol, 2 mM L-glutamine, 50 U ml^{-1} penicillin, 40 ug ml^{-1} streptomycin, and 1000 U ml^{-1} LIF (Chemicon). We seeded the ES cells on plates coated with gelatin (Chemicon), and transfected them with Lipofectamine 2000 (Invitrogen).

2.4.3 Western blotting

48 h after transfection, we harvested the cells and lysed them in Laemmli buffer. The primary antibodies included β-catenin (Sigma), SF2/ASF (mAb AK96), T7 tag (Novagen), and V5 tag (Invitrogen). The secondary antibodies were goat anti-mouse IgG (H+L) HRP-conjugated (Pierce), labeled with yellow-fluorescent Alexa Fluor 532 dye (Invitrogen), or with IRDye 800CW (LI-COR). For detection we used an ECL kit (Roche), an Image Reader FLA-5100 (FujiFilm Medical Systems), or an Odyssey Infrared Imaging System (LI-COR), respectively.

2.4.4 RNA isolation and RT-PCR

To isolate total RNA, we used Trizol (Invitrogen) and treatment with RQ1 DNase I (Promega). For first-strand cDNA synthesis, we used random hexamers and Super Script II reverse transcriptase (Invitrogen). For regular PCR we used AmpliTaq (Roche); to amplify all the SF2/ASF isoforms we used rtTh (Roche) and Vent (New England Biolabs) DNA polymerases. For radioactive PCR, we added γ-^{32}P-dCTP and amplified for 24 cycles. We separated the PCR products on 6% non-denaturing polyacrylamide gels, and detected them with the Image Reader FLA-5100. Primers: GAPDH-F (5′-AAGGTGAAGGTCGGAGTCAACGG-3′), GAPDH-R (5′CCACTTGATTTTGG AGGGATCTC-3′); SF2-e1F (5′-ACATCGACCTCAAGAATCGCCGC-3′), SF2-e4R (5′-GGGCAGGAATCCACTCCTATG-3′), SF2-e3F (5′-CACTGGTGTCGTGGAG TTTGTACGG-3′), SF2-e3R (5′-TCCACGACACCAGTGCCATCTCG-3′); V5-F (5′-GGCAAGCCCATCCCTAACCC-3′); Rluc-F (5′-GACTTCGAAAGTTTATGATCC-3′), Rluc-R (5′-GCTCATAGCTATAATGAAATGCC-3′).

2.4.5 Cell fractionation

We lysed cells in gentle lysis buffer (10 mM HEPES pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.5% (v/v) NP-40). We pelleted the nuclei at 2300 g for 5 min at 4 °C, and transferred the supernatant (cytoplasm) to another tube. We washed the nuclei once with the same buffer and repelleted them. To extract RNA, we added Trizol to the pellet and the first supernatant.

2.4.6 Luciferase reporter assay

We lysed HeLa cells using passive lysis buffer (Promega) and measured the levels of Renilla luciferase using Promega's Dual Luciferase Assay Kit and a Monolight 2010 luminometer (Analytical Luminescence Laboratory). To extract the RNA, we added Trizol to the remaining lysates.

2.4.7 *In vitro* translation assay

We prepared translation-competent HeLa cell-free extracts as described [30]. We linearized the Renilla luciferase reporter construct with XhoI and used it as a template for *in vitro* transcription with T7 RNA polymerase using an mMessage mMachine Kit (Ambion). We added a poly (A) tail using a Poly (A) Tailing Kit (Ambion). Translation reactions included 20 ng of reporter mRNA with or without 200 ng of recombinant SF2/ASF protein, purified from bacteria or 293E cells [26, 49], and were incubated at 37 °C for 30 min. We added 50 µl of passive lysis buffer (Promega) to stop the reactions. We measured luciferase activity with a Dual Luciferase Assay Kit (Promega).

2.4.8 Sucrose gradient assay

We used one 150-mm plate of cells for each assay. We prepared 10% (w/v) and 50% sucrose in 20 mM Tris-HCl pH 7.4, 100 mM KCl, 5 mM $MgCl_2$. We split transfected cells once, 12 h before harvesting. We treated HeLa cells with 50 µg ml^{-1} cycloheximide at 37 °C for 20 min. We washed the cells with ice-cold PBS containing 50 µg ml^{-1} cycloheximide and lysed in 500 µl of polysome-extraction buffer (20 mM Tris pH 7.5, 5 mM $MgCl_2$, 100 mM KCl, 0.5% (v/v) NP-40, 100 U of RNasin (Promega)). Where indicated, we added puromycin (100 µg ml^{-1}) 1 h before harvesting, and omitted cycloheximide. We spun the lysates at 13,000 g for 10 min, after a 10-min incubation on ice. Then, we layered 500 µl of each cytoplamic lysate onto 10-50% sucrose gradients and centrifuged at 4 °C in a Sorvall SW41 rotor at 36,000 rpm for 2 h. We collected fractions from the top using a BioComp gradient master, while measuring the OD at 254 nm. We treated fractions with 1% (w/v) SDS and 150-200 µg ml^{-1} Proteinase K (Roche).

We extracted RNA with phenol/chloroform/isoamyl alcohol (25:24:1), treated it with RQ1 DNase I (Promega), and analyzed it by RT-PCR [6, 50].

2.5 Acknowledgments

We thank Bert Vogelstein (Johns Hopkins University) for generously providing the DicerEx5/Ex5 RKO cell line, Greg Hannon (Cold Spring Harbor Laboratory) for sharing the $Dicer^{-/-}$ and $Dicer^{+/-}$ ES cells, and Vincent Racaniello (Columbia University) for the gift of HCV and CrPV IRES plasmids. We thank Chenghai Xue for statistical analysis, Yang Yu for advice on polysome gradients, and Jun Zhu for helpful discussions and communicating unpublished results. This work was supported by NCI grant CA13106.

2.6 Figures and Figure Legends

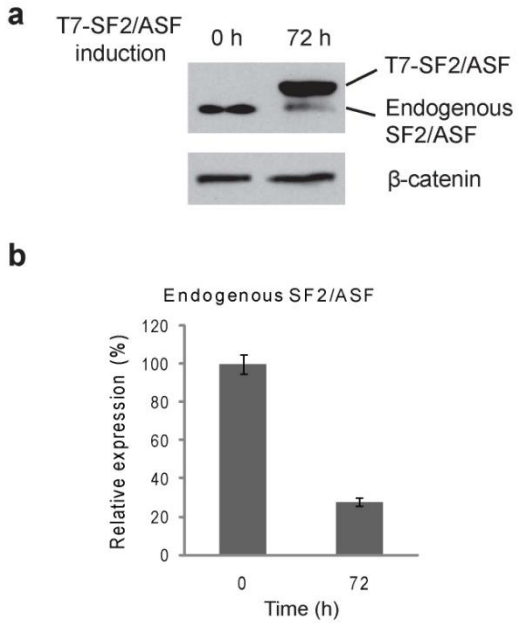

Figure 2.1 HeLa tet-off cells with inducible SF2/ASF overexpression. (**a**) Western blot analysis of SF2/ASF before and after induction, using an antibody that recognizes both endogenous and epitope-tagged SF2/ASF. (**b**) Quantification of endogenous SF2/ASF protein before and after induction. Error bars show standard deviations (SD); n = 3.

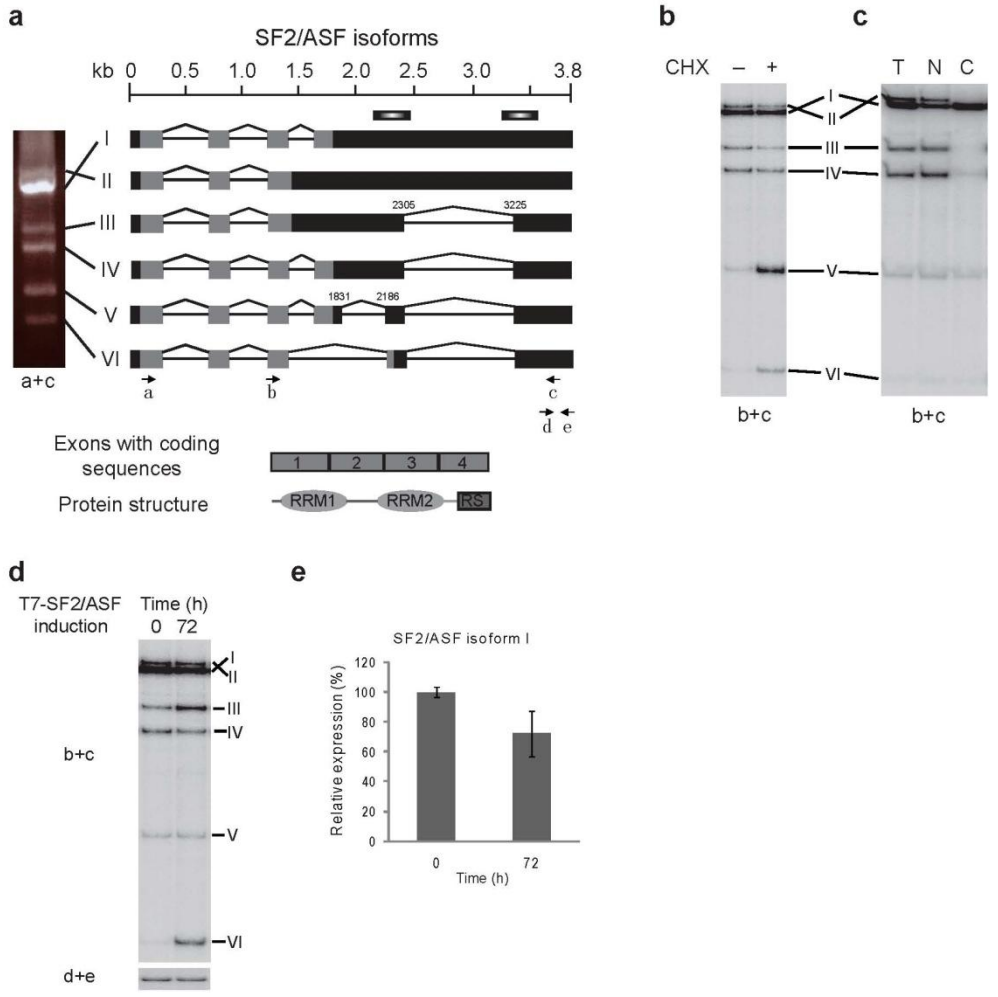

Figure 2.2 Alternative splicing of SF2/ASF. (**a**) Six alternative splicing isoforms were identified by RT-PCR with primers a and c, followed by cloning and sequencing. Their structures are shown in the diagrams, with the genomic scale shown at the top. The numbers above isoforms III and V represent the positions of the first and last nucleotide of the novel introns. The primers are indicated by arrows below the isoform diagrams. Grey represents protein-coding regions; black denotes the UTRs. The two bars below the genomic scale represent the positions of UCRs [20-23]. The correspondence between exon sequences and the domain structure of the protein—including two RNA-recognitions motifs (RRM) and an arginine/serine-rich (RS) domain—is shown at the bottom of the panel. (**b**) Cycloheximide treatment was used to inhibit NMD. Radioactive RT-PCR with primers b and c was performed to detect the changes of all the alternative splicing isoforms. (**c**) Cell fractionation was performed to separate nucleus and cytoplasm. RNA from cells before fractionation and from nuclear and cytoplasmic fractions was extracted for radioactive RT-PCR with primers b and c. T: total; N: nucleus; C: cytoplasm. (**d**) RT-PCR of RNAs from the same cell samples as in Figure 1, before and after SF2/ASF induction. The bottom panel shows amplification of a region common to all the isoforms, using primers d and e. (**e**) Quantification of the SF2/ASF protein-coding mRNA, isoform I, before and after induction. Error bars show SD; n = 3. t-test, $P < 0.04$

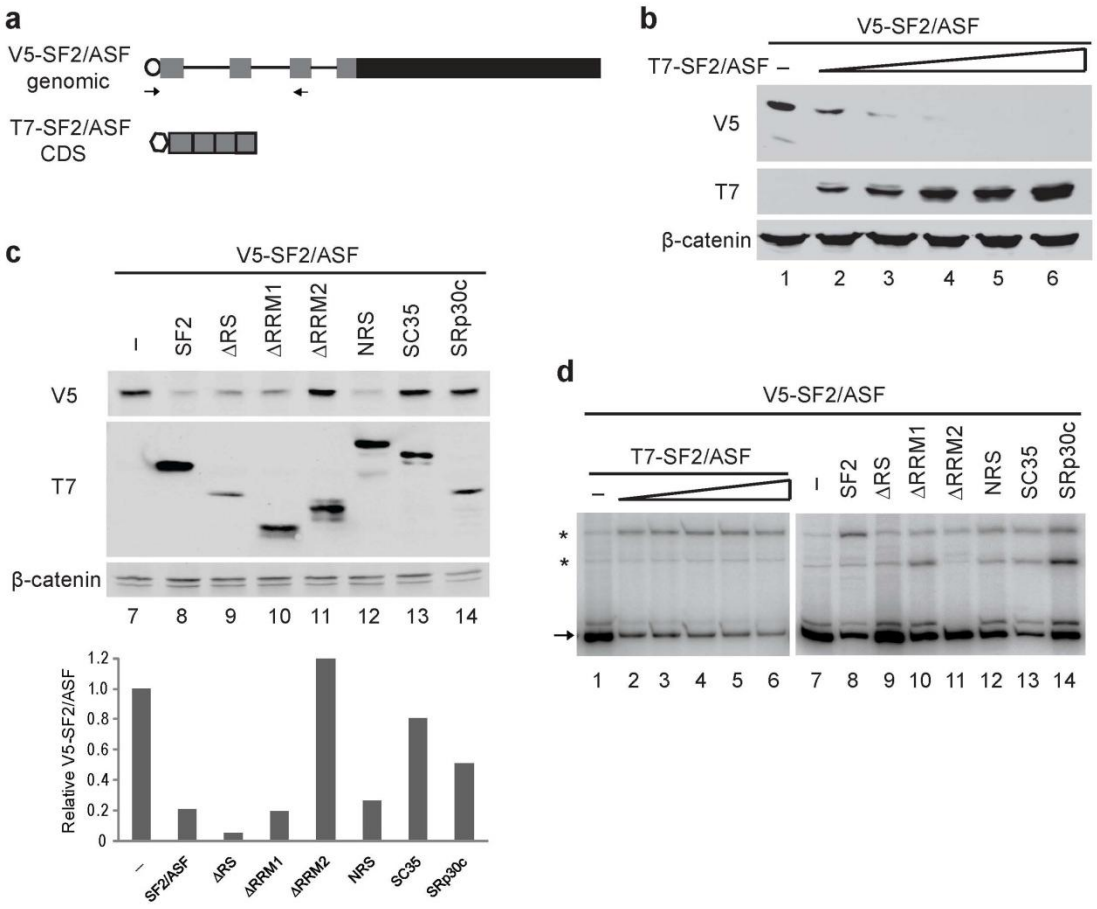

Figure 2.3 Expression of SF2/ASF from a genomic construct. (**a**) Diagrams of the V5-tagged SF2/ASF genomic construct and T7-tagged SF2/ASF cDNA construct. The V5 epitope tag is indicated by an open circle, and the T7 tag by an open hexagon. RT-PCR primers used in panel **d** are indicated by arrows. (**b**) V5-tagged genomic SF2/ASF was co-transfected with increasing amounts of T7-SF2/ASF cDNA into HeLa cells. After 48 h, protein and RNA were isolated, and Western blotting was performed with both V5 and T7 antibodies, with β-catenin as a normalization control. (**c**) Genomic V5-SF2/ASF was co-transfected with various T7-tagged SF2/ASF mutants and other SR protein cDNAs. Western blotting was performed with both V5 and T7 antibodies. The histogram shows the quantification of the relative V5-SF2/ASF expression level. The level of V5-tagged SF2/ASF was measured and normalized to that of each T7-tagged protein, with wild-type T7-SF2/ASF as the standard. The level of V5-SF2/ASF co-transfected with empty vector was set at 1. (**d**) RT-PCR of V5-SF2/ASF with one primer in the V5 tag and the other in SF2/ASF exon 3. The band corresponding to spliced mRNA is indicated by an arrow. *, RNAs that retained one or more introns.

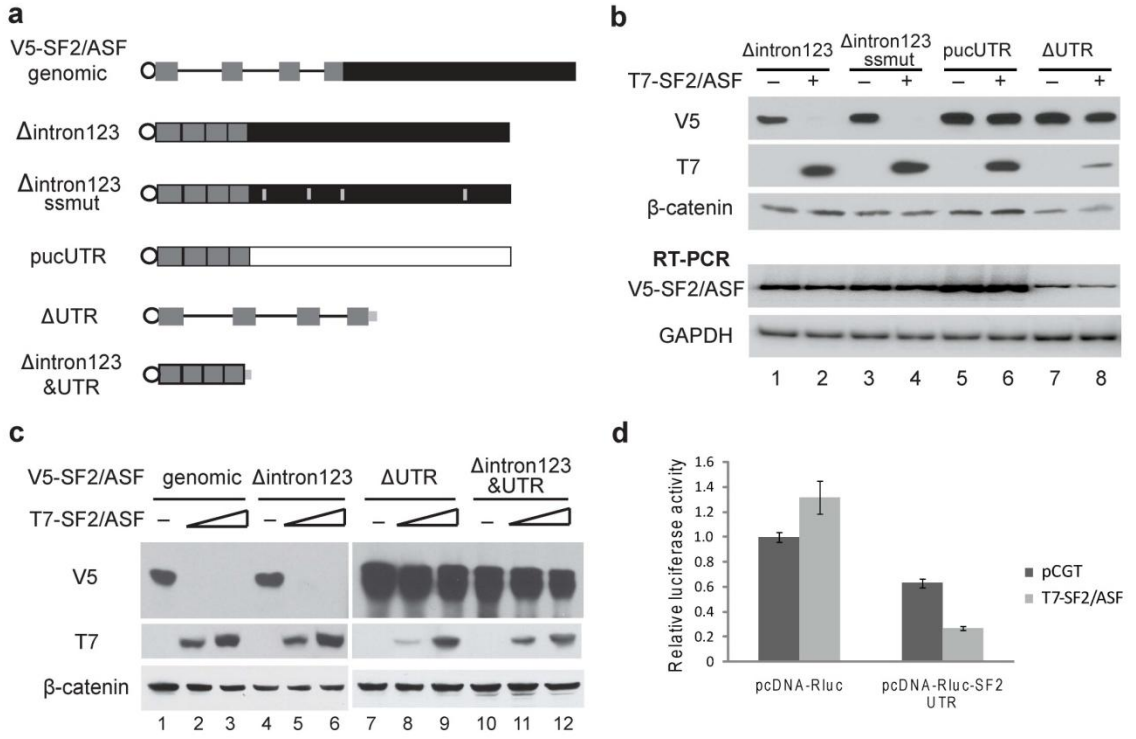

Figure 2.4 The 3′UTR is necessary and sufficient for SF2/ASF autoregulation. (**a**) Diagrams of the genomic V5-SF2/ASF mutants. Grey represents the coding region; black represents the natural 3′UTR; white represents a heterologous sequence of the same length; and the thin light-grey bar represents the 3′UTR sequences from the vector; the V5 tag is denoted by an open circle; the grey vertical lines in the 3′UTR represent inactivating mutations of the alternative 5′ and 3′ splice sites. (**b**) HeLa cells were co-transfected with V5-SF2/ASF genomic mutants and T7-SF2/ASF cDNA. Western blotting was performed to detect SF2/ASF expressed from the genomic construct using V5 antibody, from the cDNA using T7 antibody, and endogenous β-catenin was detected as a loading control. RT-PCR was carried out using the same primers as in **Fig. 3d**, with GAPDH as a reference. Deletion of the 3′UTR results in much more efficient translation, so in lanes 7 and 8 we transfected only 1/5 as much reporter plasmid, and loaded 1/2 as much total protein. (**c**) HeLa cells were co-transfected with V5-SF2/ASF genomic mutants and increasing amounts of T7-SF2/ASF cDNA. Western blotting was performed to detect SF2/ASF expressed from the genomic construct using V5 antibody, from the cDNA using T7 antibody, and endogenous β-catenin was detected as a loading control. (**d**) Luciferase reporter assay. The 3′UTR of SF2/ASF was fused to a Renilla luciferase reporter gene. The reporter was co-transfected into HeLa cells with control vector or SF2/ASF cDNA. Luciferase activity was measured and normalized to the luciferase mRNA level determined by radioactive RT-PCR. The relative luciferase activity of pcDNA-Rluc in the absence of SF2/ASF was set at 100%. The error bars show SD; n = 3. t-test, $P<0.0001$.

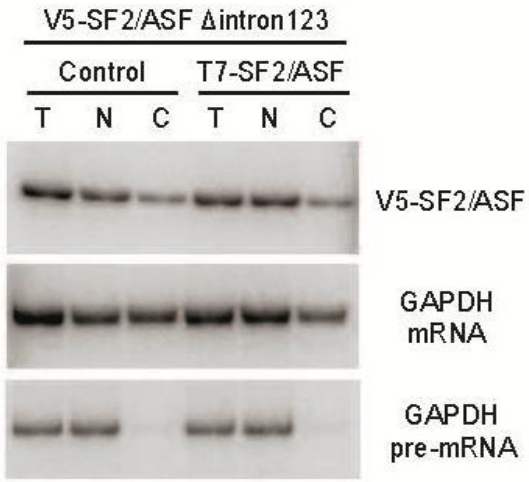

Figure 2.5 SF2/ASF overexpression does not inhibit export of its own mRNA Subcellular fractionation of HeLa cells co-transfected with V5-SF2/ASF Δ intron123 and T7-SF2/ASF cDNA. RT-PCR from total (T), nuclear (N), and cytoplasmic (C) RNAs was performed as in **Fig. 2b**. GAPDH pre-mRNA was amplified using one intronic and one exonic primer, as a control for the fractionation.

