THE UNIVERSITY
of EDINBURGH

Investigation of the physiological roles of SRSF1-mediated translation

Fiona Haward

THE UNIVERSITY *of* EDINBURGH

Doctor of Philosophy

The University of Edinburgh

2017

Declaration

I declare that this thesis is a result of my own work, unless otherwise stated, where the contributor has been fully acknowledged. This thesis has not been submitted in any form for any degree, qualification or similar at any other University or Institution.

Fiona Haward,

September, 2017

Acknowledgements

The past four years have been some of the most exciting and enjoyable, but also some of the most challenging. The work undertaken to complete this thesis would not have been even half possible without the wealth of support I have received from numerous people, both within the IGMM and closer to home.

Firstly, I would like to thank my supervisor Javier Cáceres for persuading me to embark on this project, when, in his words, he could convince no other to do so. Throughout which he has provided invaluable scientific direction and discussion regarding a field to which his knowledge is unparalleled. Of course, I also have to mention the infamously well-catered, and sometimes long, Cáceres Lab meetings!

I also owe a huge amount of gratitude to Magdalena Maslon, who has provided me with the most incredible support during my PhD and is instrumental in the huge amount I have learnt during this time; it has been a privilege to work alongside her. She was particularly patient and caring when the project was tough or I just didn't understand, especially in certain (mathematical!) departments. I would also like to thank the other members of the Cáceres Lab, past and present, for all the help, enjoyment and answers to obscure questions that they have provided over the years. I would especially like to acknowledge Dasa Longman for the excellent lab and life advice, the latter has guided me in many a low moment.

Elsewhere within the IGMM, I would like to acknowledge Ian Adams for collaborating with the mice work for this project and providing an incredible knowledge for mouse development and genetics. Furthermore, I would like to thank Dr Andrew Wood for all the CRISPR advice during the project, without which, progress would have fallen short.

On a personal note, I would like to thank Ailsa Revuelta, Katy McLaughlin, Flora Dix and Louise Cleal, who have provided countless laughs, and Christine Mordstein for the early morning discussions, which (occasionally) featured RNA. Also to Adam Douglas, whose encouragement, kindness and support was unrivalled at times when I needed it the most. Finally, I would like to thank my family, in particular my parents and brother, whose unwavering care, support and humour (and tolerance of constant phonecalls!) has been more than I could have wished for, I owe much of this effort to them.

Abstract

The serine/arginine-rich (SR-) family proteins constitute a diverse group of pre-mRNA splicing factors that are essential for viability. They can be characterised based on the presence of one or two RRMs and an RS domain. A subset, of which SRSF1 is the prototype, is capable of nucleocytoplasmic shuttling; a process governed by continual cyclic phosphorylation of the RS domain. In contrast, SRSF2, another member of the SR family, is unable to shuttle due to the presence of a nuclear retention sequence (NRS) at the C-terminus of its RS domain. When this NRS is fused to SRSF1, it prevents nucleocytoplasmic shuttling of the SRSF1-NRS fusion protein. In addition to its nuclear roles, SRSF1 is directly associated with the translation machinery and can activate mRNA translation of target transcripts via an mTOR-dependent mechanism. The specific mRNA translational targets that SRSF1 serves to regulate encode numerous factors including RNA processing factors and cell-cycle proteins.

The aim of this work is to study the physiological relevance of SRSF1 cytoplasmic functions, as previous data have relied on overexpression systems. CRISPR/Cas9 editing was used to knock-in the NRS naturally present in SRSF2 at the SRSF1 genomic locus, creating an SRSF1-NRS fusion protein. After numerous attempts, it was only possible to obtain a single viable homozygous clone in mouse embryonic stem cells (mESCs), despite being able to successfully tag the genomic SRSF1 locus. This strongly suggests that the ablation of SRSF1 shuttling ability is highly selected against in mESCs. To assess the physiological importance of SRSF1 nucleocytoplasmic shuttling during development, a mouse model for SRSF1-NRS was also developed. SRSF1-NRS homozygous mice are born at correct Mendelian ratios, but are small in size and present with severe hydrocephalus. Finally, proteomics was used to identify interactors of endogenous cytoplasmic SRSF1 and those that bind the NRS of SRSF2 to gain insights into the mechanism of nuclear retention for non-shuttling SR proteins. In summary, this work analyses the physiological relevance of cytoplasmic SRSF1 function and the consequences of the SRSF1-NRS allele in mouse development.