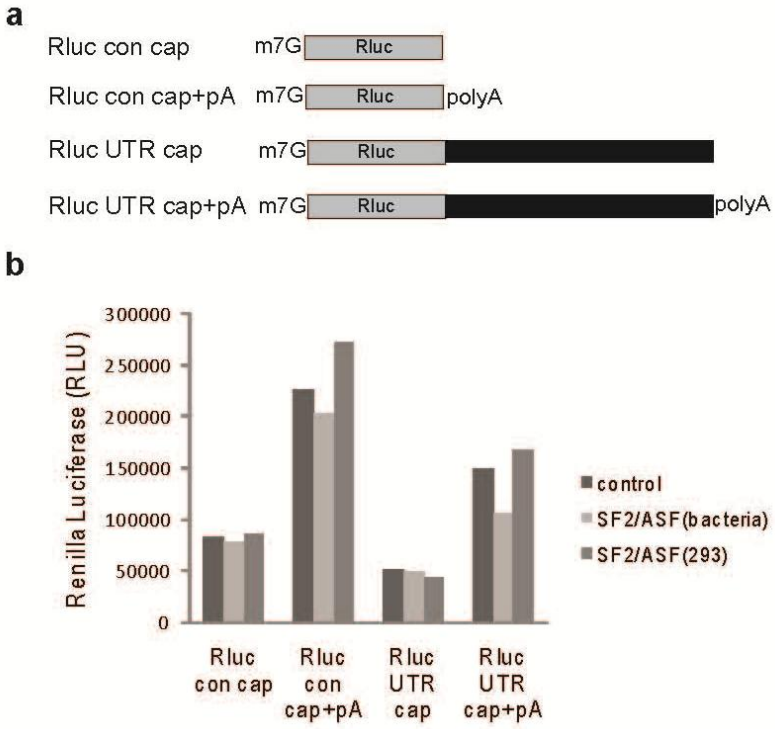

Figure 2.6 SF2/ASF does not show autoregulation in an in vitro translation assay (**a**) Diagrams of the Renilla-luciferase reporter mRNAs, with or without the SF2/ASF 3′UTR (black box). The mRNAs were generated by in vitro transcription and either 5′-capped only, or both capped and polyadenylated. (**b**) The mRNAs were incubated in translation-competent HeLa cell extract. Recombinant SF2/ASF proteins purified from either bacteria or 293E cells were present in the indicated reactions. Luciferase activity was measured after incubation for 30 min.

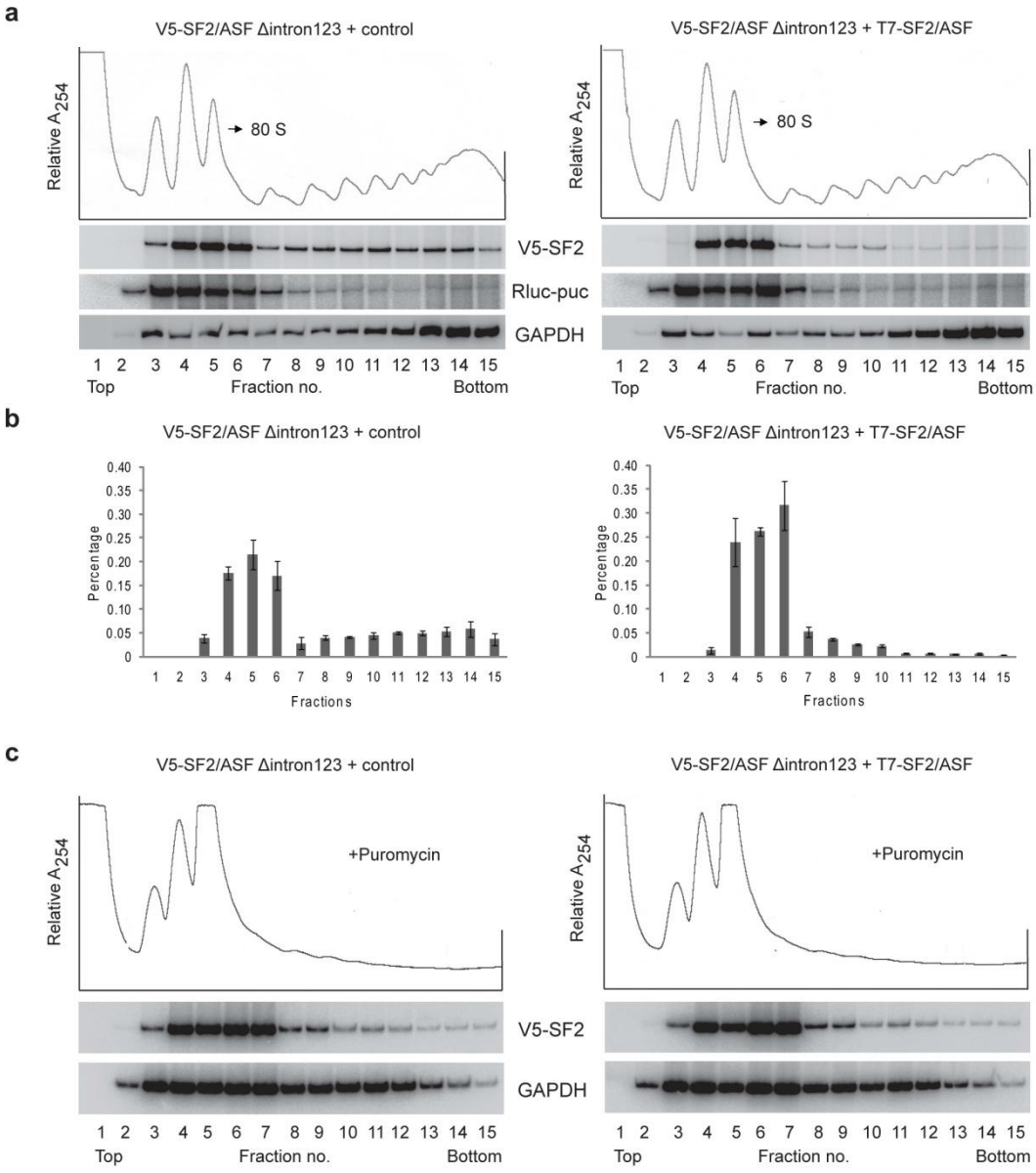

Figure 2.7 SF2/ASF reduces the polysome association of its own mRNA. (**a**) Sucrose-gradient fractionation of cytoplamic extracts from HeLa cells expressing V5-SF2/ASF Δintron123, Rluc-pucUTR, with (right panel) or without (left panel) T7-SF2/ASF cDNA. Top, UV absorbance (254 nm) profile. Middle and bottom panels, RNA extracted from each fraction was analyzed by radioactive RT-PCR to amplify V5-SF2/ASF, Rluc-pucUTR, and endogenous GAPDH mRNAs. (**b**) Quantitation of V5-SF2/ASF mRNA distribution in polysome gradients. Relative mRNA levels in each fraction were calculated as percentage of the total levels from all the fractions. Error bars indicate SD; n = 3. (**c**) Transfected HeLa cells were treated with puromycin for 1 h prior to lysis and fractionation. Gradient fractionation and analysis were done as in **a**.

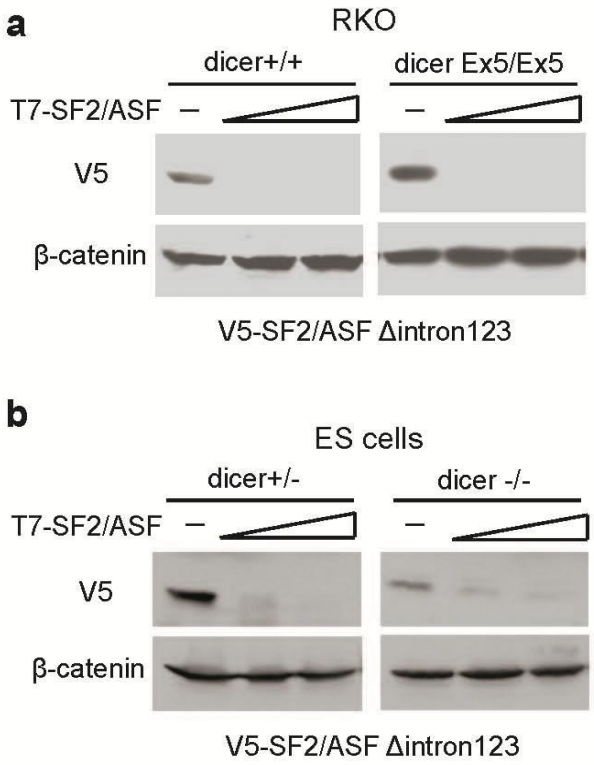

Figure 2.8 SF2/ASF autoregulation is resistant to disruption of the miRNA-processing pathway. (**a**) V5-SF2/ASF was co-transfected without or with two different amounts of T7-SF2/ASF into either wild type or Dicer-disrupted RKO cells [32]. (**b**) The same constructs were co-transfected into either heterozygous or Dicer-null knockout mouse ES cell [33].

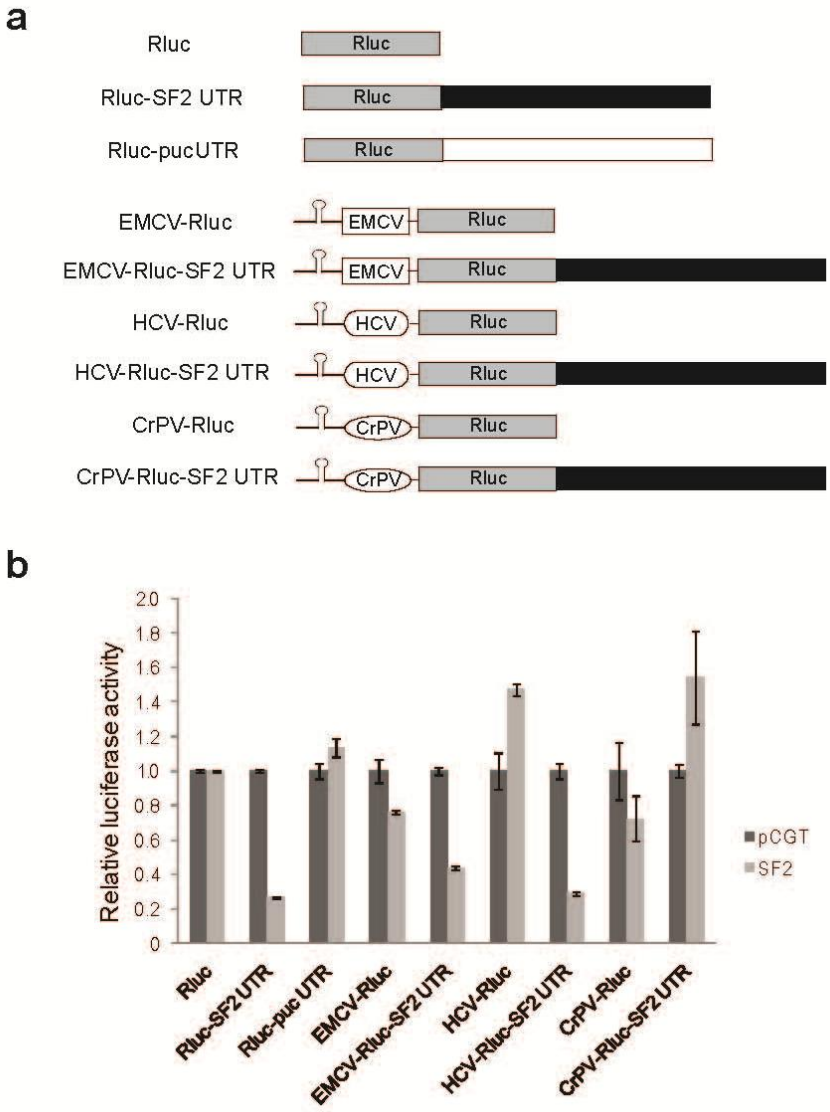

Figure 2.9 IRES-dependent translation assay. (**a**) Diagrams of EMCV, HCV, and CrPV IRES Renilla-luciferase reporter constructs, with or without the SF2/ASF 3′UTR (black box) or a heterologous sequence of the same length (white box). A hairpin was placed upstream of each IRES to inhibit cap-dependent translation. (**b**) Luciferase assay. The various Renilla-luciferase reporter constructs were co-transfected into HeLa cells with empty vector or T7-SF2/ASF cDNA. Luciferase activity was normalized to the luciferase mRNA level, measured by radioactive RT-PCR, as in **Fig. 4c**, and the percent change in the presence of SF2/ASF, compared to the activity in the absence of SF2/ASF, was plotted. The error bars show SD; n = 3.

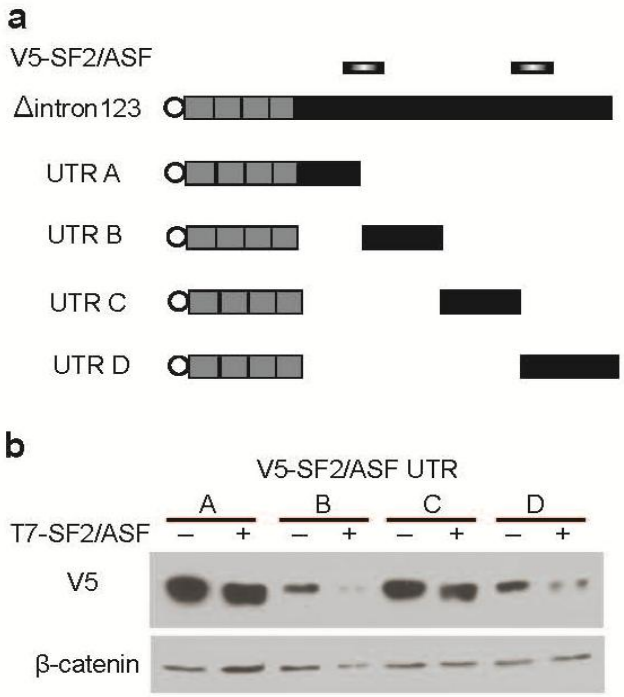

Figure 2.10 Multiple fragments in the SF2/ASF 3′UTR mediate auto-downregulation. (**a**) Diagrams of the V5-SF2/ASF constructs with different regions of the 3′UTR. The two bars on the top represent the positions of UCRs. (**b**) Each of the four V5-SF2/ASF constructs with partial 3′UTRs was co-transfected with T7-SF2/ASF cDNA or vector control. Western blotting was performed to detect the V5-tagged proteins and endogenous β-catenin.

Chapter 3

Mechanisms of activation and repression by the alternative splicing factor Fox-1

3.1 Introduction

Fox-1 was first identified by interaction with ataxin-2 [1], and therefore it has an alternative name: a2bp1 (ataxin-2 binding protein 1). A CAG repeat expansion in the ataxin-2 gene causes the neurodegenerative disease spinocerabellar ataxia, SCA2 [2]. Later, the ortholog of Fox-1 in *Caenorhabditis elegans* was shown to be critical for early embryonic patterning [3]. On the other hand, the UGCAUG hexamer was identified as a critical *cis*-element for regulation of alternative exons in several well-characterized minigenes [4-8]. However, the *trans*-acting factor that binds to this element and is responsible for alternative splicing regulation had not yet been discovered. Jin et al. first demonstrated that a zebrafish homologue of the putative RNA-binding protein Fox-1 specifically binds to GCAUG sequences in vitro and regulates splicing of alternative exons through this element [9]. Since this discovery, Fox-1 family proteins have been shown to regulate alternative splicing of a variety of genes with conserved (U)GCAUG element(s).

Fox-1 family proteins are characterized by a highly conserved RNA Recognition Motif (RRM) in human, mouse, zebrafish, fruitfly and nematode [10]. They all have one RRM in the middle and less conserved N- and C- terminal domains unique to Fox proteins. Unlike other splicing factors, which usually have very degenerate binding sites, the Fox-1 family proteins specifically bind to a (U)GCAUG element. There are three paralogues in mammals. Fox-1 is specifically expressed in brain, skeletal muscle, and heart [9, 11]. Fox-2, also known as RBM9 (RNA-binding motif protein 9), is more ubiquitously expressed [11]. Fox-3 is less well studied, and was recently discovered to be NeuN (neuronal nuclei), an antigen that has been used widely as a reliable marker for post-mitotic neuronal cells [12]. Both Fox-1 and Fox-2 have multiple isoforms, which are generated both by alternative promoters and alternative cassette exons [11, 13]. Both N- and C-terminal fragments of the encoded protein isoforms are highly diversified, and some isoforms also lack the second half of the RRM. Therefore, some of these isoforms have different activities or completely lose the splicing function [13]. The alternative

splicing of Fox-1 has been shown to be regulated during neuronal depolarization, as a way to modulate the activity of its target genes [14].

Fox-1 family proteins regulate alternative splicing positively or negatively in a position-dependent manner. They enhance exon inclusion when binding to the downstream intron of an alternative exon, while enhancing exon skipping when binding to the upstream intron. Several model target genes have been extensively studied by using reporter minigenes, such as mitochondrial ATP synthase γ-subunit (F1γ) [9], calcitonin/calcitonin-gene-related peptide (CGRP) [15], CaV1.2 L-Type calcium channel [16], fibronectin [9], non-muscle myosin II heavy chain-B (NMHC-B) [13], epithelial cell-specific fibroblast growth factor receptor 2 (FGFR2) [17], and Fox-1 and Fox-2 themselves [18]. Global analysis was utilized more recently to evaluate the splicing regulatory networks of the Fox-1 family proteins. The combination of microarray, CLIP-seq, high-throughput RT-PCR platform, and computational analysis by different groups gave rise to global identification of dramatically more endogenous target genes of Fox-1/2, and confirmed the so-called RNA map, meaning that the splicing activity (activation or repression) depends on the location of the UGCAUG element in a predictable manner [19-21].

Despite the progress in the identification of Fox-1/2 endogenous targets, little is known about the mechanisms of how Fox-1 proteins regulate alternative splicing positively or negatively. Most studies have focused on the repressing effect. Using the calcitonin/CGRP minigene as a model, Zhou et al. showed that Fox-1/2 prevent SF1 from binding to the branch point and repress the formation of spliceosomal E′ complex through binding to the upstream intron. The proteins also interfere with binding of Tra2β and SRp55 to ESEs (Exonic Splicing Enhancers) via a UGCAUG binding site in the exon, and block the formation of spliceosomal E complex [22]. Fukumura et al. used the exon 9 of F1γ as a model system, and they found that Fox-1 represses exon 9 by inhibiting the splicing of the downstream intron 9, which is a U1-snRNP-independent and U2-dependent splicing substrate [23]. Fox-2 has been shown to interact with hnRNP H1 and to help hnRNP H1 and F proteins to repress exon IIIc of an FGFR2 minigene by

anagonizing the binding of SF2/ASF [24]. The C-terminal portion of Fox-1 is critical for exon repression in the F1γ minigene [25]. However, the importance of this domain is not known in the case of exon activation as well as exon repression with U1-dependent introns. The only study which provided a mechanistic clue to the enhancing activity of Fox-1/2 was the discovery of an interaction between Fox-1 and U1-C (a U1-snRNP-specific subunit) in a yeast two-hybrid assay [26].

We have started to systematically explore the mechanism of Fox-1 regulation of alternative splicing. We used an MS2-tethering assay to evaluate the requirement of Fox-1 domains in both exon activation and repression. We found that the C-terminal portion is the only fragment that is critical for exon activation when tethered to the downstream intron. However, both the RRM and the C-terminal domain are required for exon repression when tethered to the upstream intron. We also used co-immunoprecipitation combined with mass spectrometry to identify proteins that interact with the Fox-1 C-terminal domain, and have begun to study their roles in Fox-1 activity. This work is currently in progress.

3.2 Results

3.2.1 The C-terminal domain of Fox-1 is sufficient for exon activation when tethered to the downstream intron

To elucidate the mechanisms of exon activation and repression by Fox-1, we first asked whether the RNA-binding domain and the flanking domains can be separated. We used an MS2-tethering assay to address this question. We replaced the RNA-binding domain of Fox-1 with the bacteriophage MS2 coat protein, which specifically binds to a 21-nt RNA stem-loop. We inserted this RNA element in the intron downstream of exon 7 in a human *SMN2* minigene [27] (**Fig. 3.1a**). The minigene lacks natural Fox-1 binding motifs. The alternative exon was predominantly skipped when co-tranfected with MS2 protein or Fox-1 alone. However, the MS2-Fox1(N,C) construct, in which the Fox-1 RRM was replaced by MS2, strongly induced exon 7 inclusion (**Fig. 3.1b**). This result

shows that Fox-1 can enhance exon inclusion when tethered to the downstream intron by MS2, and the RRM and the flanking domains can be separated. We next asked which domains are required for this effect. We only kept either the N- or C-terminal domain fused with MS2, and co-transfected them with the *SMN2* minigene. The results showed that the C-terminal domain but not the N-terminal domain was sufficient to induce exon 7 inclusion (**Fig. 3.1b**). We further subdivided the C-terminal domain into two halves. The second half, MS2-Fox1Cb, was fully functional in exon activation. The first half, MS2-Fox1Ca, was less active, promoting exon inclusion to a lesser extent (**Fig. 3.1b**). However, western blotting showed that the expression level of MS2-Fox1Ca was lower than those of the other mutants (**Fig. 3.1b**). It is possible that its weaker activity in exon activation was simply due to lower protein expression.

3.2.2 Defective mutants Fox-1N and Fox-1Ca are mislocalized, but correct localization is not sufficient to recover the activity on exon inclusion

We used indirect immunofluorescence against the Flag tag to examine the localization of the different transiently expressed mutants. While MS2 protein gave a diffuse pattern in both nucleus and cytoplasm, MS2-Fox1(N,C) was localized in the nucleus, the same as the wild type Fox-1 protein (**Fig. 3.2**). The two mutant proteins that were able to induce exon 7 inclusion, MS2-Fox1C and MS2-Fox1Cb, were also localized in the nucleus. However, both MS2-Fox1N and MS2-Fox1Ca, which completely or partially lost the activation potential, distributed in both nucleus and cytoplasm compartments (**Fig. 3.2**). This result raised the possibility that Fox1N and Fox1Ca could not enhance exon inclusion because of their mislocalization.

To address this question, we forced MS2, MS2-Fox1N and MS2-Fox1Ca to localize in the nucleus by fusing the SV40 nuclear localization signal (NLS) to the C-terminus of the mutant proteins. Immunofluorescence confirmed their nuclear localization (**Fig. 3.3b**). When co-transfected with the *SMN2* minigene, Fox1N-NLS still failed to promote exon inclusion. The Fox1Ca-NLS mutant was a little more active in exon activation when localized in the nucleus, compared to the mutant without the NLS, but still was not fully functional (**Fig. 3.3a**). However, this could be due to the low

expression level. Even though we tried to transfect more plasmid, the expression level of this mutant was still lower than the others. It is possible that this mutant protein is not stable and undergoes rapid degradation. From this result, we conclude that the N-terminal fragment of Fox-1 is not functional in enhancing exon inclusion.