Lay Abstract

Every cell of an organism contains the same genetic material consisting of DNA. DNA contains units known as genes, which are transcribed, processed, and eventually expressed as proteins that execute numerous cellular functions. Interspersed within the protein-coding DNA, termed exons, some sequence does not encode for protein. These sequences are called introns.

During transcription, the full DNA sequence of the gene is copied into a molecule called messenger RNA (mRNA), which is the template for protein synthesis. Therefore, to produce a continuous, functional mRNA transcript, the cell removes the introns and joins the exons together in a process called mRNA splicing. To perform splicing, the cell requires certain proteins called splicing factors. One essential splicing factor is SRSF1, which can regulate specific mRNA transcripts.

The distinct molecular machineries required to produce mRNA or protein molecules are in different compartments of the cell. First, mRNA is synthesised and spliced in the nucleus and is then exported into the cytoplasm to be translated into protein. SRSF1 is an unusual splicing factor with multiple roles in gene expression. In the nucleus, SRSF1 promotes mRNA splicing, and in the cytoplasm, it facilitates translation of this same spliced mRNA into protein. However, the role of SRSF1 in protein synthesis is not well understood.

This work aims to investigate the roles of SRSF1 function in protein synthesis and its relevance for physiology. Both cellular and whole organism models have been engineered to create a version of SRSF1 that is permanently retained in the nucleus and therefore unable to move to the cytoplasm to act on protein synthesis. This model has been used to investigate the physiological consequences on gene expression when SRSF1 is retained in the nucleus.

Table of Contents

List of Figures

Introduction

Investigation of cytoplasmic functions of SRSF1 using CRISPR-mediated targeting in mouse ES cells:

Investigating the mechanisms of nuclear retention of a non-shuttling SR protein:

Investigation ofSRSF1 function in the cytoplasm using mouse models:

List of Tables

List of Abbreviations

Abbreviation	Definition
3'UTR	3'untranslated region
5'UTR	5'untranslated region
AS	Alternative splicing
ATP	Adenosine triphosphate
cDNA	Complementary DNA
CLIP	UV cross-linking and immunoprecipitation
CRISPR	Clustered regularly interspersed palindromic repeats
DAPI	4',6'-diamidino-2-phenylinodole
DMEM	Dulbecco's Modified Eagle Media
DNA	Deoxyribonucleic acid
EJC	Exon Junction Complex
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
gDNA	Genomic DNA
gRNA	Guide RNA
GMEM	Glasgow Minimum Essential Medium
GFP	Green Fluorescent protein
IP	immunoprecipitation
IRES	Internal ribosome entry site
LFQ	Label-free quantification
LIF	Leukaemia inhibitory factor
MEFs	Mouse embryonic fibroblasts
mESCs	Mouse embryonic stem cells
miRNA	microRNA
mRNP	Messenger ribonucleoprotein
mRNA	Messenger RNA
MS	Mass spectrometry
NMD	Nonsense-mediated decay
NRS	Nuclear Retention Signal

NSCs	Neural stem cells
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
RBP	RNA binding protein
RNA	Ribonucleic Acid
RRM	RNA recognition motif
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulphate
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoproteins
siRNA	Small interfering RNA
TBS	Tris buffered saline
WT	Wild type

Chapter 1

Introduction

1.1 Gene expression regulation and RNA binding proteins

Gene expression is a highly complex process, which is subject to an unprecedented degree of post-transcriptional regulation. In response to cell-specific intra and extracellular cues, it is rapidly manipulated to shape the proteome. RNA binding proteins (RBPs) are key orchestrators in these intricate processes and crucial to maintain cellular homeostasis (Figure 1.1).

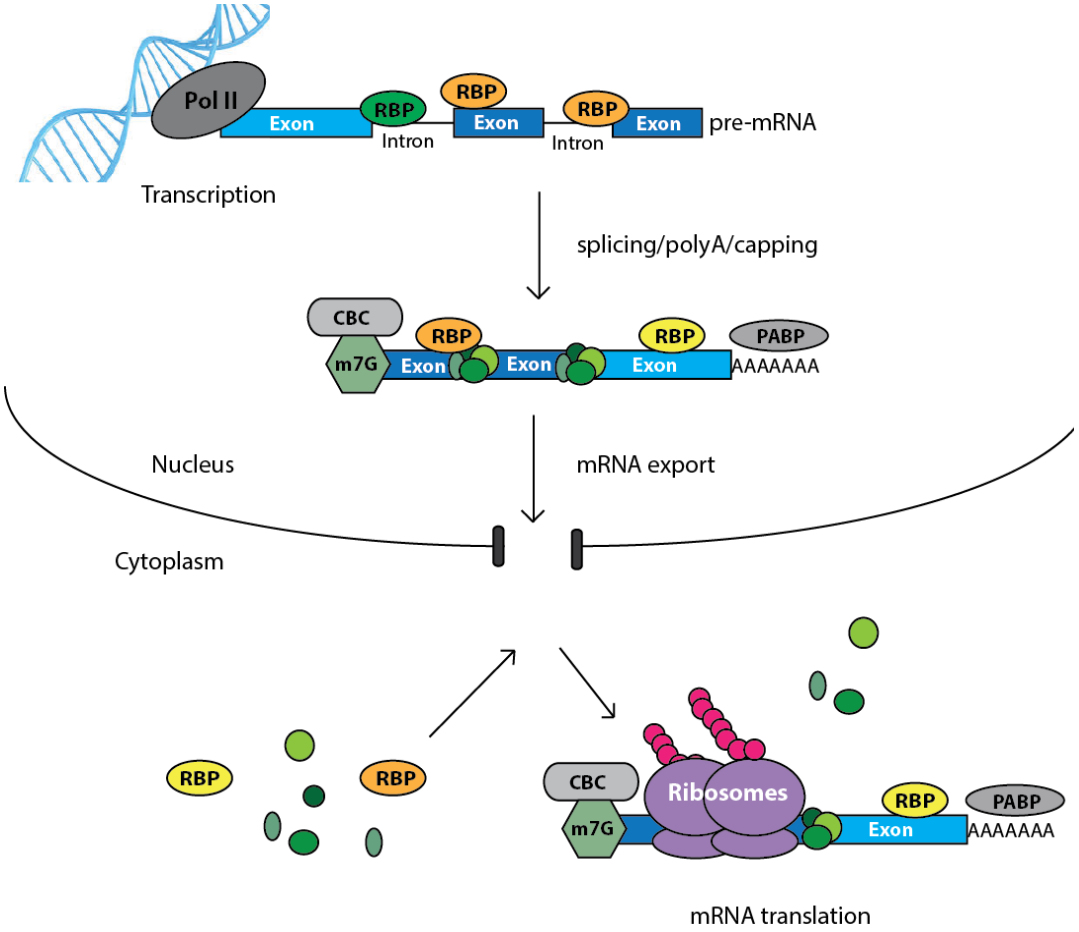


Figure 1.1: RNA-binding proteins (RBPs) in gene expression regulation. RBPs act in a concerted manner to regulate every stage of gene expression, from transcription to mRNA translation, immediately as the nascent mRNA is synthesised. The resulting messenger ribonucleoprotein (mRNP) complex is highly dynamic, undergoing numerous remodelling steps in a transcript-specific manner. *Adapted from:* (Calabretta & Richard, 2015).

In human cells, over a thousand RBPs have been identified containing diverse binding domains to regulate many species of RNA including non-coding transcripts from their synthesis to degradation (Ray *et al.*, 2013; Calabretta & Richard, 2015). Furthermore, recent

studies to systematically map RNA binding sites revealed that many of these are highly disordered and contain binding sites for post-translational modifications such as phosphorylation (Castello *et al.*, 2016). Together the RBP and its cargo RNA form highly dynamic mRNP particles, which are continually remodelled as the RNA cargo is processed. An important group of RBPs and RNA processing factors are the serine/arginine rich (SR) proteins. Different processes in gene expression are highly coupled and SR proteins are crucial at each step as they associate with a target mRNA from transcription to translation.