3.2.3 Both the RRM and the C-terminal domain are important for exon repression when tethered to the upstream intron

Fox-1 usually promotes exon skipping when binding to the upstream intron of alternative cassette exons. We again used the MS2-tethering assay to examine its activity in exon repression. We modified the β-globin PB1 minigene [19] to address this question. We mutated the three Fox-1 binding sites in the upstream intron, as well as the single site in the exon. Then we inserted an MS2 binding site in the intron upstream of the alternative exon (**Fig. 3.4a**). The alternative exon was predominantly included when co-tranfected with MS2 protein or Fox-1 alone. The MS2-Fox1(N,C) mutant, which can enhance exon inclusion when tethered to the downstream intron (see above), could not promote exon skipping when tethered to the upstream intron (**Fig. 3.4b**). It is possible that the RRM is still required for the repressive function. Therefore, we fused the MS2 protein to the N-terminus of the whole Fox-1 protein. This mutant indeed enhanced exon skipping partially (**Fig. 3.4b**). Like all canonical RRMs, the one in Fox-1 utilizes the β-sheet to bind nucleic acids. Two exposed phenylalanine residues in the RNP (ribonucleoprotein) submotifs are essential for binding to RNA, by intercalating with single-stranded bases [28]. We mutated the two phenylalanines to alanines, which was reported to eliminate RNA binding [28-29], and tested whether it can still enhance exon skipping when tethered by MS2. To our surprise, the two-amino-acid mutation completely abolished the exon-repression activity, even though the protein still bound to the minigene through MS2 tethering (**Fig. 3.4b**). We speculate that the RRM, including specific residues directly interacting with RNA, is also involved in protein-protein interactions or RNA-protein nonspecific interactions, which are essential for exon-repression activity but not for exon-activation function. To test whether the N-terminal fragment is necessary for the repressive effect, we deleted it and kept the RRM and the

C-terminal domain after the MS2 protein. This mutant was fully functional in inducing exon skipping, similar to the full-length protein (**Fig. 3.4b**). When we further deleted the C-terminal domain and only kept the RRM fused to MS2, the protein completely lost the activity (**Fig. 3.4b**). Therefore, the N-terminal domain is not required for exon repression, while both the RRM and the C-terminal domain are essential.

3.2.4 Co-immunoprecipitation and mass spectrometry to identify Fox-1C interacting proteins

To explore the mechanisms responsible for exon activation by the C-terminal fragment of Fox-1, we performed co-immunoprecipitation and mass spectrometry to identify proteins that interact with MS2-Fox1C. We used MS2 as a negative control. As both mutants have a T7 tag right before the MS2 protein, we used T7 monoclonal antibody coated Dynabeads for the immunoprecipitation. The immunoprecipitated proteins were eluted and separated by SDS-PAGE and stained with Coomassie Blue. There were clear differences between the MS2-Fox1C immunoprecipitate and the MS2 negative control (**Fig. 3.5a**). We cut out eight gel bands that were only seen in the MS2-Fox1C sample, and the corresponding gel slices of the MS2 sample. Proteins were eluted from the gel slices and analyzed by mass spectrometry. Multiple protein candidates were identified, including several RNA-binding proteins. We initially focused on three candidate proteins: hnRNP H1, Raly, and TFG.

To confirm the protein interactions identified by mass spectrometry, we repeated the co-immunoprecipitation using T7 antibody followed by western blotting. To exclude the possibility that the protein interactions are mediated by binding to the same RNA, rather than reflecting direct contacts between the proteins, we also treated the cell lysates with nuclease before IP. As both MS2 and MS2-Fox1C have a Flag tag at the N-terminus, we used Flag antibody for western blotting to confirm the IP efficiency. As shown in **Fig. 3.5b**, TFG showed strong interaction with Fox1C, both without and with nuclease treatment. Raly gave a stronger signal when the lysate was pre-treated with nuclease, in total, post-IP and IP samples. A similar phenomenon has been observed in our lab with other RNA-binding proteins that show increased amounts by western blotting when the

lysate is pre-treated with nuclease. We hypothesize that these proteins are probably associated with large RNP complexes that precipitate with cell debris. After nuclease treatment, the RNP complexes are disrupted, and therefore more proteins are present in the supernatant. Nevertheless, both Raly and TFG interacted with Fox1C specifically, and the binding was not disrupted by nuclease treatment. We have technical difficulties in detecting hnRNP H1 by IP-western, because the protein co-migrates with the IgG heavy chain. We plan to co-express MS2 or MS2-Fox1C with V5-tagged hnRNP H1 and then perform IP-western to confirm their interaction.

3.2.5 Knockdown of hnRNP H/F, Raly and TFG showed only modest inhibition of exon activation induced by MS2-Fox1C

We carried out siRNA knockdown experiment to test the importance of the observed protein-protein interactions. As hnRNP H1 and hnRNP F are very similar and appear to have redundant functions, we knocked down both proteins at the same time. We also included Ataxin-2, which led to the initial discovery of Fox-1, as a control. We transfected siRNAs into HeLa cells, and after 48 h we transfected the reporter minigene together with either MS2 or MS2-Fox1c. We also transfected GFP at the same time as an internal control to normalize for transfection efficiency. We collected samples after another 24 h, and extracted both RNA and protein. Most of the proteins were knocked down efficiently, except for hnRNP H1, which only decreased by about 50% (**Fig. 3.6a**). After hnRNP H/F, Raly and TFG knockdown, the inclusion level of *SMN2* exon 7 showed modest decreases, compared to the control luciferase siRNA transfected cells (**Fig. 3.6b**). This result suggests that these proteins may assist Fox-1 in its exon activation function through interaction with the C-terminal domain. But as the effects were not dramatic, additional experiments, quantification, and knockdown combinations will need to be done to further investigate the potential involvement of these proteins in Fox-1-induced exon inclusion.

3.3 Discussion

We characterized the function of Fox-1 using an MS2-tethering assay. We found that the C-terminal domain, especially the second half, is important for both Fox-1 localization and exon activation function. It is recently reported that the sequence at the C-terminal end comprises a putative hPY-NLS, a nuclear localization signal recognized by karyopherin β2 [10, 30]. This sequence is conserved in both Fox-1 and Fox-2, from *C. elegans* to human, and it is also present in some other splicing factors, such as hnRNP A1, hnRNP D, hnRNP F, etc. This is consistent with our localization results. It has also been noted that some Fox-1 isoforms lack the second half of the C-terminal fragment because of frame-shifting, and therefore are not functional in splicing regulation [13-14].

The Fox-1 C-terminal domain alone is sufficient to promote alternative exon inclusion when tethered to the downstream intron. This suggests that like modular transcription factors and some splicing factors, such as SR proteins and hnRNP proteins, Fox-1 also has separate nucleic-acid binding and functional domains. However, both the RRM and the C-terminal domain are required for alternative exon repression when tethered to the upstream intron. This observation implies that different mechanisms are involved in Fox-1 activation versus repression. It is possible that the RRM is also involved in protein-protein interactions, which could be critical for exon repression but not for exon activation. Another alternative scenario is that Fox-1 also needs to interact with other RNA sequences—for example elsewhere on the pre-mRNA, or with an snRNA--besides the engineered MS2-binding site, for its repressive activity, especially considering that the amino acid substitutions in the canonical RNA-binding surface of the RRM completely abolished the repression activity. It is not unprecedented that different domains of a protein are required in different activity contexts. For example, all SR proteins have one or two RRMs and one RS domain for protein-protein interaction. SR proteins function in two distinct aspects of splicing. On one hand, they are required for constitutive splicing and spliceosome assembly. On the other hand, they also regulate alternative splicing when binding to exonic enhancer elements [31]. The RS domain alone is sufficient for enhancer-dependent splicing, but not for constitutive splicing in

MS2-tethering *in vitro* splicing assays; the latter reaction requires a full-length SR protein [32]. On the other hand, under some conditions, the RS domain is dispensable for general splicing in vitro [33-34].

We used co-immunoprecipitation and mass spectrometry to identify Fox-1 C-terminal domain interacting proteins. We selected three prominent candidates for further analysis: hnRNP H1, Raly, and TFG.

hnRNP H1 belongs to the superfamily of heterogeneous nuclear ribonuclewproteins (hnRNPs). It has three RRMs and an extensive glycine-rich region near the C-terminus. hnRNP H1 binds to intronic oligo-(G) sequences and regulates alternative exons negatively or positively, depending upon the context. hnRNP H1 is very similar to another hnRNP family member, hnRNP F. They have redundant functions in regulating alternative splicing. It was recently shown that hnRNP F and hnRNP H interact with the tissue-specific splicing factor Fox-2, and this interaction involves the C-terminal domain of Fox-2 [24]. Our experiments showed that Fox-1 can also interact with hnRNP H1. Considering that the C-terminal domains of Fox-1 and Fox-2 are more than 50% conserved, our findings confirm and extend the previous study, and also establish the specificity of the method we used to identify Fox-1 interacting proteins.

Raly was originally identified as an autoantigen that cross-reacts with EBNA-1 of Epstein-Barr virus in infectious mononucleosis [35]. It is a member of the hnRNP superfamily, and is also called hnRNP C-like 2. It has a single RRM at the N-terminus, and a short poly-Gly stretch near the C-terminus. It is associated with the spliceosomal C complex [36], and is therefore suspected to be involved in pre-mRNA splicing. However, there is no direct functional evidence of such involvement, and no extensive characterization of this protein so far.

TFG is also called TRK-fused gene. Its 5′-end sequence is fused to the 3′-end of NTRK1 by chromosomal rearrangement, generating the TRK-T3 oncogene, which is associated with thyroid papillary carcinoma [37]. TFG is part of signal transduction pathways. It interacts with the SH2 domain of SHP-1, a protein-tyrosine phosphatase, and

modulates its activity [38]. It also interacts with TANK and NEMO, two proteins involved in the NF-κB pathway [39]. There is no evidence that TFG can bind RNA or regulate RNA metabolism. However, based on our mass spectrometry result and validation, it interacts with Fox-1 very strongly. Therefore, we included TFG in our functional study.

We validated the interactions of Fox1C with TFG and Raly by IP-western, both with and without nuclease treatment. siRNA knockdown of hnRNP H+F, Raly, or TFG showed modest inhibition of exon activation induced by Fox1C, suggesting that they are involved in modulating Fox-1 activity. However, this result is preliminary, and needs further investigation.

3.4 Future Perspectives

I plan to do more experiments to validate the roles of hnRNP H1, Raly, and TFG in Fox-1 induced alternative exon inclusion. I will do triplicate experiments and quantitate the exon inclusion levels to determine whether the effects are statistically significant. I will also individually add back siRNA-resistant cDNAs of the three proteins and test whether this can rescue the full inhibition on exon inclusion, so as to exclude the possibility of non-specific RNAi effects. I also plan to titrate the amount of transfected MS2-Fox1C, and choose a point such that *SMN2* exon 7 is only partially included. Then I will co-transfect the cDNAs of hnRNP H1, Raly, or TFG, and test whether any of them can enhance the exon inclusion mediated by Fox-1. Hopefully these experiments will provide stronger evidence of whether any of these proteins indeed enhances the Fox-1 effect on exon activation.

It will also be useful to carry out mass spectrometry of the whole IP instead of proteins extracted from individual gel slices, so that we can get a complete picture of all proteins that interact with Fox-1C. It is also of interest to compare the differences in protein interactions between Fox1C and Fox1ΔN (which have both the RRM and the C-terminal domain), which may help to dissect the distinct functions of Fox-1. It will be

ideal if we can tether the protein to either downstream or upstream intron of the alternative exon, and perform mass spectrometry to look for proteins that interact with Fox-1 in these specific contexts. In theory we can co-transfect higher amounts of the *SMN2* minigene together with the MS2-fused mutant constructs and carry out IPs to immunoprecipitate the MS2 mutant proteins, as well as other proteins that interact with Fox-1 when it binds to this particular position in the context of splicing. However, one technical difficulty is that we will also pull down a lot of non-specific proteins that bind to the minigene RNA in both MS2 control and MS2-Fox-1 samples. This will make it difficult to dissect which proteins show real quantitative differences between the two IPs and are the true interacting partners. This background noise may be reduced by using a quantitative mass spectrometry method, called IDIRT (Isotopic Differentiation of Interactions as Random or Targeted) or SILAC (Stable Isotope Labeling with Amino acids in Culture). In this way, the real interacting proteins will be enriched in the MS2-Fox1 IP sample, while the non-specific binding proteins will have similar amounts in both the MS2 negative control and the MS2-Fox1 sample.

3.5 Methods

3.5.1 Plasmids

We inserted the MS2-binding site (5′-GCGTACACCATCAGGGTACGC-3′) into a minigene by site-directed mutagenesis. The sequences flanking the RRM in pcDNA-Flag-Fox1 [19] were mutated to introduce NcoRI and BamHI sites, and then the RRM was replaced by T7-tagged MS2 coding. We subcloned MS2-Fox1C, MS2-Fox1N and MS2-Fox1Ca into pcDNA-Flag via BamHI/BglII and XhoI sites. We deleted the first half of C-terminal fragment in MS2-Fox1C by site-directed mutagenesis to create MS2-Fox1Cb. We inserted an SV40 NLS (5′-CCTAAGAAGAAACGTAAGGTC-3′) at the C-terminus before the stop codon by mutagenesis to construct MS2-NLS, MS2-Fox1N-NLS, and MS2-Fox1Ca-NLS. We cleaved MS2-Fox1C with BamHI and XhoI to release the C-terminal fragment of Fox1, and subcloned the full-length Fox-1 via BglII and XhoI sites

to create MS2-Fox1, or subcloned both the RRM and C-terminal domain to construct MS2-Fox1ΔN. We mutated the two phenylalanines to alanines by site-directed mutagenesis.

3.5.2 Cell culture and transfection

We cultured HeLa cells in DMEM supplemented with 10% (v/v) FBS, 100 U ml^{-1} penicillin and 100 µg ml^{-1} streptomycin. We used Fugene 6 (Roche) to transfect plasmids and Oligofectamine (Invitrogen) to transfect siRNAs. For the knockdown experiment, we transfected siRNAs 48 h prior to plasmid transfection. We collected samples after another 24 h. The siRNA sequences were: hnRNP H1 5′-CAAACAACGUUGAAAUGGA-3′; hnRNP F 5′-CGACCGAGAACGACAUUUA-3′; Raly 5′-UAACGUACCUGUCAAG CUC-3′; TFG: 5′-GAGGAAAACUUCUGAGUAA-3′; Ataxin-2 5′-GCAAAUAUGAG GAUGGUUC-3′.

3.5.3 Western blotting

We harvested the cells and lysed them in Laemmli buffer. The primary antibodies included β-catenin (Sigma), Flag tag (Sigma), hnRNP H1 (CSHL facility), hnRNP F (Santa Cruz), Raly (Abcam), TFG (Abcam), and Ataxin-2 (BD). The secondary antibodies were goat anti-mouse or anti-rabbit IgG (H+L) HRP-conjugated (Pierce), or labeled with IRDye 800CW (LI-COR). For detection we used an ECL kit (Roche), or an Odyssey Infrared Imaging System (LI-COR).

3.5.4 RNA isolation and RT-PCR

To isolate total RNA, we used Trizol (Invitrogen) and treated with RQ1 DNase I (Promega). For first-strand cDNA synthesis, we used random hexamers and ImProm-II reverse transcriptase (Promega). For radioactive PCR, we used AmpliTaq (Roche), added γ-^{32}P-dCTP and amplified for 23 cycles. We separated the PCR products on 6% non-denaturing polyacrylamide gels, and detected them with an Image Reader FLA-5100 (Fuji). PCR primers: βglobin-F (5′- AGGAGAAGTCTGCCGTTACTG-3′); βglobin-R (5′- ATAACAGCATCAGGAGTGGAC-3′); SMN2-F (5′-TAATACGACTCACTATA GG-3′); SMN2-R (5′-TAACGCTTCACATTCCAGATCTGTC-3′); GFP-F (5′-CGA

TCAAGCTTGCCACCATGAGCAAGGGC-3′); GFP-R (5′-CATTAACCCTCACTAA AGGGAATTCCAGCTTGTGGCCGAG).

3.5.5 Immunofluorescence

We fixed HeLa cells with 4% (v/v) para-formaldehyde in PBS for 30 min, 48 h after transfection. We permeabilized cells by incubation in 0.2% (v/v) Triton X-100 for 5 min. We then incubated the cells with Flag antibody (5 µg/ml; Sigma) for 1 h, washed with PBS, and then incubated with Alexa Fluor 594-conjugated goat anti-rabbit IgG antibody (1:1000; Invitrogen) for 1 h. We imaged the cells with a fluorescence microscope (Axioskop, Carl Zeiss).

3.5.6 Immunoprecipitation and mass spectrometry

Dynabeads Protein G was washed twice with Citrate-Phosphate Buffer (25 mM citric acid, 50 mM Na_2HPO_4, pH 5.0), and incubated with T7-tag monoclonal antibody for 1 h at room temperature (2 ml culture supernatant per 10 µl beads). The beads were washed twice in 500 µl Citrate-Phosphate Buffer, and twice in 0.2 M triethanolamine pH 8.2 (Sigma). The beads were resuspended in 1 ml of 20 mM dimethyl pimelimidate (Sigma) in 0.2 M triethanolamine pH 8.2, and incubated at 20 °C for 30 min with gentle mixing to crosslink the antibody to the beads. To stop the reaction, the beads were resuspended in 1 ml of 50 mM Tris-HCl, pH 7.5, and mixed for 15 min at room temperature. Finally, the beads were washed three times in PBS. Four 15-cm plates of HeLa cells were lysed in 4 ml of lysis buffer (0.3% (v/v) NP-40, 200 mM NaCl, 50 mM Tris, pH 7.4, 1 mM DTT, 0.1 mM EDTA, 0.1 mM EGTA, with freshly added 1 mM sodium vanadate, 50 mM sodium fluoride and protease inhibitor cocktail), and sheared by passing sequentially through a syringe with 20G, 22G and 26G needles, three times each. Nuclease was added (1 U/ml RNase cocktail (Ambion), 500 U/ml Benzonase (Novagen) and 2 mM $MgCl_2$) and incubated on ice for 30 min. Lysates were then centrifuged at 13,000g for 20 min at 4 °C. The supernatant was passed through a 0.45-µm syringe filter (with HT Tuffryn membrane, Pall Corporation). 80 µl beads (1:1 suspension) was added to the cleared lysates and incubated for 1 h at 4 °C. After washing five times in lysis buffer, the beads

were resuspended in Laemmli buffer and loaded on an SDS-PAGE gel. For mass spectrometry analysis, the samples were run on a NuPAGE Novex Bis-tris Mini Gel (Invitrogen) and stained with Gelcode Blue Stain Reagent (Pierce).

3.6 Acknowledgements

We thank Zuo Zhang for the initial design at the beginning of the project. We thank Oliver Fregoso for the kind guidance on immunoprecipitation, and Cristian Ruse for mass spectrometry analysis. We thank Zhaozhu Qiu and Stephen Hearn for the microscopic imaging.

3.7 Figures and Figure Legends

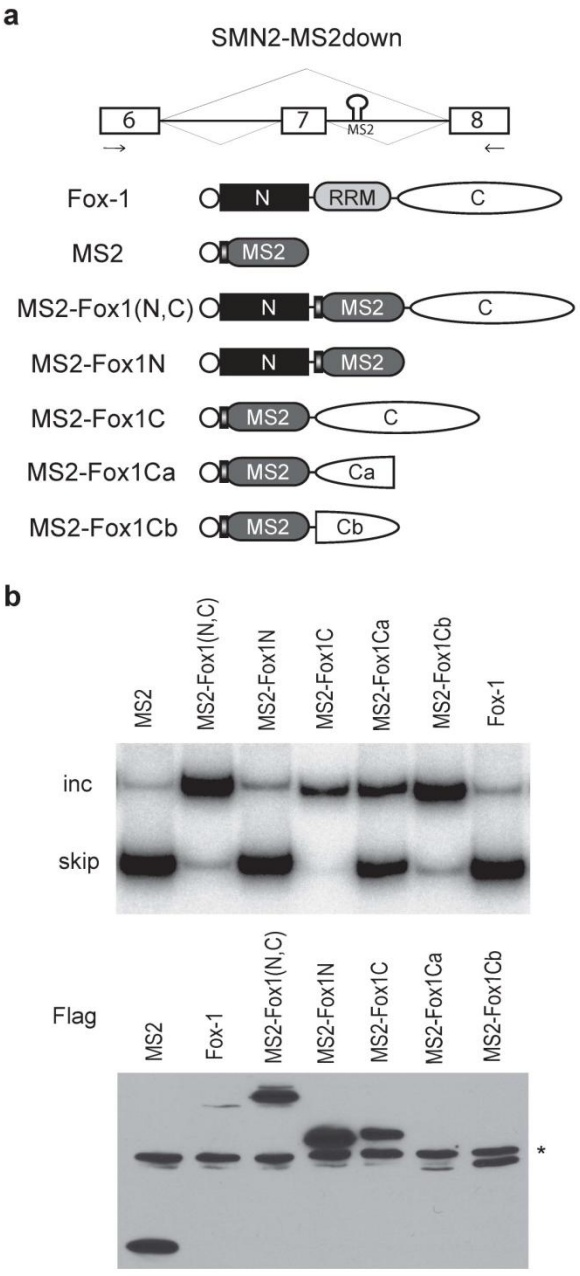

Figure 3.1 The C-terminal domain of Fox-1 is sufficient for exon activation when tethered to the downstream intron. (**a**) Diagrams of the modified *SMN2* minigene and MS2-fused Fox-1 mutants. MS2-binding site was inserted in the intron downstream of exon 7. The primers for RT-PCR are indicated by arrows below the exons. Fox-1 has one RRM (grey) flanked by N-terminal (black square) and C-terminal (white oval) domains. All the protein mutants have a Flag epitope tag at the N-terminus, indicated by an open circle. MS2 protein also has a T7 tag at the N-terminus represented by a bar. (**b**) HeLa cells were co-transfected with the SMN2-MS2down reporter minigene and different mutant proteins. RNA and proteins were extracted 48 h after transfection. Radioactive RT-PCR was performed to detect the changes of the alternative splicing isoforms (upper gel). Western blot analysis using Flag antibody was performed to show the expression of the protein mutants (lower gel). *, non-specific band.

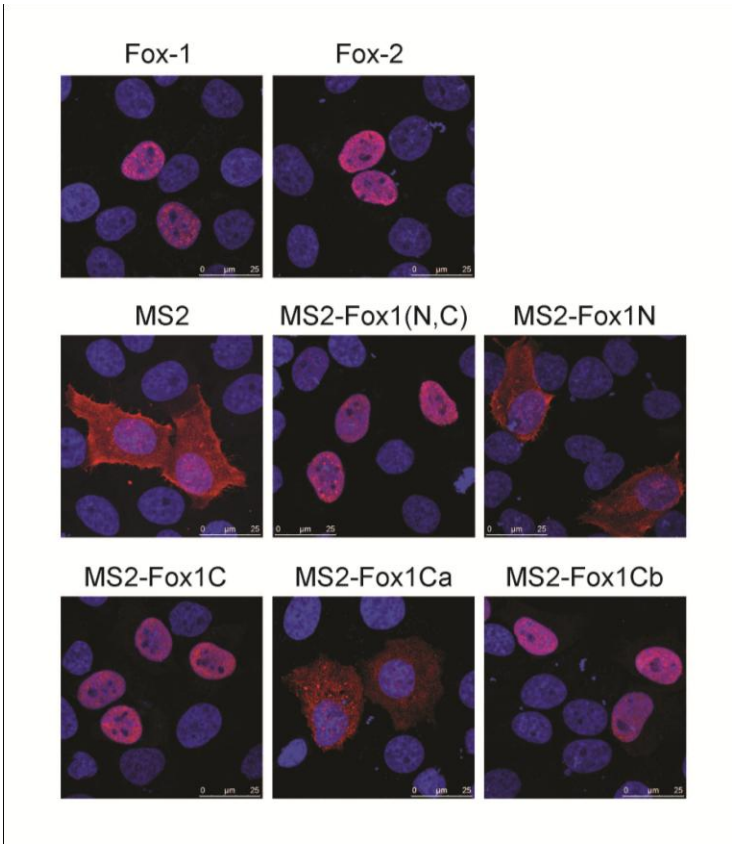

Figure 3.2 Subcellular localization of the Fox-1 mutant proteins. Indirect immunofluorescence of HeLa cells transfected with mutant proteins was performed using Flag antibody.

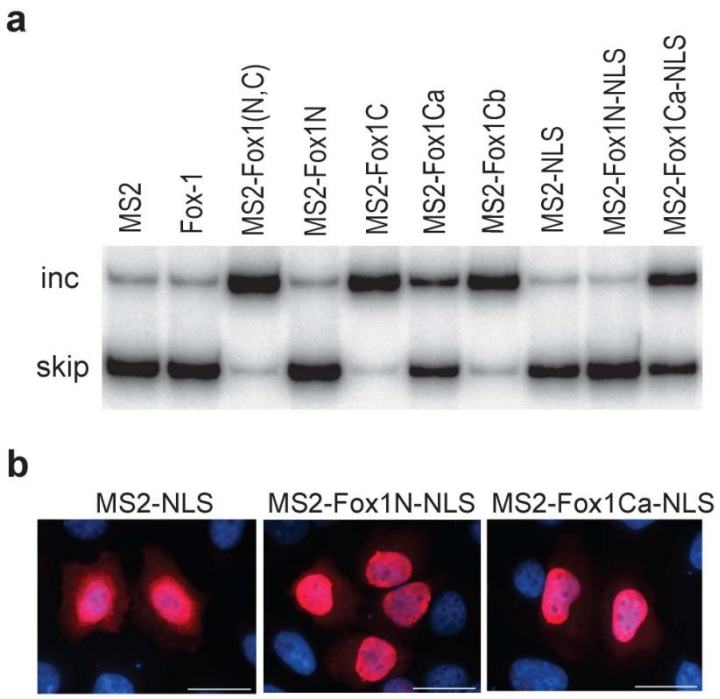

Figure 3.3 Mis-localization is not the reason for loss of function. (**a**) HeLa cells were co-transfected with the SMN2-MS2down reporter minigene and different mutant proteins. RNA was extracted 48 h after transfection and analyzed as in Figure 1B. (**b**) Indirect immunofluorescence of HeLa cells transfected with mutant proteins containing an SV40 NLS was performed using Flag antibody.

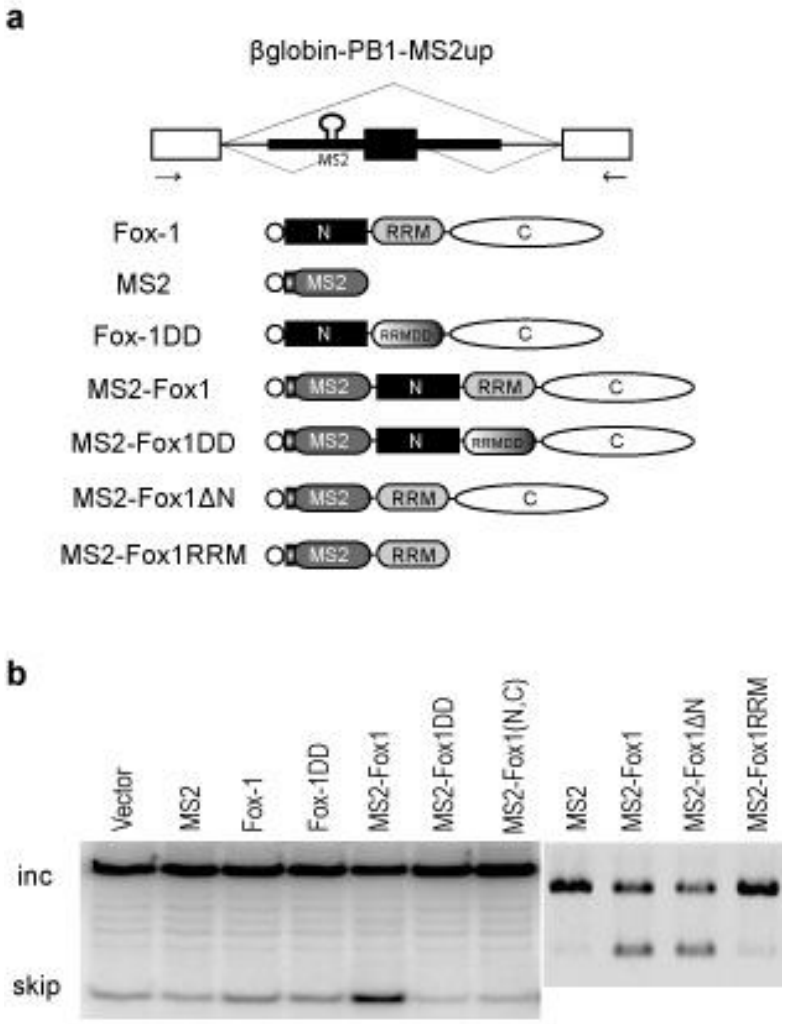

Figure 3.4 The RRM is required for exon repression when tethered to the upstream intron. (**a**) Diagrams of the modified βglobin-PB1 minigene and MS2-fused Fox-1 mutants. The three Fox-1 binding sites in the upstream intron and one binding site in the alternative exon were mutated, and an MS2-binding site was inserted in the upstream intron. The primers for RT-PCR are indicated by arrows below the exons. All the protein mutants have a Flag epitope tag, which is indicated by an open circle. The RRM with two phenylalanine to alanine mutations is designated RRMDD. (**b**) HeLa cells were co-transfected with the βglobin-PB1-MS2up reporter minigene and different mutant proteins. RNA was extracted 48 h after transfection. Radioactive RT-PCR was performed to detect the changes of the alternative splicing isoforms.

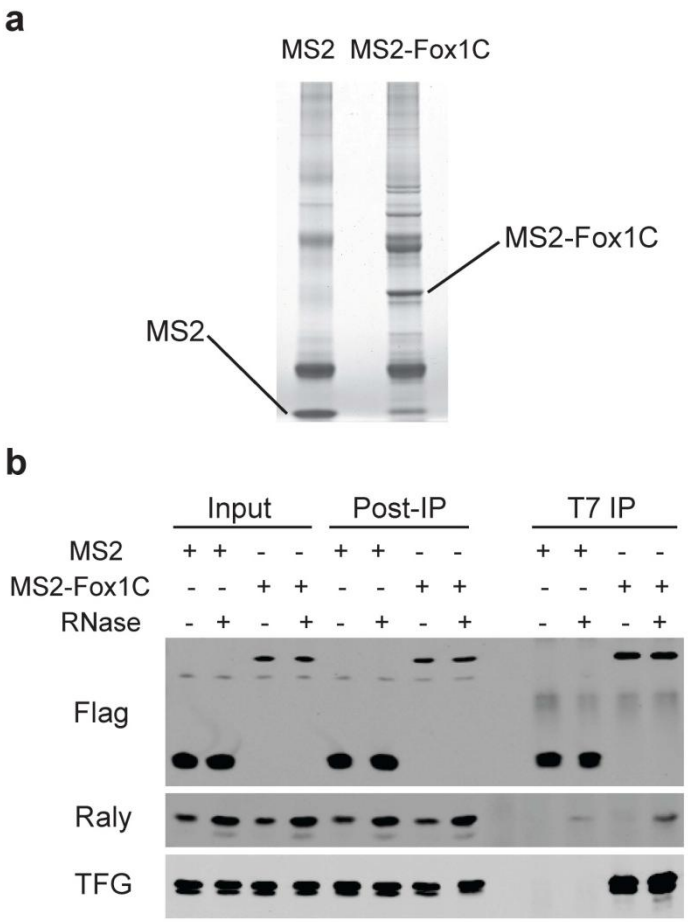

Figure 3.5 The C-terminal domain of Fox-1 interacts specifically with Raly and TFG. (**a**) HeLa cells were transfected with MS2 or MS2-Fox1C. The proteins were immunoprecipitated from whole-cell lysates using T7 monoclonal antibody coated Protein G Dynabeads. The immunoprecipitated proteins were separated on a gradient SDS-PAGE gel and stained with Coomassie Blue. (**b**) Co-immunoprecipitation using T7 antibody and western blotting with Flag, TFG and Raly antibodies. IPs were performed either with or without nuclease treatment. Whole-cell lysates, post-IP lysates and IPs were blotted with the indicated antibodies.

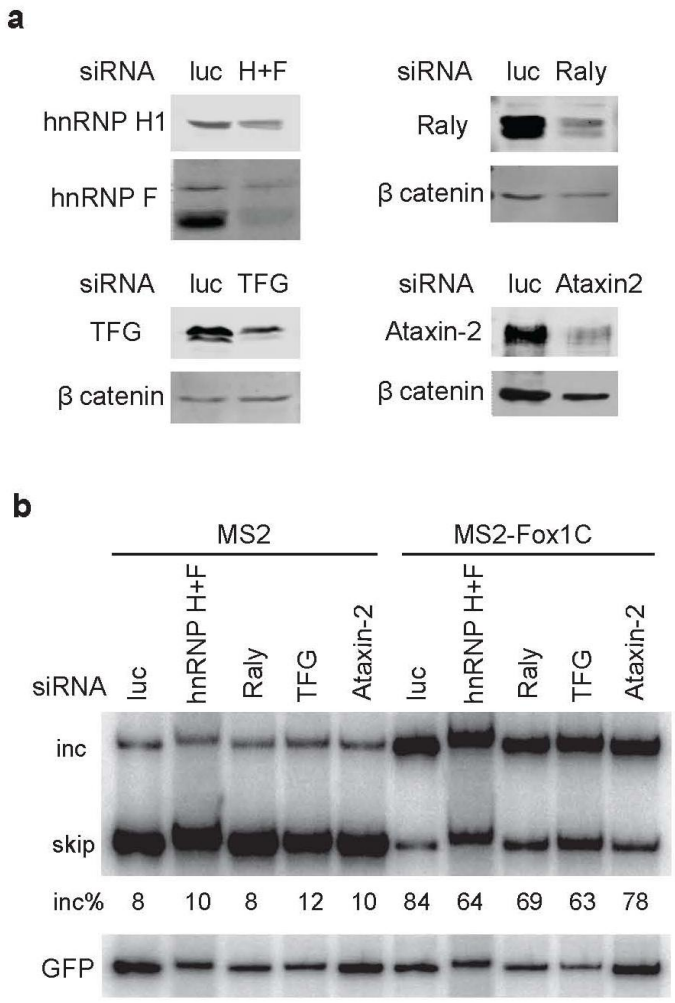

Figure 3.6 Knockdown of hnRNP H/F, Raly and TFG partially inhibits exon activation induced by MS2-Fox1C. HeLa cells were transfected with the indicated siRNAs 48 h before transfection of SMN2-MS2down minigene and MS2/MS2-Fox1C plasmids. GFP was co-transfected as an internal control. After another 24 h, total proteins and RNAs were extracted. (**a**) Western blotting analysis using the indicated antibodies. (**b**) Radioactive RT-PCR of *SMN2* and GFP. Samples were reloaded according to the GFP level.

Chapter 4

Global assessment of alternative splicing regulation by massively parallel paired-end mRNA sequencing

4.1 Introduction

Ultra-high-throughput mRNA-Seq [1-2] has emerged as a promising alternative to splicing microarrays [3-4] for AS studies by avoiding cross-hybridization issues and improving sensitivity. Published studies [1-2, 5-7] have mostly used SE-mRNA-seq, in which ~30 nucleotides at the 5′ end of each fragmented transcript are sequenced to obtain millions of reads. Due to the small read-length generated by current platforms (e.g., Illumina/Solexa), a limitation of the technology is the very small proportion (~5%) of reads that span exon junctions [2, 5], impeding the detection of splicing isoforms that are exclusively represented by exon junctions, such as the exon-skipping product in cassette-type AS,. A methodological upgrade of SE-mRNA-Seq that can mitigate this problem is PE-mRNA-Seq, which determines the sequences at both ends of each fragment, thereby providing additional structural information about the transcript. Here we address the unique challenges in the analysis of PE-mRNA-Seq data and present the first application of the technology to AS regulation. We identify the targets of Fox-2 (also known as RBM9), a highly conserved splicing factor enriched in brain, heart, and muscle, which has been implicated in several neuromuscular diseases [8-11].

4.2 Results

A PE read determines the two ends of each fragment, but transcript-structure ambiguity remains when AS occurs in the middle of a fragment. To help resolve this ambiguity, only fragments of a defined size are selected for sequencing (e.g., a ~250-nt gel band corresponding to a fragment size of 184 +/- 20 nt before ligation of adaptors). The size variation of selected fragments is much smaller than the median size of typical human exons (126 nt for constitutive exons and 108 nt for cassette exons). We developed an algorithm, named Bayesian Analysis of Splicing Isoform Structure (BASIS), which takes advantage of this stringent size constraint to infer the probability of all splicing isoforms compatible with each PE-read (**Fig. 4.1** and Methods). After alignment of the 5′ reads (read1) and 3′ reads (read2) to the genome or our exon-junction database, BASIS

first estimates the size distribution of all fragments using large exons (**Fig. 4.1a**), and a prior probability of each AS isoform from an AS database using directly observed junction reads. For each PE read, it then enumerates all compatible isoforms and infers the posterior probability of each isoform (**Fig. 4.1 b and c**). A PE read with at least one compatible transcript is considered a legitimate PE (L-PE) read. For reads without unique mapping for both ends, or without transcript support, the 5′ and 3′ reads are treated separately, and considered illegitimate SE (IL-SE) reads. The 5′ and 3′ reads before structural inference are also named SE reads. The L-PE reads weighted by inferred posterior probabilities with respect to compatible isoforms, and the IL-SE reads are then combined to estimate splicing changes for each AS event, similar to previous analyses of SE-mRNA-Seq data [2].

To evaluate the performance of BASIS, we first performed simulations. In each simulation experiment, we fragmented RefSeq transcripts *in silico* into a particular size to obtain 5 million "simulated PE reads" and controlled the variation of fragment sizes to mimic real PE-mRNA-Seq data. Among the ~4.6 million PE reads with both ends uniquely mapped to the genome, 97.5-99.8% were legitimate when the average fragment size was <184 nt (93.6% for fragment size 284 nt; **Fig. 4.2a**). Importantly, among the ~4.8 million reads including both L-PE and IL-SE reads, we observed a dramatic increase in the proportion of observed or inferred junction reads, from 0.7 million (14.6%) when the fragment size was 44 nt, to 2.4 million (49.2%) when the fragment size was 284 nt, strongly supporting the benefit of PE-mRNA-Seq to recover more exon junctions (**Fig. 4.2b**). To assess the accuracy of BASIS, we focused on AS-junction reads not directly observed but inferred by the model (i.e., PE reads that span an alternatively spliced junction). As shown in **Fig. 4.2c**, the number of inferred AS-junction reads increased from ~1,000 to 0.5 million, when the fragment size varied from 44 nt to 284 nt (left axis). The proportion of correctly inferred AS-junction reads ranged from 75.2 to 84.6% (for sizes 44-184 nt), with a moderate decrease for larger fragment sizes (right axis). These simulation results suggest that PE-mRNA-Seq can greatly increase the detection of alternatively spliced exon junctions, with high accuracy even at the level of individual reads.

We next performed PE-mRNA-Seq using the Illumina/Solexa platform to study regulated alternative splicing by Fox-2. Fox-2 belongs to a conserved family of splicing factors that recognize a well defined motif and have many targets identified in previous studies [10-17]. Because Fox-2 is the only paralog expressed in HeLa cells, we compared HeLa cells that express Fox-2 (dubbed the "Fox-2" sample) with shRNA-treated HeLa cells in which Fox-2 was knocked down ("No-Fox" sample) [10]. For each sample, we used cytoplasmic poly(A)+ RNA to generate sequencing libraries, and sequenced three lanes (technical replicates) to increase the read coverage. In total, we obtained 60 million and 48 million PE reads for the Fox-2 and No-Fox samples, respectively (**Table 4.1**). About 62-65% of the reads could be mapped unambiguously to the genome or our exon-junction database, including 3.9-4.2 million junction reads (6-7%) per sample. These and other general statistics are consistent with previous studies (**Fig. 4.3**) [2, 5]. Applying the BASIS method estimated the size of fragments to peak at 202-205 nt, with an unexpected minor peak at ~55 nt, which was consistent with Bioanalyzer analysis (**Fig. 4.4**). We also obtained a total of 97.1 million L-PE + IL-SE reads, including 13-16% junction reads, representing a two-fold gain by the BASIS analysis (**Table 4.1**). Further examination of the gene-expression level confirmed that Fox-2 was specifically knocked down by 3.2 fold (**Fig. 4.5**).

To compare PE-mRNA-Seq to SE-mRNA-Seq, we examined the number of exons and exon junctions recovered with different read-coverage thresholds. SE reads before BASIS analysis—equivalent to SE-mRNA-Seq with exactly the same sequencing depth—are ideal for such a comparison. As shown in **Fig. 4.6a**, the number of recovered exons did not discernibly change after BASIS analysis using different thresholds (left panel), or over the entire range of gene-expression levels (right panel). In contrast, BASIS analysis almost doubled the number of recovered exon junctions (**Fig. 4.6b**): 29-31% of exon junctions are supported by ≥ 10 fragments after BASIS, compared to 16-18% before BASIS (left panel). Again, this increase was consistently observed for genes with different expression levels (right panel).

Based on the greatly increased coverage of exon junctions, we expected that PE-mRNA-Seq would be more sensitive to detect splicing changes. Indeed, both technical replicates and PE sequencing increased the number of identified Fox-2-dependent exons: 292 exons showed a significant splicing change with three lanes per sample, compared to 42 or 140 exons with one or two lanes per sample, respectively (false discovery rate, or FDR < 0.1, Fisher's exact test, with Benjamini correction [18]) (**Fig. 4.6c**, left panel). In the comparison of PE- vs. SE-mRNA-Seq, BASIS analysis of single-lane-per-sample data had the greatest benefit in the FDR range between 0.05 and 0.2, for which the number of significant exons almost doubled (**Fig. 4.6c**, right panel). When each sample was sequenced with replicates, the gain of statistical power was shifted to the left (smaller FDR), because SE-mRNA-Seq data started to detect exons with moderately significant changes, while high-confidence predictions remained difficult. Taken together, these observations consistently confirm the higher sensitivity of PE-mRNA-Seq over SE-mRNA-Seq.

To estimate the sensitivity of PE-mRNA-Seq quantitatively, we examined 114 Fox-target exons that were previously validated by RT-PCR [10-17] and were included in our AS database. Among them, 44 exons (39%) showed a significant Fox-2-dependent splicing change in the PE-mRNA-Seq data (P<0.05, Fisher's exact test). Because these known Fox-2 targets were identified in various cell types, we considered the 32 exons we previously validated experimentally in HeLa cells [7], and found 14 (44%) with a significant change.

To evaluate the accuracy of PE-mRNA-Seq, we defined a high-confidence set of Fox-2 targets with FDR ≤ 0.1, in addition to a requirement for a reciprocal proportional change of exon inclusion $|\Delta I| \geq 0.1$ [19] (i.e., ≥10% change in exon inclusion level). With these stringent criteria, we identified 126 cases of cassette exons (95 activated exons and 31 repressed exons, **Table 4.2**), 17 cases of tandem cassette exons (multiple consecutive exons included or skipped, **Table 4.3**), and 4 cases of mutually exclusive exons (**Table 4.4**). We tested 19 cases representing the entire range of statistical significance by RT-PCR, and validated 18 cases; we also tested three additional cases with a smaller ΔI, and

all three were validated (**Fig. 4.7** and **Fig. 4.8**). Therefore, the accuracy of PE-mRNA-Seq is estimated to be ~95% (18/19 or 21/22), confirming our observations about consistent and specific changes in read number in the AS region, but not in constitutively spliced exons and exon junctions (**Fig. 4.7**).

Fox-2-dependent splicing does not necessarily imply direct regulation by binding to the target RNA. Nevertheless, Fox proteins very specifically recognize a UGCAUG element [9, 14] and an RNA functional map has been derived to predict target exon inclusion or skipping, depending on the position of the binding sites [10-11, 16]. Therefore, the presence of the sequence element in regions consistent with the map should be indicative of direct targets. Indeed, *de novo* motif analysis of cassette exons with Fox-2-dependent inclusion revealed a single, highly significant motif with a UGCAUG core in the downstream intronic sequences (**Fig. 4.9a**, E value$<1.6\times10^{-11}$) [20-21], suggesting that Fox-2 is the primary and direct regulator of these exons. More detailed analysis of the UGCAUG element revealed distinct patterns of motif enrichment consistent with the previous map: there was a 9-fold enrichment in intronic sequences downstream of the 5′ splice sites of Fox-2-activated exons, and 11- and 4- fold enrichment in exons and in intronic sequences upstream of the 3′ splice sites of Fox-2-repressed exons, respectively (**Fig. 4.9b**). Importantly, our data also suggested an extension of the map, by showing an enrichment of the motif in intronic sequences flanking the downstream constitutive exons—similar to observations made for the neuron-specific splicing factor Nova [22-23]. In total, 71 of the 126 Fox-dependent cassette exons have at least one UGCAUG element in the exon or in the flanking introns near the 5′ or 3′ splice sites, representing a 2.7-fold enrichment compared to all cassette exons as a control ($P<5.2\times10^{-18}$, Fisher's exact test). More stringently, 26 high-confidence Fox-2 dependent exons have conserved UGCAUG elements and were predicted as Fox-1/2 targets in our previous study [10], representing an 8-fold enrichment ($P=2.1\times10^{-16}$, Fisher's exact test). Combining these observations with the fact that some Fox targets have more distal binding sites [15], a majority of the exons identified by PE-mRNA-Seq represent direct Fox targets, and the conservation of the putative Fox binding sites further argues for their functional relevance under physiological conditions.