1.2 Pre-mRNA Splicing

Eukaryotic genomes are comprised of exons and introns, which are coding and not protein-coding respectively. On transcription, the pre-mRNA is extensively and precisely modified to remove introns and produce an exonic mRNA transcript by splicing. RNA splicing and the concept of introns was simultaneously discovered in 1977 by several groups, and rapidly identified as a ubiquitous process which has shaped the evolution of higher organisms (Berget *et al.*, 1977; Chow *et al.*, 1977; Goldberg *et al.*, 1977; Klessig, 1977). The concept of "genes in pieces" revolutionised molecular biology and was instrumental in reshaping the central dogma into the highly regulated principles regarding gene expression that are currently understood (Gilbert 1978).

Early work also identified the consensus sequences for splicing, the basic mechanisms of intron removal involving branch-site interactions and the formation of a lariat intermediate via transesterification (Gilbert, 1978; Sharp, 1985). The basis for this reaction mechanism evolved from group II self-splicing introns, which act as ribozymes with self-RNA as the catalyst (Sharp, 1985). Group II introns can adopt a tertiary structure that brings the active groups required for splicing in close proximity (Wahl *et al.*, 2009). Such class of introns is likely to be the predecessor of spliceosomal introns, which are present in all eukaryotes, albeit with different abundance. The exact evolutionary mechanisms that gave rise to spliceosomal introns remain controversial and unclear at least for higher organisms, with much historical debate between the "introns-early" and "introns-late" hypotheses. The current consensus favours neither directly, but suggests that intron invasion coincided with the endosymbiosis events that triggered the onset of eukaryogenesis (Irimia & Roy, 2014). Intron number then rapidly expanded and persisted throughout evolution to facilitate the vital genome expansion and diversification that gave rise to higher eukaryotes. Early spliceosomal introns were capable of self-splicing similar to their group II ancestors, but over time this ability was lost and eukaryotes evolved a suite of highly conserved *trans*-regulatory factors to mediate

catalysis, which became an early spliceosome (Rogozin *et al.*, 2012; Irimia & Roy, 2014). Intron gain into a eukaryotic genome has recently been demonstrated experimentally for *Saccharomyces cerevisiae* by reverse splicing reactions, suggesting that intron gain in eukaryotes is possible via this mechanism (Lee & Stevens, 2016).

For almost all splicing events to occur, the spliceosomal machinery assembles extremely proficiently in response to consensus sequences in the pre-mRNA transcript. In eukaryotes, the most important are the highly conserved 5' and 3' splice sites, corresponding to GU and AG consensus sequences respectively (Lee & Rio, 2014). The polypyrimidine tract and the branch point sequences are also significant, though less well conserved and defined by sequence alone. Throughout the splicing reaction, these consensus sites are verified at checkpoints by the spliceosome components to ensure fidelity of cleavage (Will & Lührmann, 2011). During the evolution of higher eukaryotes, conservation of the exonic region of the 5'SS has increased. For example, it is more strictly conserved in humans than yeast, which lack such complex regulation by auxiliary factors at splice sites in comparison to humans (Ast, 2004).

Despite the strong sequence requirements for canonical splicing, intron retention can occur in humans when consensus sequences are not recognised (Sibley *et al,* 2016). In addition, non-canonical splicing events can occur at a low frequency at cryptic splice sites. Such splice sites comprise a consensus akin to those at canonical splice sites, but are located far from the latter, imbedded in introns. Considering the volume of intronic sequence in vertebrate genomes, it is unsurprising that this can occur by chance (Sibley *et al*, 2016).

1.2.1 The Spliceosome and the splicing cycle

The spliceosome is the huge macromolecular machine that acts to catalyse the removal of introns and join exons to form a functional mRNA transcript. In most eukaryotes, including *S. cerevisiae*, the splicing process occurs co-transcriptionally concurrent with emergence of the nascent pre-mRNA transcript (Kornblihtt *et al.*, 2013; Wallace & Beggs, 2017). Therefore it is unsurprising that chromatin state, nucleosome positioning and associated chromatin modifiers also exhibit an interplay with splicing and can alter its outcome (Schwartz & Ast, 2010).