4.3 Discussion

To conclude, we present the first application of PE-mRNA-Seq to study AS regulation, and demonstrate its advantage over existing technologies. PE-mRNA-Seq was recently used to study gene expression [24], but analysis of AS using the technology was not addressed. We developed a simple but very effective statistical model to infer transcript structural information in PE data, which resulted in substantially improved sensitivity (~2 fold) and high accuracy (~95% validation rate). However, extended motif analysis also suggested that the sequencing depth in the current study did not reach saturation in Fox-2 target identification. This is probably due in part to our specific experimental system being based on HeLa cells, in which both Fox-2 protein and many targets are only moderately expressed (**Fig. 4.10**). Future investigation in brain and muscles will likely to yield more functional targets.

4.4 Methods

4.4.1 Sample preparation, PE mRNA-Seq and RT-PCR

HeLa cells expressing or lacking Fox-2 were generated as previously described [10]. To reduce splicing precursors and intermediates, cells were fractionated to enrich cytoplasmic RNA. Specifically, we lysed cells in gentle lysis buffer (10 mM HEPES pH 7.4, 10mM NaCl, 3 mM MgCl2, 0.5 % (v/v) NP-40), and pelleted the nuclei at 2300 g for 5 min. We extracted cytoplasmic RNA from the supernatant by Trizol and treated with DNase I (Promega). We used 4 μg cytoplasmic RNA from each sample to prepare PE-mRNA-Seq libraries following the manufacturer's instructions (Illumina, San Diego, CA). Briefly, poly (A)-enriched RNA was prepared using the oligo-dT mRNA purification kit (Invitrogen), and then subjected to fragmentation (Ambion). Double-stranded cDNA was synthesized and ligated to adaptor oligos at both ends. The samples were subjected to agarose gel electrophoresis and the cDNA was extracted from gel slices at the position of around 250bp-length (QIAGEN). Then the samples were amplified by PCR for 15 cycles and purified by QIAquick PCR spin column (QIAGEN). It was then

added to Illumina flow cell and cluster generation was carried on an Illumina cluster generation station. Flow cells were then subjected to paired end sequencing on an Illumina GA IIx sequencer with version 2 or 3 chemistry. Data was analyzed using the Illumina data analysis pipeline (v 1.0-1.4). Each of the two samples was sequenced in three Solexa lanes to increase the coverage. Semi-quantitative RT-PCR was performed as previously described [10], with primers given in **Table 4.5**.

4.4.2 Simulation of PE-mRNA-Seq

In each simulation, we chose an average fragment size (between 44 nt and 284 nt, adaptor not included). The size of each simulated fragment was then generated from a normal distribution with the specified mean, and a standard deviation of 12 nt (i.e., 95% of fragments are in the +/- 24 nt range). The start position of a fragment on a randomly picked RefSeq transcript was then generated from a uniform distribution.

4.4.3 Compilation of exons, exon junctions and AS events

We compiled human exons, exon junctions, and AS events based on our splicing database dbCASE [25], and Refseq and UCSC Known Gene transcripts [26]. To estimate gene-expression level, we used a nonredundant set of 330,729 exons comprising all AG/GT internal exons from dbCASE, and additional exons from all Refseq/UCSC Known Gene transcripts that are missing in dbCASE (e.g., terminal exons and reliable non-canonical exons). Similarly, 1,246,562 exon junctions were compiled from a combination of dbCASE, Refseq, and UCSC Known Gene transcripts.

4.4.4 Reads mapping

Raw 5′ and 3′ reads from Illumina/Solexa PE-mRNA-Seq were mapped independently to hg18 using size 22 nt to 32 nt by Eland (Illumina). We required unambiguous mapping of $\geq$ 22 nt with $\leq$ 2 mismatches. Reads were also mapped to exon junctions, requiring $\leq$ 2 mismatches in $\geq$22 nt, with $\geq$ 4 nt overlap on each side. The mapping to the genome and exon junctions was then combined to get the best mapped loci. Only reads unambiguously mapped to unique loci were included for further analysis.

4.4.5 Inference of transcript structures

The probabilistic model of gapless considers all PE reads with both ends mapped uniquely and located within exons, exon pairs, and trios in our splice isoform database, to infer the probability of all compatible isoforms. Since the average exon size is 150 nt, i.e., 450 nt for exon trios, and the typical fragment size is ~145-245 nt (corresponding to a 200-300 nt gel band with adapters included, see **Fig. 4.4a**), about 97% of fragments from known transcripts are covered in the database, as shown in **Fig. 4.2a**.

For each PE read, all unique isoforms (paths) P_k between the two ends are enumerated, and the mRNA transcript size in the path between the two ends, l_k, is recorded. The posterior probability of the read coming from path P_k is

$$\Pr(P_k \mid l) = \frac{\Pr(l \mid P_k)\Pr(P_k)}{\sum_j \Pr(l \mid P_j)\Pr(P_j)} = \frac{\Pr(l_k)\Pr(P_k)}{\sum_j \Pr(l_j)\Pr(P_j)}$$

(1)

where $\Pr(l_k)$ is the probability of observing a fragment with length l_k. This size distribution was estimated from exons ≥ 400 nt, without internal alternative splicing, so that there is no ambiguity of splicing variants for all PE reads with both ends in the exon. The threshold of exon size is a conservative choice, given a rough estimate of ~185 nt fragments, to avoid truncations of the distribution. $\Pr(P_k)$ is the prior distribution, which is in proportion to the relative abundance of each isoform. This was estimated using directly observed SE junction reads (with a pseudo count 1). The prior is especially important for usage of alternative splice sites that are close to each other, such that the size constraint does not give sufficient resolution. In other cases when splicing paths yield a different number of exons (e.g., cassette exons), the size constraint is generally sufficient for discrimination, and the prior does not appear to play a dominant role.

4.4.6 Quantification of gene expression

We defined a set of 225,967 “core” exons by i) removing exons with inclusion level <0.5 and ii) removing overlapping exons. Gene-expression level was estimated by counting

the number of uniquely mapped L-PE and IL-SE reads in the core exons, normalized by the total length of core exons in each gene and the total number of reads (RPKM [5]).

4.4.7 Evaluation of splicing changes

For each AS event, we considered the two isoforms including or excluding the alternative exon. The number of reads supporting the inclusion or exclusion isoform in two conditions, Fox-2 (A) and NoFox (B), is denoted as N_{IA}, N_{EA}, N_{IB} and N_{EB}, respectively. Fisher's exact test was used to evaluate the significance of the splicing change [2]. Only AS events with $N_{IA}+N_{EA}+ N_{IB} +N_{EB} \geq 20$ and $N_{IA}* N_{EB}+ N_{EA}*N_{IB} > 0$ were tested. The latter criterion ensures that it is possible to evaluate the odds ratio. The *P*-values obtained by Fisher's exact test were corrected using the Benjamini-Hochberg procedure to obtain a false discovery rate (FDR).

Besides the statistical significance of a splicing change, we also estimated the magnitude of the reciprocal proportional change of splicing, ΔI, similar to the ASPIRE method [23, 27]. We denote the ratio of gene-expression level in the two conditions by $r_G=N_{GA}/N_{GB}$, and the relative abundance of the two isoforms as I_A, E_A, I_B, E_B, respectively. By definition,

$$\begin{cases} I_A + E_A = 1 \\ I_B + E_B = 1 \end{cases}$$

(2)

and

$$\begin{cases} I_A / I_B = R_I = \dfrac{N_{IA}/N_{GA}}{N_{IB}/N_{GB}} = \dfrac{1}{r_G}\dfrac{N_{IA}}{N_{IB}} \\ E_A / E_B = R_E = \dfrac{N_{EA}/N_{GA}}{N_{EB}/N_{GB}} = \dfrac{1}{r_G}\dfrac{N_{EA}}{N_{EB}} \end{cases}$$

(3)

After simple manipulation,

$$\begin{cases} I_A = \dfrac{R_I(1-R_E)}{R_I - R_E} = \dfrac{N_{IA}\left(N_{EB} - \dfrac{N_{EA}}{r_G}\right)}{N_{IA}N_{EB} - N_{EA}N_{IB}} \\ I_B = \dfrac{1-R_E}{R_I - R_E} = \dfrac{N_{IB}r_G\left(N_{EB} - \dfrac{N_{EA}}{r_G}\right)}{N_{IA}N_{EB} - N_{EA}N_{IB}} \end{cases}$$

(4)

and

$$\Delta I = \frac{R_I + R_E - 1 - R_I R_E)}{R_I - R_E} = \frac{\left(N_{IA} - r_G N_{IB}\right)\left(N_{EB} - \dfrac{N_{EA}}{r_G}\right)}{N_{IA}N_{EB} - N_{EA}N_{IB}}$$

(5)

As a constraint of consistency, we also require
$(R_I - 1)(R_E - 1) = (N_{IA} - r_G N_{IB})(N_{EA} - r_G N_{EB}) < 0$.

We applied this method to cassette exons, tandem cassette exons, and mutually exclusive exons. Since the letter two types involve multiple alternative exons in each event, some of the identified cases might be actually cassette exons that dominate the observed splicing change. Therefore, we filtered tandem cassette exons by removing those for which a more significant *P*-value was observed for the corresponding cassette-type AS event (if it exists). For mutually exclusive exons, we tested the extent of mutually exclusive splicing based on mRNA/EST sequences using Fisher's exact test (by building a contingency table counting transcripts supporting #both included, #exon1 only, #exon2 only, #both skipped). Candidate AS events were removed unless a *P*-value <0.05 was reached.

4.4.8 Motif analysis

Motif analysis was performed for cassette exons, using sequences from the alternative exon, and 200-nt sequences flanking the splice sites of upstream and downstream introns. Motif conservation was measured by a branch-length score (BLS) as previously described [10].

4.5 Acknowledgements

This work was supported by NIH grant GM74688 to M.Q.Z. and A.R.K., and in part by a grant from National Science Foundation of China (60702002) to C.X.. The PE-mRNA-Seq data will be deposited to the GEO short read achieve.

4.6 Figures and Figure Legends

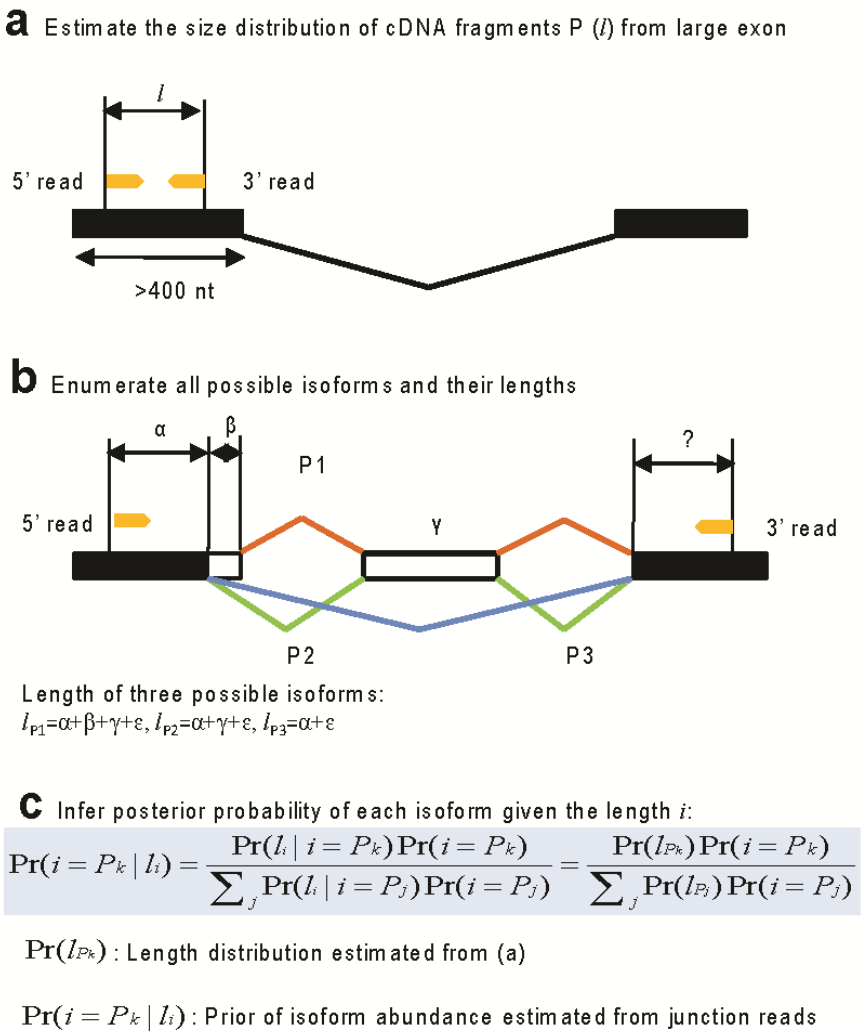

Figure 4.1 Schematic representation of the BASIS model to infer the transcript structure defined by paired-end reads. (**a**) The size distribution of cDNA fragments is estimated from exons ≥ 400 nt without internal splicing. (**b**) For each PE read with both ends uniquely mapped, all possible paths (isoforms) between the two ends are enumerated according to the AS database. (**c**) The posterior probability of each isoform is inferred, using the size of each isoform (**b**) and a prior probability of the abundance of each isoform estimated from directly observed junction reads.

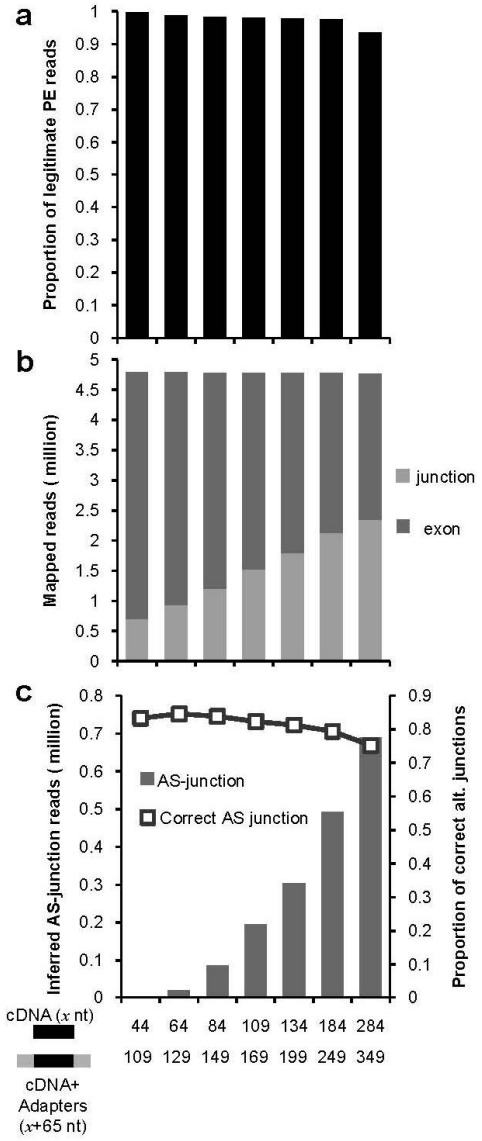

Figure 4.2 Evaluating the gapless method using simulation data. Each simulation differs in the average fragment size, while other parameters are the same. (**a**) The proportions of legitimate PE reads among all PE reads with both ends mapped uniquely. (**b**) The number of exon and exon-junction reads obtained in each simulation. (**c**) The number of reads that are inferred to span alternatively spliced exon junctions (left axis). The inferred structure was compared with the actual coordinates to calculate the correct rate (right axis).

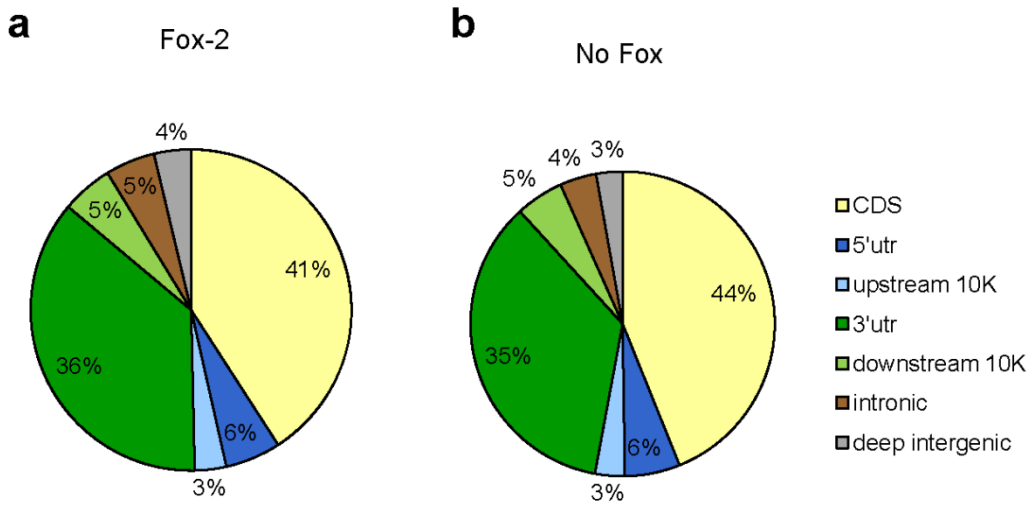

Figure 4.3 Distribution of SE reads over the genome. (**a**) Fox-2. (**b**) No-Fox. Genic regions, defined by RefSeq and UCSC Known Gene transcripts, are broken down into 5′ untranslated regions (UTRs), coding sequences (CDS), 3′ UTRs, and introns. The reads located within 10 kb upstream or downstream of genes are counted separately.

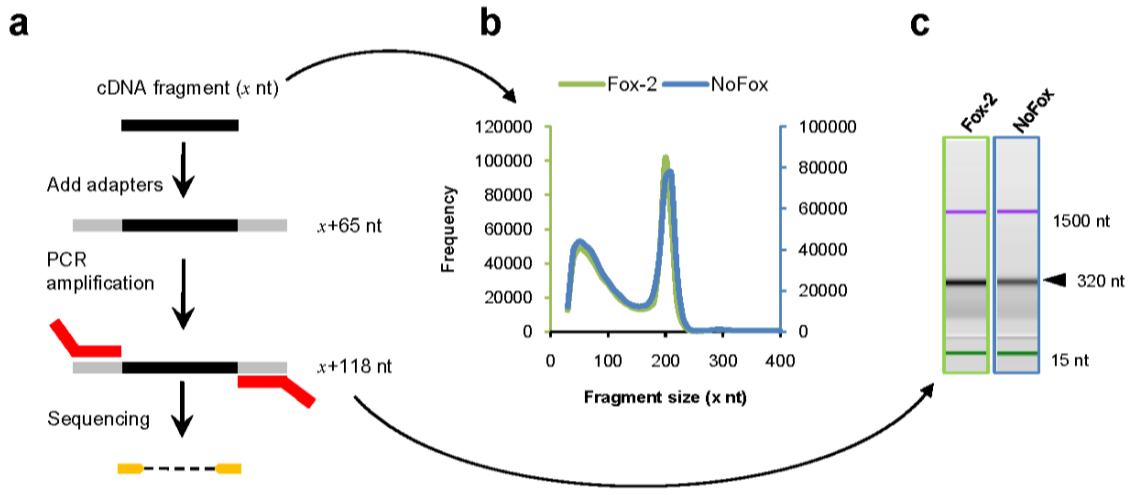

Figure 4.4 Transcript fragment-size estimation using real data. (**a**) A schematic representation of the simplified PE mRNA-Seq library preparation, highlighting the change in the size of cDNA species in the library due to adapters (gray) and sequencing primers (red) added to transcript fragments (black). (**b**) The size of transcript fragments estimated from PE reads located in exons ≥400 nt, for the Fox-2 (left axis) and No-Fox (right axis) samples. (**c**) Results of Bioanalyzer analysis using the final library. The peak size of transcript fragments, including adaptors and primers, is indicated. This peak corresponds to the size estimated in (**b**).

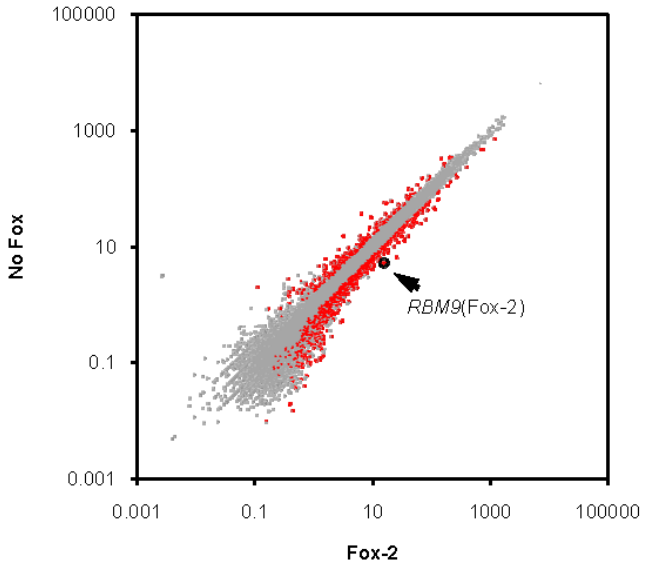

Figure 4.5 Gene expression level in the Fox-2 (x-axis) and No-Fox (y-axis) samples, estimated from the PE mRNA-seq data. Among all genes (gray points), 853 genes showing significant changes between the two samples ($P<0.05$, Binomial test, with Bonferroni correction, log2 fold changes >1.5) are highlighted in red. Fox-2, which is 3.2-fold down-regulated in shRNA treated cells, is indicated by an arrowhead.

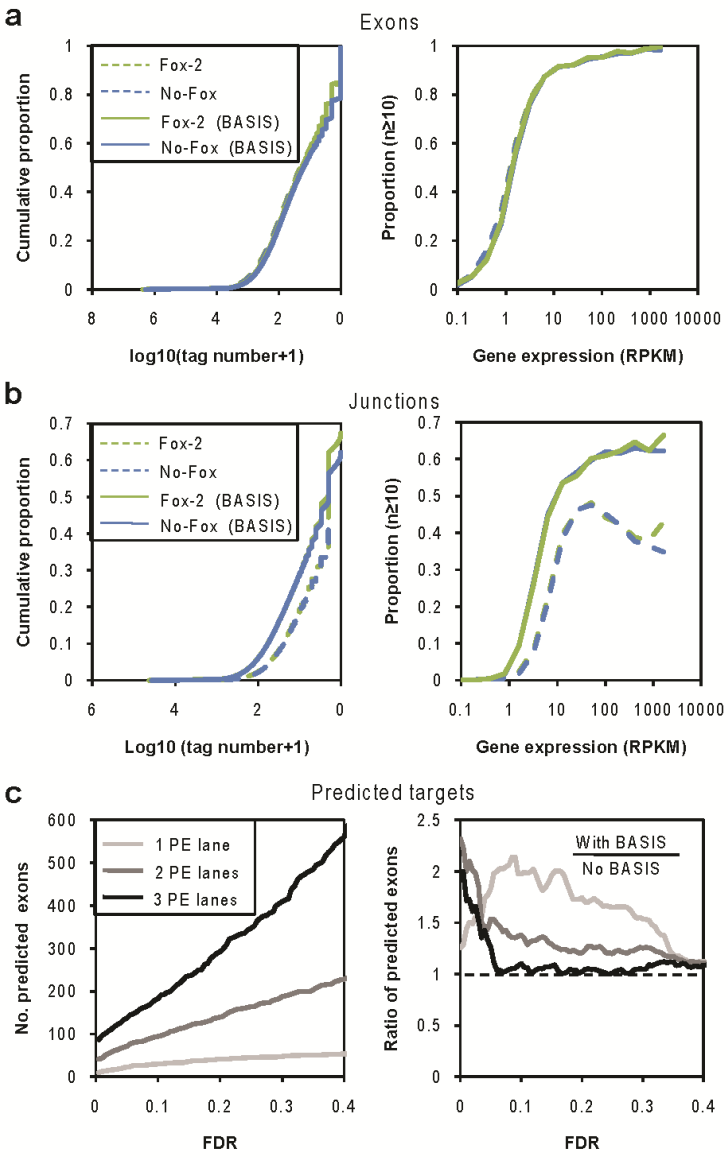

Figure 4.6 Comparison of PE-mRNA-seq and SE-mRNA-Seq for the detection of exons, exon junctions, and regulated target exons. (**a**) Detected exons in each sample. For each exon, the number of reads was counted to calculate the cumulative proportion of detected exon (y-axis) with varying thresholds (x-axis) (left panel). Exons are also binned according to gene-expression level (measured by RPKM [14]) and for each bin the proportion of detected exons with ≥ 10 reads was calculated (right panel). (**b**) similar to (**a**), but the detected exon junctions are shown. (**c**) Predicted target exons with Fox-2 dependent splicing. Left panel: the number of significant exons detected after structure inference, using one-lane, two-lane, and three-lane data, respectively. Right panel: The ratio of the number of significant exons detected after structure inference to that detected before structure inference, using one-lane, two-lane, and three-lane data, respectively.

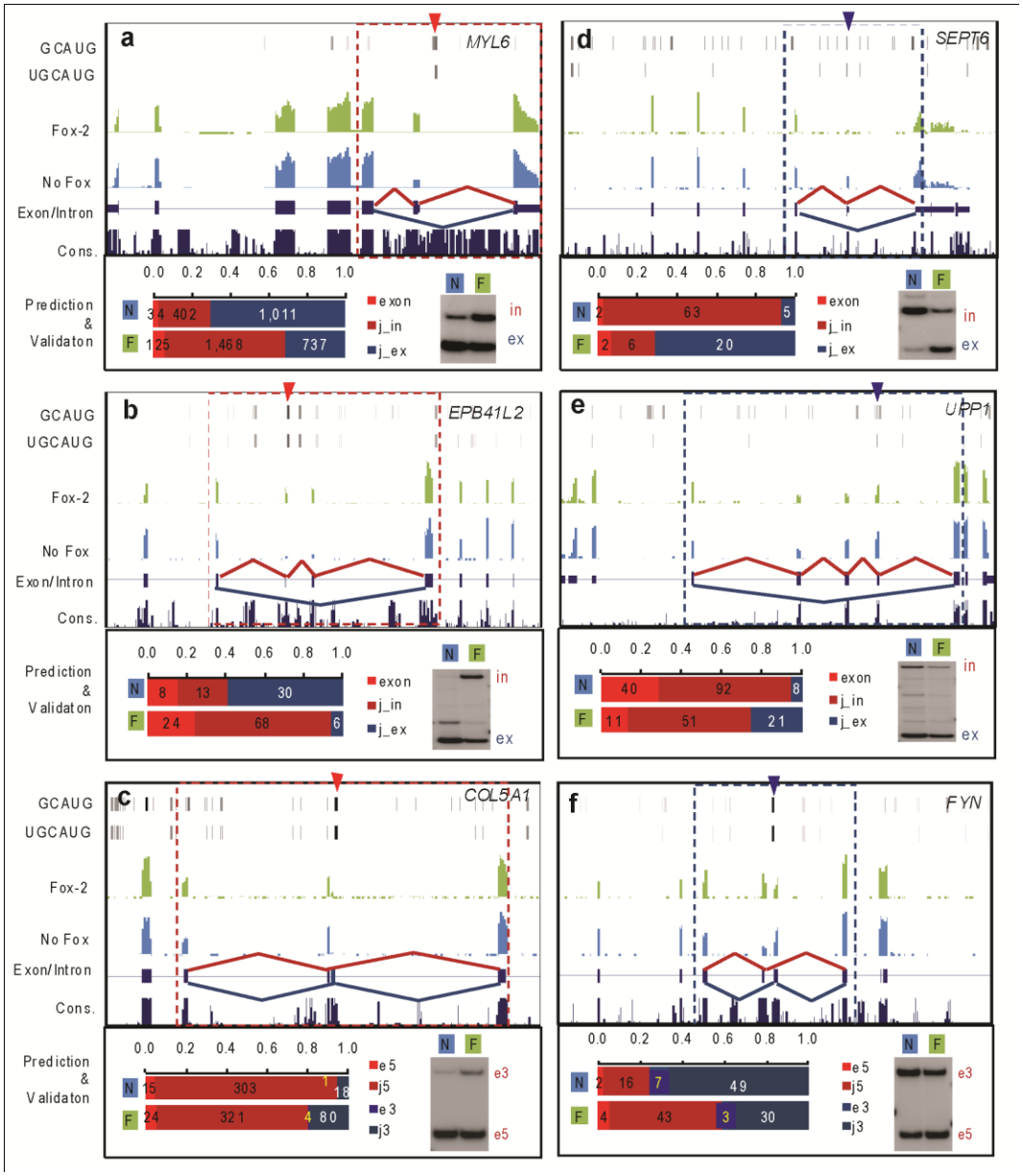

Figure 4.7 Examples of Fox-2 target exons identified by PE-mRNA-Seq and validated by RT-PCR Each panel shows an exon with predicted Fox-2 dependent exon activation (**a-c**) or repression (**d-f**). The Locations of GCAUG and UGCAUG elements, followed by read-coverage profiles of Fox-2 and No-Fox samples, exon and intron structure, and sequence conservation in vertebrates are shown at the top. The Fox-2 binding site(s) predicted to be important for splicing regulation are indicated by an arrowhead. The AS pattern is shown in the region highlighted by a box with dotted lines. At the bottom, the numbers of exon tags and exon-junction tags are shown for each sample. The results of RT-PCR are shown on the right. j_in: exon junction(s) of the inclusion isoform. j_ex: exon junction(s) of the exclusion isoform. j5 and e5: 5′ junction and exon, j3 and e3: 3′ junction and exon.

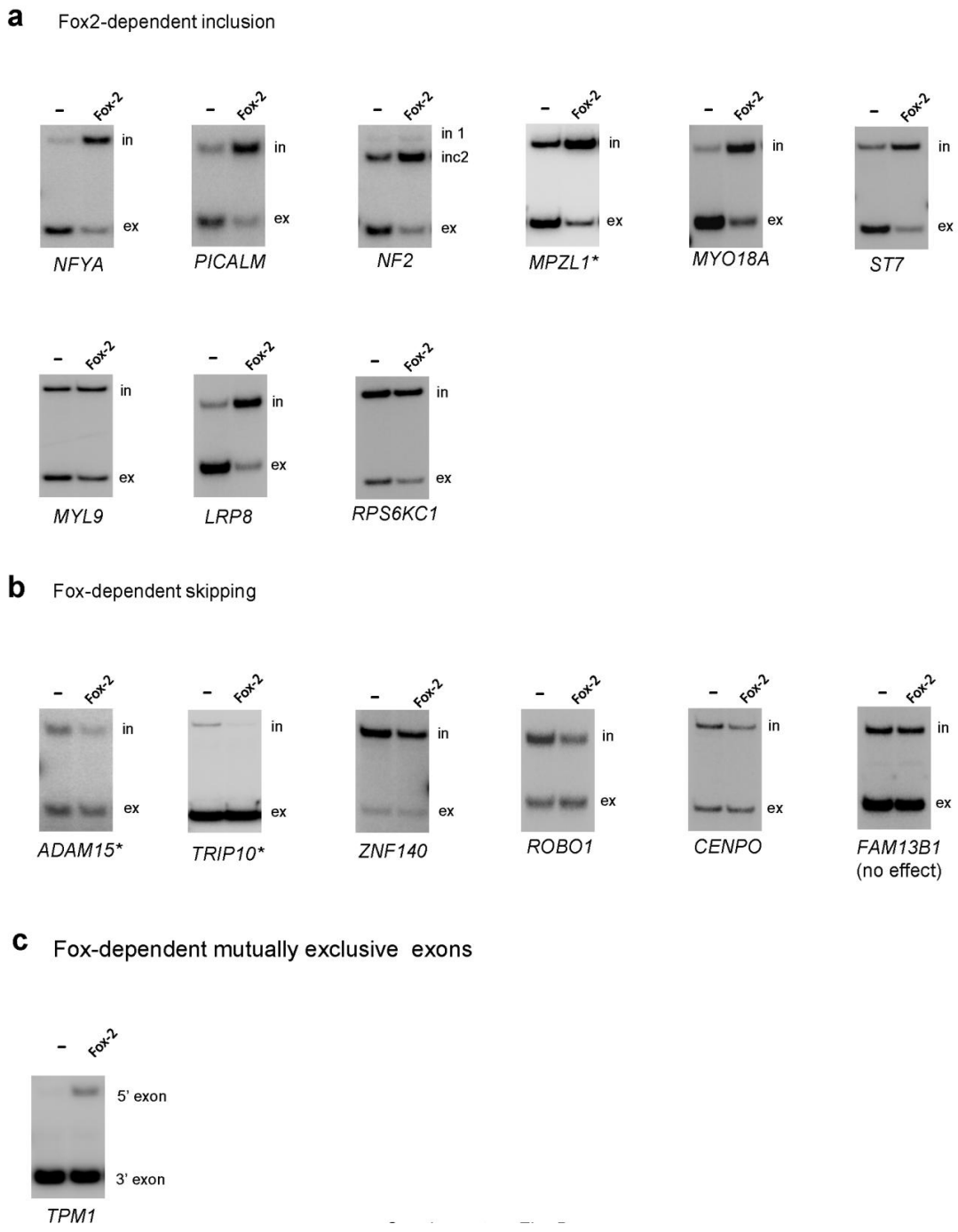

Figure 4.8 Additional examples of Fox-2 targets identified from PE mRNA-Seq data (**a**) Cassette exons with Fox-2 dependent exon inclusion. (**b**) Cassette exons with Fox-2 dependent exon exclusion. (**c**) Mutually exclusive exons. In each panel, the bands corresponding to the expected isoforms are indicated.

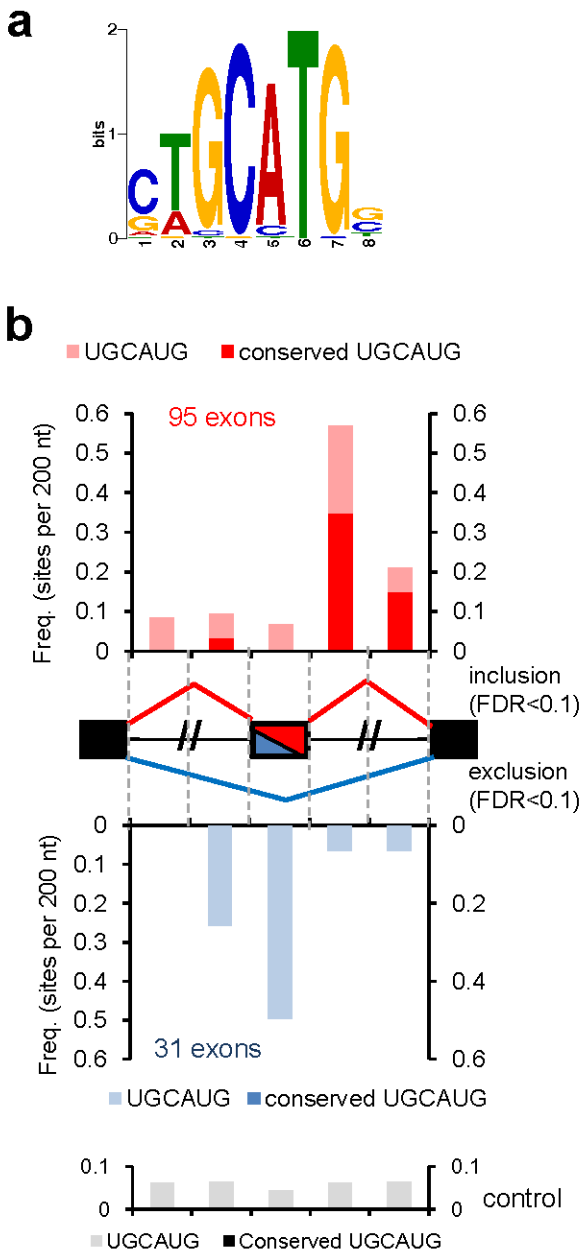

Figure 4.9 Analysis of Fox targets recovered the motif de novo and extended the RNA map. (**a**) *De novo* motif analysis using the intronic sequences downstream of the alternative exon activated by Fox-2. (**b**) An RNA map that predicts exon inclusion or skipping depending on the position of the Fox-2 binding sites. The number of all UGCAUG elements or conserved UGCAUG elements is shown, for exons activated (red) or repressed (blue) by Fox-2, respectively. The frequency of the motif in all cassette exons is shown at the bottom as a control.

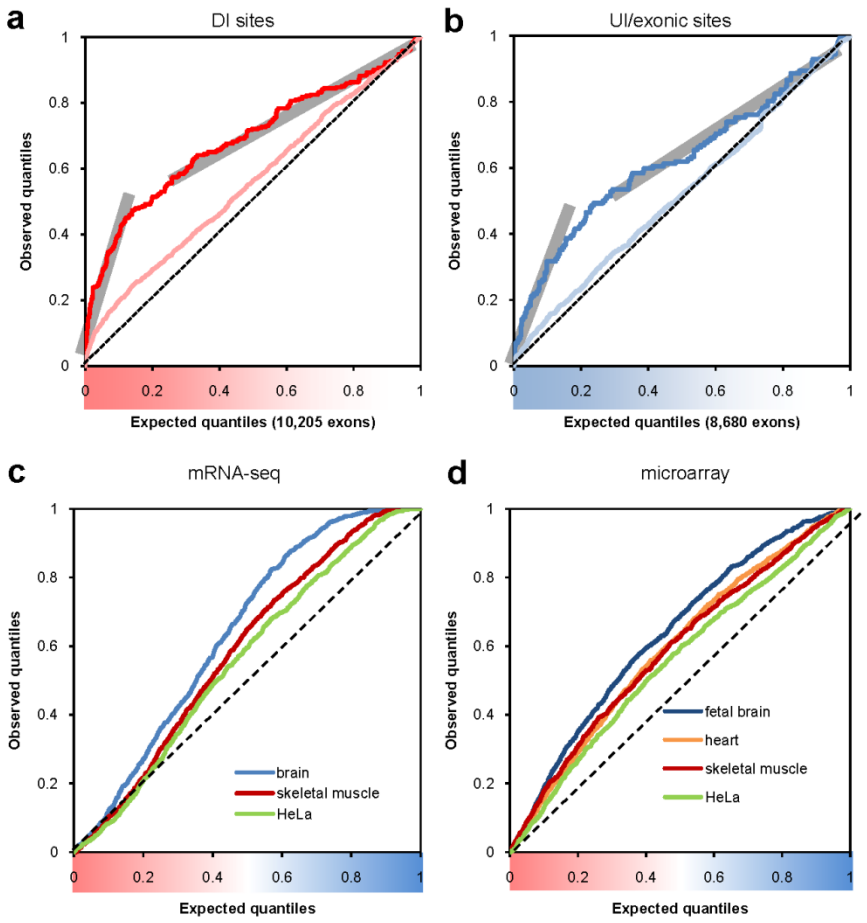

Figure 4.10 The detection of Fox target exons is not saturated at the current sequencing depth. (**a**) The cumulative distribution of activating UGCAUG elements, i.e., sites in downstream introns (DI), for exons ranked by the significance of Fox-2 dependent activation. The light-red curve represents the distribution of all UGCAUG elements, and the red curve represents the conserved UGCAUG elements. (**b**) Similar to (**a**), except that repressive UGCAUG elements (i.e., sites in exons and upstream introns, or UI), are used and exons are ranked by the significance of Fox-2-dependent repression. The light-blue curve represents the distribution of all UGCAUG elements, and the blue curve represents the conserved UGCAUG elements. (**c**) Fox target genes show higher relative expression in brain and skeletal muscle than in HeLa cells. Each curve represents a tissue or cell line profiled by mRNA-Seq. All genes are sorted by expression levels in descending order (x-axis). The cumulative distribution is plotted on the y-axis. A shift to the top left from the diagonal represents a higher relative expression (ranks) for the Fox target genes. The HeLa sample was generated in this study, and the brain and skeletal muscle data are from ref. [14]. (**d**) Similar to (**c**), but gene-expression level is estimated from whole-transcript microarray data [12].

4.7 Tables

Table 4.1: Summary of PE-mRNA-Seq reads in HeLa cells

Sample	Total PE reads	SE reads before BASIS analysis			L-PE+IL-SE Reads after BASIS analysis		
		all	genome	junction	all	Non-junction	junction
Fox-2	60,025,819	74,313,065	70,102,868	4,210,197	53,882,182	46,724,607	7,157,575
No-Fox	48,194,727	62,463,908	58,565,260	3,898,648	43,207,319	36,364,886	6,842,433

Table 4.2: Fox-2-dependent cassette exons

Symbol	chrom	chromosome start	chromsome end	strand	exon size	direction of change	FDR	UGCAUG
MYL6	chr12	54840294	54841624	+	45	1	3.52E-114	1
S100A4	chr1	1.52E+08	1.52E+08	-	49	1	7.79E-89	2
S100A4	chr1	1.52E+08	1.52E+08	-	39	1	2.02E-54	2
CLSTN1	chr1	9718530	9723900	-	57	1	4.36E-38	0
GOLIM4	chr3	1.69E+08	1.69E+08	-	84	1	1.25E-16	1
FOXM1	chr12	2844110	2845947	-	45	1	9.32E-16	0
PICALM	chr11	85365314	85369918	-	24	1	1.63E-14	0
LIMCH1	chr4	41315962	41341399	+	36	1	1.22E-11	1
MPRIP	chr17	17024042	17029564	+	63	1	1.37E-11	1
MYL6	chr12	54840294	54841624	+	35	1	2.77E-11	1
EPB41L1	chr20	34249195	34263711	+	411	1	9.02E-11	1
TPM2	chr9	35674729	35675335	-	76	1	2.61E-10	0
ST7	chr7	1.17E+08	1.17E+08	+	69	1	1.04E-09	0
MYO18A	chr17	24433460	24437720	-	45	1	2.88E-09	0
UAP1	chr1	1.61E+08	1.61E+08	+	51	1	4.90E-09	1
NUMA1	chr11	71399480	71404953	-	42	1	1.29E-08	1
UAP1	chr1	1.61E+08	1.61E+08	+	48	1	2.21E-08	1
SEPT6	chrX	1.19E+08	1.19E+08	-	45	-1	1.21E-07	1
SEPT6	chrX	1.19E+08	1.19E+08	-	45	-1	1.96E-07	1
MAP3K7	chr6	91302777	91313826	-	81	-1	3.04E-07	2
SPNS1	chr16	28900292	28901366	+	156	1	3.25E-07	0
NFYA	chr6	41154746	41159908	+	87	1	3.25E-06	1

NF2	chr22	28407428	28424591	+	45	1	5.51E-06	1
APP	chr21	26276528	26294367	-	57	1	1.15E-05	1
NASP	chr1	45844741	45851506	+	1017	1	1.93E-05	0
MYL9	chr20	34606676	34611065	+	162	1	2.08E-05	1
NSFL1C	chr20	1383612	1386918	-	157	1	2.12E-05	1
SPAG9	chr17	46407131	46409580	-	39	1	2.23E-05	0
MAP2K7	chr19	7874839	7880780	+	48	1	3.94E-05	0
NME4	chr16	387209	389123	+	187	1	6.66E-05	0
UBR4	chr1	19323622	19325735	-	97	1	8.55E-05	0
SEPT6	chrX	1.19E+08	1.19E+08	-	62	-1	1.23E-04	1
PAM	chr5	1.02E+08	1.02E+08	+	321	1	1.60E-04	0
C16orf13	chr16	625282	626278	-	257	1	1.69E-04	1
TPM1	chr15	61140450	61141528	+	76	1	1.74E-04	0
FMNL1	chr17	40667222	40669393	+	18	1	1.72E-04	1
ZNF140	chr12	1.32E+08	1.32E+08	+	57	-1	2.29E-04	0
ENAH	chr1	2.24E+08	2.24E+08	-	63	1	2.29E-04	0
RAC1	chr7	6398080	6406343	+	57	1	2.27E-04	0
NDEL1	chr17	8304052	8312195	+	35	1	2.90E-04	1
KIAA1128	chr10	86227311	86265158	+	85	1	2.98E-04	1
MYL9	chr20	34606676	34611065	+	67	1	3.13E-04	1
PFDN5	chr12	51975593	51977974	+	437	1	3.36E-04	1
NFYA	chr6	41154746	41159908	+	84	1	4.02E-04	1
LRP8	chr1	53509487	53514012	-	39	1	4.09E-04	1
MARK3	chr14	1.03E+08	1.03E+08	+	27	1	4.12E-04	0
NME4	chr16	387209	389123	+	178	1	4.60E-04	0