The spliceosome consists of multiple RNA-protein complexes termed snRNPs and a plethora of associated factors. The latter include the serine/arginine (SR) proteins which are inherent to the splicing process and form a core network of interactions within the spliceosome

(Akerman, *et al*. 2015; Long & Cáceres 2009). SR proteins are highly relevant for the work presented in this thesis and will be discussed latterly in further detail (Section 1.4 and 1.6).

1.2.1.1 Components of the Spliceosome

The components of the major spliceosome were originally isolated by glycerol gradient fractionation followed by affinity chromatography in the presence of ATP. These components were active in splicing reactions in variable combinations on a transcript specific basis (Grabowski *et al.*, 1985; Sharp *et al.*, 1987). Further work verified the existence two distinct spliceosomes in mammalian cells: the major U2-dependent spliceosome and the minor U12- dependent spliceosome, which act to remove major or minor introns, respectively (Will & Lührmann, 2011).

The minor intron class are scarce (<0.5% of all vertebrate introns) and were not discovered until nearly two decades after major introns. Contrary to original findings describing the AT/AC consensus sequence (after which they are sometimes named), it was later established that minor introns require a longer and more strictly constrained nucleotide stretch at the 5'ss and strict branch point. The minor spliceosome is absent from lower eukaryotes and composed of alternate spliceosome components: U11, U12, U4atac and U6atac, and U5, which is also present in the major spliceosome (reviewed in: Patel & Steitz 2003; Turunen *et al*. 2013). Although it mainly functions to remove these specific, rare introns, it can also remove those spliced by the major spliceosome (Rogozin *et al.*, 2012). Incidentally, the last eukaryotic common ancestor (LECA) contained both core and auxiliary factors (including SR proteins) for both major and minor spliceosomal machineries (Irimia & Roy, 2014). Despite this, ancestral spliceosomal introns were most likely of the major type, although U12 introns do exist in lower organisms (Rogozin *et al.*, 2012).

The core machinery of the major U2-dependent spliceosome consists of five snRNP particles U1, U2, U4, U5 and U6. Each of these has both an RNA component, termed snRNA, and a proteinaceous scaffold (Wahl *et al.*, 2009). The mechanism for pre-mRNA splicing has been highly characterised into a complex cycle (Figure 1.2) that requires over 150 proteins in two catalytic steps (Papasaikas & Valcarcel 2016; Zhou *et al*. 2002). Dynamic interaction networks of such protein (and RNA) facilitators are central to spliceosome function and drive transition between different spliceosomal complexes (Hegele *et al.*, 2012).

Aside the snRNPs and the SR protein family in mammals (discussed later), there are numerous auxiliary factors necessary for splicing to occur. In budding yeast these have been

extensively characterised and some act as non-energy dependent moieties serving as scaffolding or recruiter proteins; although a significant proportion are RNA helicases of the SF2 superfamily (Cordin & Beggs 2013; Wahl & Lührmann 2015). Eight of such helicases, named Sub2, Prp5, Prp28, Brr2, Prp2, Prp16, Prp22 and Prp43 in yeast, are essential in eukaryotes and act to monitor splicing fidelity at various splicing checkpoints and continually remodel the spliceosome throughout the cycle (Figure 1.2) (Cordin & Beggs, 2013; Bourgeois *et al.*, 2016).