C16orf13	chr16	625282	626278	-	163	1	4.57E-04	1
R3HDM1	chr2	1.36E+08	1.36E+08	+	69	1	5.21E-04	1
MBNL1	chr3	1.54E+08	1.54E+08	+	36	-1	6.32E-04	2
MYO9B	chr19	17181395	17182660	+	48	1	9.61E-04	1
ABI1	chr10	27099180	27106175	-	15	1	9.99E-04	0
SQSTM1	chr5	1.79E+08	1.79E+08	+	157	1	1.26E-03	0
TARBP2	chr12	52184749	52185900	+	128	1	1.30E-03	1
UPP1	chr7	48100885	48113203	+	47	-1	1.74E-03	0
DIAPH1	chr5	1.41E+08	1.41E+08	-	27	1	2.02E-03	1
NF2	chr22	28407428	28424591	+	60	1	2.15E-03	1
BAT2D1	chr1	1.7E+08	1.7E+08	+	103	-1	2.27E-03	0
RAB11FIP3	chr16	478852	486859	+	135	1	2.93E-03	1
FOXM1	chr12	2844110	2845947	-	45	1	3.30E-03	0
ETF1	chr5	1.38E+08	1.38E+08	-	124	1	3.34E-03	1
LTBP3	chr11	65063767	65064428	-	141	1	3.54E-03	0
R3HDM1	chr2	1.36E+08	1.36E+08	+	255	1	3.89E-03	1
C16orf68	chr16	8623081	8627155	+	266	1	3.96E-03	0
C16orf68	chr16	8623081	8627155	+	139	1	4.79E-03	0
UPP1	chr7	48100885	48113203	+	159	-1	5.68E-03	0
HDHD2	chr18	42916708	42930868	-	62	-1	5.87E-03	0
JARID1C	chrX	53266767	53271113	-	43	1	6.22E-03	1
EVI5L	chr19	7826901	7829216	+	33	1	6.99E-03	1
ANXA11	chr10	81922543	81925890	-	178	1	8.91E-03	0
FMNL3	chr12	48317991	48327058	-	115	1	1.02E-02	1
PAFAH1B1	chr17	2443991	2488363	+	205	1	1.07E-02	0

GTF3C2	chr2	27419679	27433134	-	208	1	1.25E-02	0
CPNE1	chr20	33684131	33716213	-	143	1	1.35E-02	0
MRPL3	chr3	1.33E+08	1.33E+08	-	81	-1	1.47E-02	0
CENPO	chr2	24869837	24876216	+	114	-1	1.65E-02	0
PBX1	chr1	1.63E+08	1.63E+08	+	113	-1	1.73E-02	1
ODF2	chr9	1.3E+08	1.3E+08	+	240	1	1.96E-02	0
ODF2	chr9	1.3E+08	1.3E+08	+	235	1	1.95E-02	0
CAST	chr5	96084099	96088989	+	66	1	2.08E-02	0
FHL2	chr2	1.05E+08	1.05E+08	-	51	-1	2.46E-02	0
ABI1	chr10	27080533	27094252	-	87	1	2.55E-02	1
PAK4	chr19	44308267	44352236	+	73	-1	2.67E-02	1
ATP13A3	chr3	1.96E+08	1.96E+08	-	90	1	2.82E-02	0
CCDC99	chr5	1.69E+08	1.69E+08	+	77	1	2.91E-02	1
DEPDC1	chr1	68719711	68722360	-	852	-1	3.09E-02	3
RPAIN	chr17	5266813	5276738	+	112	1	3.49E-02	0
NBEAL2	chr3	47008334	47010513	+	81	1	3.50E-02	0
LSM14B	chr20	60134755	60138402	+	117	1	3.62E-02	1
C16orf13	chr16	624720	626278	-	257	1	3.84E-02	1
TLE3	chr15	68145407	68153999	-	322	1	4.02E-02	2
APLP2	chr11	1.29E+08	1.3E+08	+	168	-1	4.27E-02	0
PAK4	chr19	44308267	44352236	+	83	-1	4.36E-02	1
DNAJC5	chr20	62032648	62035840	+	74	-1	4.43E-02	0
MORF4L2	chrX	1.03E+08	1.03E+08	-	49	-1	4.40E-02	1
FAM13B1	chr5	1.37E+08	1.37E+08	-	66	-1	4.86E-02	0
CD46	chr1	2.06E+08	2.06E+08	+	93	1	4.87E-02	0

HPS4	chr22	25191421	25192227	-	54	1	5.21E-02	1
DEGS1	chr1	2.22E+08	2.22E+08	+	65	1	5.26E-02	0
SMG7	chr1	1.82E+08	1.82E+08	+	150	1	5.31E-02	0
C16orf13	chr16	624720	626278	-	163	1	5.36E-02	1
RAD51L3	chr17	30457518	30458578	-	135	1	5.40E-02	0
C20orf7	chr20	13737488	13745191	+	98	1	5.97E-02	1
CBLB	chr3	1.07E+08	1.07E+08	-	144	1	6.05E-02	0
MON1A	chr3	49922556	49925795	-	486	-1	6.03E-02	0
IHPK2	chr3	48706206	48707857	-	131	1	6.31E-02	0
FYN	chr6	1.12E+08	1.12E+08	-	165	1	6.60E-02	0
JTB	chr1	1.52E+08	1.52E+08	-	83	-1	6.94E-02	0
GOLGA4	chr3	37371596	37382678	+	63	-1	7.12E-02	0
C16orf68	chr16	8630088	8636669	+	41	-1	7.08E-02	0
EED	chr11	85652862	85665827	+	134	-1	7.23E-02	0
PLOD2	chr3	1.47E+08	1.47E+08	-	63	1	7.47E-02	1
G6PC3	chr17	39507574	39508344	+	119	-1	7.53E-02	0
TSPAN4	chr11	832846	840366	+	98	-1	7.93E-02	0
NT5C3	chr7	33032954	33068930	-	55	1	8.06E-02	0
ROBO1	chr3	78777988	78783768	-	27	-1	8.01E-02	0
CTSC	chr11	87685204	87707897	-	85	1	8.04E-02	0
CAST	chr5	96084099	96088989	+	106	1	8.63E-02	0
SEC31A	chr4	83997866	84003568	-	78	-1	8.65E-02	0
ALAS1	chr3	52207188	52208495	+	177	1	8.87E-02	0
IHPK2	chr3	48706206	48707857	-	114	1	9.24E-02	0
FAM109A	chr12	1.1E+08	1.1E+08	-	394	1	9.21E-02	0

PPIL5	chr14	49138838	49144588	+	67	1	9.17E-02	0
NFRKB	chr11	1.29E+08	1.29E+08	-	80	-1	9.27E-02	0
SH3KBP1	chrX	19469964	19478121	-	129	1	9.42E-02	1
RPS6KC1	chr1	2.11E+08	2.11E+08	+	93	1	9.99E-02	1

Table 4.3: Fox-2-dependent tandem cassette exons

Symbol	chrom	chromosome start	chromsome end	tandem exon number	strand	direction of change	FDR
EPB41L2	chr6	131232396	131248100	2	-	1	4.61E-09
KIAA0528	chr12	22501172	22513995	2	-	1	6.27E-04
NRD1	chr1	52074391	52078553	2	-	-1	1.89E-03
UTRN	chr6	145160653	145165975	2	+	1	3.39E-03
KIAA0528	chr12	22501172	22513995	2	-	1	5.65E-03
UPP1	chr7	48100885	48113203	3	+	-1	1.78E-02
MARK3	chr14	103027867	103039370	2	+	1	1.88E-02
ESAM	chr11	124129747	124137395	2	-	1	2.06E-02
ESAM	chr11	124129747	124137395	2	-	1	2.40E-02
LSM14B	chr20	60134755	60138402	2	+	1	2.80E-02
UBE2B	chr5	133737973	133752002	2	+	-1	4.28E-02
UPP1	chr7	48100885	48113203	3	+	-1	4.31E-02
UBE2B	chr5	133737973	133752002	2	+	-1	4.93E-02
UPP1	chr7	48100885	48113203	2	+	-1	6.48E-02
LARP4	chr12	49092455	49107958	2	+	1	8.66E-02
NAP1L4	chr11	2956090	2970096	2	-	1	8.69E-02
NAP1L4	chr11	2956090	2970096	2	-	1	8.99E-02

Table 4.4 Fox-2-dependent mutually exclusive exons

symbol	chrom	chromStart	chromEnd	strand	direction of change	FDR
TPM2	chr9	35674485	35675335	-	1	1.06E-18
TPM1	chr15	61140121	61141528	+	-1	6.65E-07
COL5A1	chr9	136857459	136866870	+	-1	5.52E-06
FYN	chr6	112124173	112130929	-	1	5.26E-03

Table 4.5 Primers used for RT-PCR validation

Name	Sequence (5′--> 3′)
MYL6 F	AGTAGAGATGCTGGTGGCAGG
MYL6 R	ATTCACACAGGGAAAGGCACG
NFYA F	CAAACAGCAATAGTTCGACAG
NFYA R	GCCAGTTGATGTGATTAGCTG
SEPT6 F	CCAGGCTGGAGGCTCACAGAC
SEPT6 R	GGCCCAGCTCTGTTGCGCAGG
PICALM F	AAAGGTTGCACCAACAACCGC
PICALM R	CATGCCTGTTGGTGTAGTAGC
NF2 F	CACAATGAGAACTCCGACAGG
NF2 R	GCTAGAGCTCTTCAAAGAAGG
ADAM15 F	CCAATCTGGTCCCTCTGAACG
ADAM15 R	TTCGGGCTTCTCACCACCGGG
TPM1 F	CGTAAGCTGGTCATCATTGAG
TPM1 R (5′E)	AATTGTTCTTCCAGCTGTCGG
TPM1 R (3′E)	CTTCTCAGCCTGAGCCTCCAG
MPZL1 F	TGGTCCTAGGTCTCACTCTGC
MPZL1 R	GCATACACCACAGACTCTGAC
TRIP10 F	CTTCAGCCAGCCCATGAACCG
TRIP10 R	CCGTTTTCGCTGCTGCTCTGG
MYO18A F	CTGCCATTGAGGATGAGATGG
MYO18A R	GCTGCCTTGGAAGGTCCCTTG
ST7 F	AAGCCTGGAGAGAGAGAAACC
ST7 R	AGCAATGGTTGTTGCTTCCTC

MYL9 F	GTTTGACCAGTCCCAGATCC
MYL9 R	CCACTTCCTCATCTGTGAAGC
LRP8 F	CACTGCAACCAGGAGCAGGAC
LRP8 R	CCAATCTTGAGGTCAGTGCAG
RPS6KC1 F	GAACAGCCGAGTACCTCATGC
RPS6KC1 R	GTCTGTTCTGTCCTTGTGTCC
ZNF140 F	CGGAAAGCACCACGGAAACGC
ZNF140 R	TCTCTTTGAGCAGGCTGAAGC
CENPO F	CTTCTGGCCTGGGTGAGCTAG
CENPO R	CTCATCTCGCAGACGCCTTAG
FAM13B1 F	AGGGGAAGCAGCGTGTGTCAG
FAM13B1 R	GACAGCTGAGCTTCTCCAGAC
ROBO1 F	CTTCAGCATCTGGCTTTATCG
ROBO1 R	TGCTGACAGCTTCGCCTCCTC
EPB41L2 F	ACAAGATGGGGACGGCAGGAG
EPB41L2 R	TTCGTCTTTCCCAACCTCTGC
UPP1 F	GAGCCAATGCAGAGAAAGCTG
UPP1 R	CCAGCTTCTTGTTAAGGTCCG
FYN F	CTTGACAATGGTGGATACTAC
FYN R (5′E)	TTGGCATCCCTTTGTGACAGG
FYN R (3′E)	CTGACCCAGCTTCTTCTCCAG
COL5A1 F	GTTTACTGCAACTTCACAGCC
COL5A1 R (5′E)	TGGGCCAAGAAGTGATTCTGG
COL5A1 R (3′E)	GCTTGTACTGACTATACCAGG

References

Chapter 1

1. Crick, F.H., On protein synthesis. Symp Soc Exp Biol, 1958. **12**: p. 138-63.

2. Berget, S.M., C. Moore, and P.A. Sharp, Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A, 1977. **74**(8): p. 3171-5.

3. Chow, L.T., et al., An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell, 1977. **12**(1): p. 1-8.

4. Sakharkar, M.K., V.T. Chow, and P. Kangueane, Distributions of exons and introns in the human genome. In Silico Biol, 2004. **4**(4): p. 387-93.

5. Jurica, M.S. and M.J. Moore, Pre-mRNA splicing: awash in a sea of proteins. Mol Cell, 2003. **12**(1): p. 5-14.

6. Patel, A.A. and J.A. Steitz, Splicing double: insights from the second spliceosome. Nat Rev Mol Cell Biol, 2003. **4**(12): p. 960-70.

7. Wahl, M.C., C.L. Will, and R. Luhrmann, The spliceosome: design principles of a dynamic RNP machine. Cell, 2009. **136**(4): p. 701-18.

8. Wang, E.T., et al., Alternative isoform regulation in human tissue transcriptomes. Nature, 2008. **456**(7221): p. 470-6.

9. Cartegni, L., S.L. Chew, and A.R. Krainer, Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet, 2002. **3**(4): p. 285-98.

10. Long, J.C. and J.F. Caceres, The SR protein family of splicing factors: master regulators of gene expression. Biochem J, 2009. **417**(1): p. 15-27.

11. Xu, X., et al., ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. Cell, 2005. **120**(1): p. 59-72.

12. Jumaa, H., G. Wei, and P.J. Nielsen, Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20. Curr Biol, 1999. **9**(16): p. 899-902.

13. Ding, J.H., et al., Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. EMBO J, 2004. **23**(4): p. 885-96.

14. Wang, H.Y., et al., SC35 plays a role in T cell development and alternative splicing of CD45. Mol Cell, 2001. **7**(2): p. 331-42.

15. Wang, J., Y. Takagaki, and J.L. Manley, Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. Genes Dev, 1996. **10**(20): p. 2588-99.

16. Peng, X. and S.M. Mount, Genetic enhancement of RNA-processing defects by a dominant mutation in B52, the Drosophila gene for an SR protein splicing factor. Mol Cell Biol, 1995. **15**(11): p. 6273-82.

17. Ring, H.Z. and J.T. Lis, The SR protein B52/SRp55 is essential for Drosophila development. Mol Cell Biol, 1994. **14**(11): p. 7499-506.

18. Dreyfuss, G., V.N. Kim, and N. Kataoka, Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol, 2002. **3**(3): p. 195-205.

19. Zhu, J., A. Mayeda, and A.R. Krainer, Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol Cell, 2001. **8**(6): p. 1351-61.

20. Okunola, H.L. and A.R. Krainer, Cooperative-binding and splicing-repressive properties of hnRNP A1. Mol Cell Biol, 2009. **29**(20): p. 5620-31.

21. Blanchette, M. and B. Chabot, Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. EMBO J, 1999. **18**(7): p. 1939-52.

22. Chan, R.C. and D.L. Black, Conserved intron elements repress splicing of a neuron-specific c-src exon in vitro. Mol Cell Biol, 1997. **17**(5): p. 2970.

23. Mayeda, A., et al., Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. EMBO J, 1994. **13**(22): p. 5483-95.

24. Blanchette, M., et al., Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. Mol Cell, 2009. **33**(4): p. 438-49.

25. Jin, Y., et al., A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. EMBO J, 2003. **22**(4): p. 905-12.

26. Underwood, J.G., et al., Homologues of the Caenorhabditis elegans Fox-1 protein are neuronal splicing regulators in mammals. Mol Cell Biol, 2005. **25**(22): p. 10005-16.

27. Jensen, K.B., et al., Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. Neuron, 2000. **25**(2): p. 359-71.

28. Jensen, K.B., et al., The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. Proc Natl Acad Sci U S A, 2000. **97**(11): p. 5740-5.

29. Ule, J., et al., An RNA map predicting Nova-dependent splicing regulation. Nature, 2006. **444**(7119): p. 580-6.

30. Zhang, C., et al., Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. Genes Dev, 2008. **22**(18): p. 2550-63.

31. Lee, J.E. and T.A. Cooper, Pathogenic mechanisms of myotonic dystrophy. Biochem Soc Trans, 2009. **37**(Pt 6): p. 1281-6.

32. Ho, T.H., et al., Muscleblind proteins regulate alternative splicing. EMBO J, 2004. **23**(15): p. 3103-12.

33. Warzecha, C.C., et al., The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. RNA Biol, 2009. **6**(5): p. 546-62.

34. Warzecha, C.C., et al., ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. Mol Cell, 2009. **33**(5): p. 591-601.

35. Wu, J.I., et al., Function of quaking in myelination: regulation of alternative splicing. Proc Natl Acad Sci U S A, 2002. **99**(7): p. 4233-8.

36. Hinman, M.N. and H. Lou, Diverse molecular functions of Hu proteins. Cell Mol Life Sci, 2008. **65**(20): p. 3168-81.

37. Aznarez, I., et al., A systematic analysis of intronic sequences downstream of 5′ splice sites reveals a widespread role for U-rich motifs and TIA1/TIAL1 proteins in alternative splicing regulation. Genome Res, 2008. **18**(8): p. 1247-58.

38. Chawla, G., et al., Sam68 regulates a set of alternatively spliced exons during neurogenesis. Mol Cell Biol, 2009. **29**(1): p. 201-13.

39. Paronetto, M.P., et al., The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. J Cell Biol, 2007. **176**(7): p. 929-39.

40. Paronetto, M.P., et al., Alternative splicing of the cyclin D1 proto-oncogene is regulated by the RNA-binding protein Sam68. Cancer Res, 2010. **70**(1): p. 229-39.

41. Nilsen, T.W. and B.R. Graveley, Expansion of the eukaryotic proteome by alternative splicing. Nature, 2010. **463**(7280): p. 457-63.

42. Bentley, D.L., Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol, 2005. **17**(3): p. 251-6.

43. Kornblihtt, A.R., Coupling transcription and alternative splicing. Adv Exp Med Biol, 2007. **623**: p. 175-89.

44. Allo, M., et al., Control of alternative splicing through siRNA-mediated transcriptional gene silencing. Nat Struct Mol Biol, 2009. **16**(7): p. 717-24.

45. Kolasinska-Zwierz, P., et al., Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet, 2009. **41**(3): p. 376-81.

46. Schwartz, S., E. Meshorer, and G. Ast, Chromatin organization marks exon-intron structure. Nat Struct Mol Biol, 2009. **16**(9): p. 990-5.

47. Tilgner, H., et al., Nucleosome positioning as a determinant of exon recognition. Nat Struct Mol Biol, 2009. **16**(9): p. 996-1001.

48. Luco, R.F., et al., Regulation of alternative splicing by histone modifications. Science, 2010. **327**(5968): p. 996-1000.

49. Xiao, S.H. and J.L. Manley, Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. Genes Dev, 1997. **11**(3): p. 334-44.

50. Xiao, S.H. and J.L. Manley, Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. EMBO J, 1998. **17**(21): p. 6359-67.

51. Misteli, T., J.F. Caceres, and D.L. Spector, The dynamics of a pre-mRNA splicing factor in living cells. Nature, 1997. **387**(6632): p. 523-7.

52. Misteli, T., et al., Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. J Cell Biol, 1998. **143**(2): p. 297-307.

53. Colwill, K., et al., The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. EMBO J, 1996. **15**(2): p. 265-75.

54. Nayler, O., S. Stamm, and A. Ullrich, Characterization and comparison of four serine- and arginine-rich (SR) protein kinases. Biochem J, 1997. **326 (Pt 3)**: p. 693-700.

55. Gui, J.F., W.S. Lane, and X.D. Fu, A serine kinase regulates intracellular localization of splicing factors in the cell cycle. Nature, 1994. **369**(6482): p. 678-82.

56. Koizumi, J., et al., The subcellular localization of SF2/ASF is regulated by direct interaction with SR protein kinases (SRPKs). J Biol Chem, 1999. **274**(16): p. 11125-31.

57. Rossi, F., et al., Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. Nature, 1996. **381**(6577): p. 80-2.

58. Tang, Z., M. Yanagida, and R.J. Lin, Fission yeast mitotic regulator Dsk1 is an SR protein-specific kinase. J Biol Chem, 1998. **273**(10): p. 5963-9.

59. Misteli, T. and D.L. Spector, Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors. Mol Biol Cell, 1996. **7**(10): p. 1559-72.

60. Novoyatleva, T., et al., Protein phosphatase 1 binds to the RNA recognition motif of several splicing factors and regulates alternative pre-mRNA processing. Hum Mol Genet, 2008. **17**(1): p. 52-70.

61. Feng, Y., M. Chen, and J.L. Manley, Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. Nat Struct Mol Biol, 2008. **15**(10): p. 1040-8.

62. Shin, C., Y. Feng, and J.L. Manley, Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. Nature, 2004. **427**(6974): p. 553-8.

63. Shin, C. and J.L. Manley, The SR protein SRp38 represses splicing in M phase cells. Cell, 2002. **111**(3): p. 407-17.

64. Shi, Y. and J.L. Manley, A complex signaling pathway regulates SRp38 phosphorylation and pre-mRNA splicing in response to heat shock. Mol Cell, 2007. **28**(1): p. 79-90.

65. Allemand, E., et al., Regulation of heterogenous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock. Proc Natl Acad Sci U S A, 2005. **102**(10): p. 3605-10.

66. van der Houven van Oordt, W., et al., The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. J Cell Biol, 2000. **149**(2): p. 307-16.

67. Shi, Y., B. Reddy, and J.L. Manley, PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. Mol Cell, 2006. **23**(6): p. 819-29.

68. Brahms, H., et al., Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B′ and the Sm-like protein LSm4, and their interaction with the SMN protein. RNA, 2001. **7**(11): p. 1531-42.

69. Miranda, T.B., et al., Spliceosome Sm proteins D1, D3, and B/B′ are asymmetrically dimethylated at arginine residues in the nucleus. Biochem Biophys Res Commun, 2004. **323**(2): p. 382-7.

70. Liu, Q. and G. Dreyfuss, In vivo and in vitro arginine methylation of RNA-binding proteins. Mol Cell Biol, 1995. **15**(5): p. 2800-8.

71. Nichols, R.C., et al., The RGG domain in hnRNP A2 affects subcellular localization. Exp Cell Res, 2000. **256**(2): p. 522-32.

72. Sinha, R., et al., Arginine Methylation Controls the Subcellular Localization and Functions of the Oncoprotein Splicing Factor SF2/ASF. Mol Cell Biol, 2010.

73. Konig, H., H. Ponta, and P. Herrlich, Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. EMBO J, 1998. **17**(10): p. 2904-13.

74. Matter, N., P. Herrlich, and H. Konig, Signal-dependent regulation of splicing via phosphorylation of Sam68. Nature, 2002. **420**(6916): p. 691-5.

75. Lynch, K.W. and A. Weiss, A model system for activation-induced alternative splicing of CD45 pre-mRNA in T cells implicates protein kinase C and Ras. Mol Cell Biol, 2000. **20**(1): p. 70-80.

76. Topp, J.D., et al., A cell-based screen for splicing regulators identifies hnRNP LL as a distinct signal-induced repressor of CD45 variable exon 4. RNA, 2008. **14**(10): p. 2038-49.

77. Melton, A.A., et al., Combinatorial control of signal-induced exon repression by hnRNP L and PSF. Mol Cell Biol, 2007. **27**(19): p. 6972-84.

78. Xie, J. and D.L. Black, A CaMK IV responsive RNA element mediates depolarization-induced alternative splicing of ion channels. Nature, 2001. **410**(6831): p. 936-9.

79. Lee, J.A., et al., Depolarization and CaM kinase IV modulate NMDA receptor splicing through two essential RNA elements. PLoS Biol, 2007. **5**(2): p. e40.

80. Xie, J., et al., A consensus CaMK IV-responsive RNA sequence mediates regulation of alternative exons in neurons. RNA, 2005. **11**(12): p. 1825-34.

81. Li, Q., J.A. Lee, and D.L. Black, Neuronal regulation of alternative pre-mRNA splicing. Nat Rev Neurosci, 2007. **8**(11): p. 819-31.

82. Hallegger, M., M. Llorian, and C.W. Smith, Alternative splicing: global insights. FEBS J, 2010. **277**(4): p. 856-66.

83. Wilhelm, B.T., et al., Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature, 2008. **453**(7199): p. 1239-43.

84. Jensen, K.B. and R.B. Darnell, CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. Methods Mol Biol, 2008. **488**: p. 85-98.

85. Yeo, G.W., et al., An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat Struct Mol Biol, 2009. **16**(2): p. 130-7.

86. Furger, A., et al., Promoter proximal splice sites enhance transcription. Genes Dev, 2002. **16**(21): p. 2792-9.

87. Kameoka, S., P. Duque, and M.M. Konarska, p54(nrb) associates with the 5′ splice site within large transcription/splicing complexes. EMBO J, 2004. **23**(8): p. 1782-91.

88. Lin, S., et al., The splicing factor SC35 has an active role in transcriptional elongation. Nat Struct Mol Biol, 2008. **15**(8): p. 819-26.

89. Maniatis, T. and R. Reed, An extensive network of coupling among gene expression machines. Nature, 2002. **416**(6880): p. 499-506.

90. McCracken, S., M. Lambermon, and B.J. Blencowe, SRm160 splicing coactivator promotes transcript 3′-end cleavage. Mol Cell Biol, 2002. **22**(1): p. 148-60.

91. Castelo-Branco, P., et al., Polypyrimidine tract binding protein modulates efficiency of polyadenylation. Mol Cell Biol, 2004. **24**(10): p. 4174-83.

92. Millevoi, S., et al., A physical and functional link between splicing factors promotes pre-mRNA 3′ end processing. Nucleic Acids Res, 2009. **37**(14): p. 4672-83.

93. Licatalosi, D.D., et al., HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature, 2008. **456**(7221): p. 464-9.

94. Masuda, S., et al., Recruitment of the human TREX complex to mRNA during splicing. Genes Dev, 2005. **19**(13): p. 1512-7.

95. Cheng, H., et al., Human mRNA export machinery recruited to the 5′ end of mRNA. Cell, 2006. **127**(7): p. 1389-400.

96. Kohler, A. and E. Hurt, Exporting RNA from the nucleus to the cytoplasm. Nat Rev Mol Cell Biol, 2007. **8**(10): p. 761-73.

97. Valencia, P., A.P. Dias, and R. Reed, Splicing promotes rapid and efficient mRNA export in mammalian cells. Proc Natl Acad Sci U S A, 2008. **105**(9): p. 3386-91.

98. Tange, T.O., A. Nott, and M.J. Moore, The ever-increasing complexities of the exon junction complex. Curr Opin Cell Biol, 2004. **16**(3): p. 279-84.

99. Le Hir, H., et al., The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J, 2001. **20**(17): p. 4987-97.

100. Luo, M.J. and R. Reed, Splicing is required for rapid and efficient mRNA export in metazoans. Proc Natl Acad Sci U S A, 1999. **96**(26): p. 14937-42.

101. Huang, Y. and J.A. Steitz, SRprises along a messenger's journey. Mol Cell, 2005. **17**(5): p. 613-5.

102. Le Hir, H. and G.R. Andersen, Structural insights into the exon junction complex. Curr Opin Struct Biol, 2008. **18**(1): p. 112-9.

103. Le Hir, H. and B. Seraphin, EJCs at the heart of translational control. Cell, 2008. **133**(2): p. 213-6.

104. Diem, M.D., et al., PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. Nat Struct Mol Biol, 2007. **14**(12): p. 1173-9.

105. Ma, X.M., et al., SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. Cell, 2008. **133**(2): p. 303-13.

106. Bhaskar, P.T. and N. Hay, The two TORCs and Akt. Dev Cell, 2007. **12**(4): p. 487-502.

107. Sanford, J.R., et al., A novel role for shuttling SR proteins in mRNA translation. Genes Dev, 2004. **18**(7): p. 755-68.

108. Michlewski, G., J.R. Sanford, and J.F. Caceres, The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. Mol Cell, 2008. **30**(2): p. 179-89.

109. Nagy, E. and L.E. Maquat, A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci, 1998. **23**(6): p. 198-9.

110. Chang, Y.F., J.S. Imam, and M.F. Wilkinson, The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem, 2007. **76**: p. 51-74.

111. Kashima, I., et al., Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev, 2006. **20**(3): p. 355-67.

112. Zhang, Z. and A.R. Krainer, Involvement of SR proteins in mRNA surveillance. Mol Cell, 2004. **16**(4): p. 597-607.

Chapter 2

1. Wang, E.T., et al., Alternative isoform regulation in human tissue transcriptomes. Nature, 2008. **456**(7221): p. 470-6.

2. Cartegni, L., S.L. Chew, and A.R. Krainer, Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet, 2002. **3**(4): p. 285-98.

3. Zhang, Z. and A.R. Krainer, Involvement of SR proteins in mRNA surveillance. Mol Cell, 2004. **16**(4): p. 597-607.

4. Huang, Y., et al., SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. Mol Cell, 2003. **11**(3): p. 837-43.

5. Lai, M.C. and W.Y. Tarn, Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. J Biol Chem, 2004. **279**(30): p. 31745-9.

6. Sanford, J.R., et al., A novel role for shuttling SR proteins in mRNA translation. Genes Dev, 2004. **18**(7): p. 755-68.

7. Hanamura, A., et al., Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. Rna, 1998. **4**(4): p. 430-44.

8. Li, X. and J.L. Manley, Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell, 2005. **122**(3): p. 365-78.

9. Li, X., J. Wang, and J.L. Manley, Loss of splicing factor ASF/SF2 induces G2 cell cycle arrest and apoptosis, but inhibits internucleosomal DNA fragmentation. Genes Dev, 2005. **19**(22): p. 2705-14.

10. Xu, X., et al., ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. Cell, 2005. **120**(1): p. 59-72.

11. Karni, R., et al., The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat Struct Mol Biol, 2007. **14**(3): p. 185-93.

12. Ghigna, C., et al., Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. Mol Cell, 2005. **20**(6): p. 881-90.

13. McGlincy, N.J. and C.W. Smith, Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? Trends Biochem Sci, 2008. **33**(8): p. 385-93.

14. Barreau, C., L. Paillard, and H.B. Osborne, AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res, 2005. **33**(22): p. 7138-50.

15. Valencia-Sanchez, M.A., et al., Control of translation and mRNA degradation by miRNAs and siRNAs. Genes Dev, 2006. **20**(5): p. 515-24.

16. Morrison, M., K.S. Harris, and M.B. Roth, smg mutants affect the expression of alternatively spliced SR protein mRNAs in Caenorhabditis elegans. Proc Natl Acad Sci U S A, 1997. **94**(18): p. 9782-5.

17. Sureau, A., et al., SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. Embo J, 2001. **20**(7): p. 1785-96.

18. Wollerton, M.C., et al., Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell, 2004. **13**(1): p. 91-100.

19. Jumaa, H. and P.J. Nielsen, The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. Embo J, 1997. **16**(16): p. 5077-85.

20. Lareau, L.F., et al., Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature, 2007. **446**(7138): p. 926-9.

21. Ni, J.Z., et al., Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev, 2007. **21**(6): p. 708-18.

22. Bejerano, G., et al., Ultraconserved elements in the human genome. Science, 2004. **304**(5675): p. 1321-5.

23. Sandelin, A., et al., Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC Genomics, 2004. **5**(1): p. 99.

24. Dickins, R.A., et al., Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. Nat Genet, 2005. **37**(11): p. 1289-95.

25. Ge, H., P. Zuo, and J.L. Manley, Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. Cell, 1991. **66**(2): p. 373-82.

26. Krainer, A.R., et al., Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and Drosophila splicing regulators. Cell, 1991. **66**(2): p. 383-94.

27. Isken, O. and L.E. Maquat, Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes Dev, 2007. **21**(15): p. 1833-56.

28. Tanguay, R.L. and D.R. Gallie, Translational efficiency is regulated by the length of the 3′ untranslated region. Mol Cell Biol, 1996. **16**(1): p. 146-56.

29. Pestova, P.V., J.R. Lorsch, and C. Hellen, The Mechanism of Translation Initiation in Eukaryotes. In Translational Control in Biology and Medicine, M Mathews, N. Sonenberg, J Hershey, ed. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press). 2007: p. pp. 87-128.

30. Bergamini, G., T. Preiss, and M.W. Hentze, Picornavirus IRESes and the poly(A) tail jointly promote cap-independent translation in a mammalian cell-free system. Rna, 2000. **6**(12): p. 1781-90.

31. Murchison, E.P. and G.J. Hannon, miRNAs on the move: miRNA biogenesis and the RNAi machinery. Curr Opin Cell Biol, 2004. **16**(3): p. 223-9.

32. Cummins, J.M., et al., The colorectal microRNAome. Proc Natl Acad Sci U S A, 2006. **103**(10): p. 3687-92.

33. Murchison, E.P., et al., Characterization of Dicer-deficient murine embryonic stem cells. Proc Natl Acad Sci U S A, 2005. **102**(34): p. 12135-40.

34. Doudna, J.A. and P. Sarnow, Translation Initiation by Viral Internal Ribosome Entry Sites. In Translational Control in Biology and Medicine, M Mathews, N. Sonenberg, J Hershey, ed. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press). 2007: p. pp. 129-154.

35. Isken, O., et al., Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. Cell, 2008. **133**(2): p. 314-27.

36. Boutz, P.L., et al., A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev, 2007. **21**(13): p. 1636-52.

37. Makeyev, E.V., et al., The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. Mol Cell, 2007. **27**(3): p. 435-48.

38. Boutz, P.L., et al., MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. Genes Dev, 2007. **21**(1): p. 71-84.

39. Tacke, R., A. Boned, and C. Goridis, ASF alternative transcripts are highly conserved between mouse and man. Nucleic Acids Res, 1992. **20**(20): p. 5482.

40. Stutz, F. and E. Izaurralde, The interplay of nuclear mRNP assembly, mRNA surveillance and export. Trends Cell Biol, 2003. **13**(6): p. 319-27.

41. Li, Y., et al., An intron with a constitutive transport element is retained in a Tap messenger RNA. Nature, 2006. **443**(7108): p. 234-7.

42. Moore, M.J., From birth to death: the complex lives of eukaryotic mRNAs. Science, 2005. **309**(5740): p. 1514-8.

43. Baek, D., et al., The impact of microRNAs on protein output. Nature, 2008. **455**(7209): p. 64-71.

44. Giorgi, C. and M.J. Moore, The nuclear nurture and cytoplasmic nature of localized mRNPs. Semin Cell Dev Biol, 2007. **18**(2): p. 186-93.

45. Michlewski, G., J.R. Sanford, and J.F. Caceres, The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. Mol Cell, 2008. **30**(2): p. 179-89.

46. Sanford, J.R., et al., Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. PLoS ONE, 2008. **3**(10): p. e3369.

47. Cáceres, J.F., et al., Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. J Cell Biol, 1997. **138**(2): p. 225-38.

48. Cazalla, D., et al., Nuclear export and retention signals in the RS domain of SR proteins. Mol Cell Biol, 2002. **22**(19): p. 6871-82.

49. Durocher, Y., S. Perret, and A. Kamen, High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. Nucleic Acids Res, 2002. **30**(2): p. E9.

50. Bor, Y.C., et al., Northern Blot analysis of mRNA from mammalian polyribosomes. Nature Protocols, 2006: p. doi:10.1038/nprot.2006.216.

Chapter 3

1. Shibata, H., D.P. Huynh, and S.M. Pulst, A novel protein with RNA-binding motifs interacts with ataxin-2. Hum Mol Genet, 2000. **9**(9): p. 1303-13.

2. Lastres-Becker, I., U. Rub, and G. Auburger, Spinocerebellar ataxia 2 (SCA2). Cerebellum, 2008. **7**(2): p. 115-24.

3. Kiehl, T.R., H. Shibata, and S.M. Pulst, The ortholog of human ataxin-2 is essential for early embryonic patterning in C. elegans. J Mol Neurosci, 2000. **15**(3): p. 231-41.

4. Huh, G.S. and R.O. Hynes, Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element. Genes Dev, 1994. **8**(13): p. 1561-74.

5. Lim, L.P. and P.A. Sharp, Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. Mol Cell Biol, 1998. **18**(7): p. 3900-6.

6. Kawamoto, S., Neuron-specific alternative splicing of nonmuscle myosin II heavy chain-B pre-mRNA requires a cis-acting intron sequence. J Biol Chem, 1996. **271**(30): p. 17613-6.

7. Hedjran, F., et al., Control of alternative pre-mRNA splicing by distributed pentameric repeats. Proc Natl Acad Sci U S A, 1997. **94**(23): p. 12343-7.

8. Modafferi, E.F. and D.L. Black, A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon. Mol Cell Biol, 1997. **17**(11): p. 6537-45.

9. Jin, Y., et al., A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. EMBO J, 2003. **22**(4): p. 905-12.

10. Kuroyanagi, H., Fox-1 family of RNA-binding proteins. Cell Mol Life Sci, 2009. **66**(24): p. 3895-907.

11. Underwood, J.G., et al., Homologues of the Caenorhabditis elegans Fox-1 protein are neuronal splicing regulators in mammals. Mol Cell Biol, 2005. **25**(22): p. 10005-16.

12. Kim, K.K., R.S. Adelstein, and S. Kawamoto, Identification of neuronal nuclei (NeuN) as Fox-3, a new member of the Fox-1 gene family of splicing factors. J Biol Chem, 2009. **284**(45): p. 31052-61.

13. Nakahata, S. and S. Kawamoto, Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. Nucleic Acids Res, 2005. **33**(7): p. 2078-89.

14. Lee, J.A., Z.Z. Tang, and D.L. Black, An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. Genes Dev, 2009. **23**(19): p. 2284-93.

15. Zhou, H.L., A.P. Baraniak, and H. Lou, Role for Fox-1/Fox-2 in mediating the neuronal pathway of calcitonin/calcitonin gene-related peptide alternative RNA processing. Mol Cell Biol, 2007. **27**(3): p. 830-41.

16. Tang, Z.Z., et al., Developmental control of CaV1.2 L-type calcium channel splicing by Fox proteins. Mol Cell Biol, 2009. **29**(17): p. 4757-65.

17. Baraniak, A.P., J.R. Chen, and M.A. Garcia-Blanco, Fox-2 mediates epithelial cell-specific fibroblast growth factor receptor 2 exon choice. Mol Cell Biol, 2006. **26**(4): p. 1209-22.

18. Damianov, A. and D.L. Black, Autoregulation of Fox protein expression to produce dominant negative splicing factors. RNA, 2010. **16**(2): p. 405-16.

19. Zhang, C., et al., Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. Genes Dev, 2008. **22**(18): p. 2550-63.

20. Yeo, G.W., et al., An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat Struct Mol Biol, 2009. **16**(2): p. 130-7.

21. Venables, J.P., et al., Cancer-associated regulation of alternative splicing. Nat Struct Mol Biol, 2009. **16**(6): p. 670-6.

22. Zhou, H.L. and H. Lou, Repression of prespliceosome complex formation at two distinct steps by Fox-1/Fox-2 proteins. Mol Cell Biol, 2008. **28**(17): p. 5507-16.

23. Fukumura, K., et al., U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. Nucleic Acids Res, 2009. **37**(6): p. 1907-14.

24. Mauger, D.M., C. Lin, and M.A. Garcia-Blanco, hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. Mol Cell Biol, 2008. **28**(17): p. 5403-19.

25. Fukumura, K., et al., Tissue-specific splicing regulator Fox-1 induces exon skipping by interfering E complex formation on the downstream intron of human F1gamma gene. Nucleic Acids Res, 2007. **35**(16): p. 5303-11.

26. Ohkura, N., et al., Coactivator-associated arginine methyltransferase 1, CARM1, affects pre-mRNA splicing in an isoform-specific manner. J Biol Chem, 2005. **280**(32): p. 28927-35.

27. Hua, Y., et al., Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. Am J Hum Genet, 2008. **82**(4): p. 834-48.

28. Auweter, S.D., et al., Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. EMBO J, 2006. **25**(1): p. 163-73.

29. Mayeda, A., et al., Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. EMBO J, 1994. **13**(22): p. 5483-95.

30. Lee, B.J., et al., Rules for nuclear localization sequence recognition by karyopherin beta 2. Cell, 2006. **126**(3): p. 543-58.

31. Long, J.C. and J.F. Caceres, The SR protein family of splicing factors: master regulators of gene expression. Biochem J, 2009. **417**(1): p. 15-27.

32. Graveley, B.R. and T. Maniatis, Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. Mol Cell, 1998. **1**(5): p. 765-71.

33. Shaw, S.D., et al., Deletion of the N-terminus of SF2/ASF permits RS-domain-independent pre-mRNA splicing. PLoS One, 2007. **2**(9): p. e854.

34. Zhu, J. and A.R. Krainer, Pre-mRNA splicing in the absence of an SR protein RS domain. Genes Dev, 2000. **14**(24): p. 3166-78.

35. Rhodes, G.H., et al., The p542 gene encodes an autoantigen that cross-reacts with EBNA-1 of the Epstein Barr virus and which may be a heterogeneous nuclear ribonucleoprotein. J Autoimmun, 1997. **10**(5): p. 447-54.

36. Jurica, M.S., et al., Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. RNA, 2002. **8**(4): p. 426-39.

37. Greco, A., et al., The DNA rearrangement that generates the TRK-T3 oncogene involves a novel gene on chromosome 3 whose product has a potential coiled-coil domain. Mol Cell Biol, 1995. **15**(11): p. 6118-27.

38. Roccato, E., et al., Analysis of SHP-1-mediated down-regulation of the TRK-T3 oncoprotein identifies Trk-fused gene (TFG) as a novel SHP-1-interacting protein. J Biol Chem, 2005. **280**(5): p. 3382-9.

39. Miranda, C., et al., The TFG protein, involved in oncogenic rearrangements, interacts with TANK and NEMO, two proteins involved in the NF-kappaB pathway. J Cell Physiol, 2006. **208**(1): p. 154-60.

Chapter 4

1. Pan, Q., et al., Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature Genet, 2008. 40(12): p. 1413-1415.

2. Wang, E.T., et al., Alternative isoform regulation in human tissue transcriptomes. Nature, 2008. 456(7221): p. 470-476.

3. Castle, J.C., et al., Differential expression of 24,426 human alternative splicing events and predicted cis-regulation in 48 tissues. Nature Genet., 2008. 40: p. 1416-1425.

4. Clark, T., et al., Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biol., 2007. 8(4): p. R64.

5. Mortazavi, A., et al., Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Meth, 2008. 5(7): p. 621-628.

6. Kuroyanagi, H., et al., Transgenic alternative-splicing reporters reveal tissue-specific expression profiles and regulation mechanisms in vivo. Nat Meth, 2006. 3(11): p. 909-915.

7. Marioni, J., et al., RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res., 2008. 18(9): p. 1509-1517.

8. Kuroyanagi, H., Fox-1 family of RNA-binding proteins. Cell Mol Life Sci, 2009. 10.1007/s00018-009-0120-5.

9. Ponthier, J.L., et al., Fox-2 splicing factor binds to a conserved intron motif to promote inclusion of protein 4.1R alternative exon 16. J Biol Chem, 2006. 281(18): p. 12468 - 12474.

10. Zhang, C., et al., Defining the splicing regulatory network of tissue-specific splicing factors Fox-1/2. Genes Dev, 2008. 22: p. 2250-2263.

11. Yeo, G.W., et al., An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat Struct Mol Biol, 2009. 16: p. 130-137.

12. Baraniak, A.P., J.R. Chen, and M.A. Garcia-Blanco, Fox-2 mediates epithelial cell-specific fibroblast growth factor receptor 2 exon choice. Mol. Cell. Biol., 2006. 26(4): p. 1209-1222.

13. Lim, L.P. and P.A. Sharp, Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. Mol. Cell. Biol., 1998. 18(7): p. 3900-3906.

14. Jin, Y., et al., A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. EMBO J., 2003. 22(4): p. 905-912.

15. Nakahata, S. and S. Kawamoto, Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. Nucleic Acids Res., 2005. 33(7): p. 2078-2089.

16. Underwood, J.G., et al., Homologues of the Caenorhabditis elegans Fox-1 protein are neuronal splicing regulators in mammals. Mol. Cell. Biol., 2005. 25(22): p. 10005-10016.

17. Deguillien, M., et al., Multiple cis elements regulate an alternative splicing event at 4.1R pre-mRNA during erythroid differentiation. Blood, 2001. 98(13): p. 3809-3816.

18. Benjamini, Y. and Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Soc. B, 1995. 57(289-300).

19. Ule, J., et al., Nova regulates brain-specific splicing to shape the synapse. Nature Genet., 2005. 37(8): p. 844-852.

20. Bailey, T. and C. Elkan. Fitting a mixture model by expectation maximization to discover motifs in biopolymers in Proc. Int. Conf. Intell. Syst. Mol. Biol. 1994. Menlo Park, California: AAAI Press.

21. Smith, A.D., P. Sumazin, and M.Q. Zhang, Identifying tissue-selective transcription factor binding sites in vertebrate promoters. Proc Natl Acad Sci U S A, 2005. 102(5): p. 1560-1565.

22. Ule, J., et al., An RNA map predicting Nova-dependent splicing regulation. Nature, 2006. 444: p. 580-586.

23. Licatalosi, D.D., et al., HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature, 2008. 456(7221): p. 464-469.

24. Han, X., et al., Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing. Proc Natl Acad Sci U S A, 2009. 106(31): p. 12741-12746.

25. Zhang, C., et al., Dual-specificity splice sites function alternatively as 5′ and 3′ splice sites. Proc. Natl. Acad. Sci. USA, 2007. 104(38): p. 15028-15033.

26. Kuhn, R.M., et al., The UCSC Genome Browser Database: update 2009. Nucl. Acids Res., 2009. 37(suppl_1): p. D755-761.

27. Ule, J., et al., CLIP identifies Nova-regulated RNA networks in the brain. Science, 2003. 302(5648): p. 1212-1215.