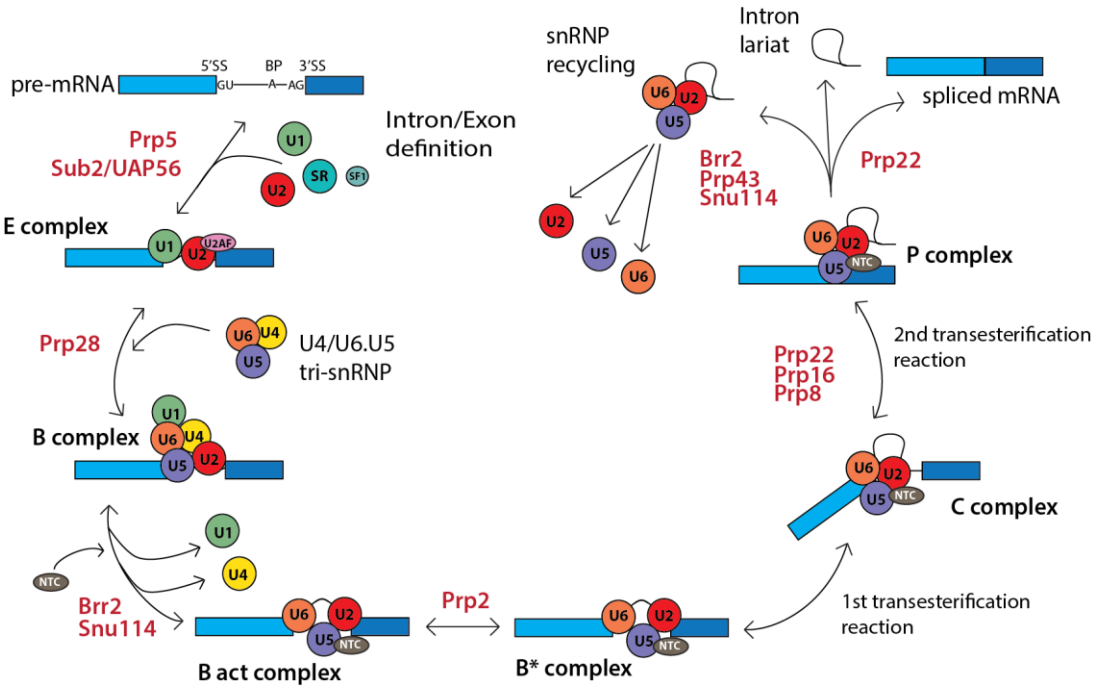


Figure 1.2: The Splicing Cycle. The nascent pre-mRNA is immediately bound by RBPs that initiate co-transcriptional splicing. Intron/exon boundaries are defined by a multitude of factors including SR proteins. The U1/U2 snRNP is recruited to drive intron removal through downstream, dynamic interactions mediated by the U4.U5/U6 tri-snRNP and various RNA helicases (indicated in red for *S. cerevisiae*). At the end of the process, snRNPs are released from the spliced transcript and recycled for the next round of splicing. The different complex formations that determine splicing-stage are also illustrated. *Adapted from* (Lee & Rio, 2014; Matera & Wang, 2014; Papasaikas & Valcárcel, 2016).

1.2.1.2 Spliceosome Assembly and Mechanistic Outline

Biochemically, pre-mRNA splicing is described by two sequential transesterification reactions. The first step of splicing is initiated when the intronic branch point adenine directs nucleophilic attack via its 2'OH group to the 5' phosphate group of the intronic guanine nucleotide of the 5'ss. This generates a lariat intermediate via a 2'-5' phosphodiester bond. Consequently, the free 3'OH at the 5'ss launches the second step by nucleophilic attack on

the 3'ss exonic phosphate group which simultaneously joins the exons to release the lariat intermediate.

The splicing process is highly dependent on at least two magnesium ions (Mg^{2+}) which, stabilise both the leaving and attacking reaction states. This is coordinated by the U2/U5/U6 snRNA, hence the U2/U5/U6 snRNPs are vital to catalyse the splicing reaction. (Hang *et al.* 2015; Papasaikas & Valcarcel 2016). U5 snRNP acts as a crucial molecular scaffold and is the only snRNP indispensable for both major and minor spliceosome function (Hang *et al.*, 2015). Other metal ions are required for each step of catalysis in both humans and yeast, during which they bind to the U6 snRNA in a conformation highly reminiscent of the ancestral group II introns. Indeed, similarly to the latter, RNA interactions (U6 snRNA) are ultimately required for catalysis (Fica *et al.*, 2013).

Spliceosome assembly is initiated when the U1 snRNA of the U1 snRNP forms base pairs with the 5'ss. In higher eukaryotes, the presence of a phosphorylated SR protein, most likely bound to a regulatory sequence in the transcript, is crucial for the U1 complex to remain stable (Long and Cáceres, 2009; Wahl, Will and Lührmann, 2009). To form the E complex, the two subunits of U2AF, U2AF1 and U2AF2 bind to the polypyrimidine tract and the AG of the 3'ss respectively. Concurrent with this, the branch point just upstream is bound by the Branchpoint Binding Protein (BBP) (Matera & Wang, 2014). This drives recruitment of the U2 snRNP and thus forms the complete E complex. Single molecule studies indicate that U2AF1/2 and the U2 snRNP exist in stoichiometry with each other in a pre-A complex, known as the I complex, formation of which expels excess factor binding at the 3'ss and triggers remodelling into the A complex in an ATP-dependent mechanism (Chen *et al.*, 2016). The E complex is committed to splicing, although splice site selection may not be fully decided until later in the cycle (Matlin *et al*. 2005; Chen *et al*. 2016).

On assembly of the pre-spliceosomal A complex, the U4/U6.U5 tri-snRNP is recruited in a pre-assembled state to trigger transition to the catalytically inactive B complex. Disruption of U1 binding at the 5'ss occurs and is replaced by the U6 snRNA which binds in its place. Both events, and thus RNP remodelling, are mediated by the helicase Prp28. The U2 snRNP is necessary for tri-snRNP association with which it directly interacts at two interfaces as shown by the cryo-EM structure in yeast (Plaschka *et al.*, 2017). For activity in splicing, the B complex must undergo significant remodelling including removal of the U1 and U2 snRNPs and recruitment of the essential factors Prp38, Snu23 and Spp381. The latter interacts with Brr2 which is the only helicase of the Ski2-like DEAH-box subfamily required for splicing, as it is responsible for ATP-dependent dissociation of the U4 snRNA and associated

factors (Cordin & Beggs, 2013). Remodelling continues as the U2 and U6 snRNA form a helix that ultimately recruits NTC and NTR (nineteen complex and the NTC-related complex respectively) proteins for activation into its catalytically active state (Plaschka *et al.*, 2017).

Once activated the B complex performs the first catalytic step of splicing and transitions to the C complex which securely holds both the intron-exon lariat intermediate and the excised exon (Matera & Wang 2014; Bertram *et al*. 2017). This is activated into the C* complex for the second catalytic step by Prp8 and Prp16, which are necessary to position the 3' splice site close to the lariat for the nucleophilic attack. The second step of splicing in humans is dependent on Prp22, Prp18 and Slu7 proteins. Prp22 is bound among other factors to Prp8 and acts indirectly to hold the 3'ss prior to the second step of splicing after which it facilitates release of the nascent transcript by binding to the newly spliced 3' exon (Zhang *et al.* 2017; Bertram *et al.* 2017). After the second catalytic step, the components of the spliceosome are released from the exon-ligated transcript and recycled for subsequent rounds of splicing. It is interesting to note that all of the steps of the splicing cycle are reversible, at least *in vitro* (Matera & Wang, 2014).

The Exon Junction Complex (EJC) is deposited on the mRNA 24 nucleotides upstream of the newly ligated exon boundary to prepare it for nuclear export, however, the complex is also capable of non-canonical binding along the length of the transcript (Le Hir *et al.*, 2001; Saulière *et al.*, 2012). The EJC is a dynamic and multicomponent protein complex that consists of the key factors: eIF4AIII, Y14, MLN51 and Magoh, although numerous other components are known to comprise the peripheral EJC in a context-dependent manner. Knockdown of the components of the EJC has transcriptome wide ranging effects on splicing and downstream regulatory processes (Wang *et al.*, 2014). Strikingly in the C* complex, the four components of the exon junction complex (EJC) eIF4AIII, MAGOH, Y14 and MLN51 are already present on the spliced transcript (Bertram *et al.*, 2017; Zhang *et al.*, 2017).

Throughout the splicing process, the nascent mRNA is continually modified before nuclear export, to prepare it for mRNA translation once it reaches the cytoplasm. On transcription, nascent pre-mRNA is modified through the addition of a 7-methlyguanosine cap to the 5'end of the transcript. This is directed by a suit of regulatory factors including the methyltransferase RNMT and the guanylyltransferase RNGT, and bound by the cap binding complex (composed of CBP80/20 in humans) (Cowling, 2010). Also concurrent with transcription is 3' end formation and polyadenlylation of the pre-mRNA transcript which marks the mRNA for downstream processing (Singh *et al.*, 2013).

1.3 Splicing Regulation

Analogous to other RNA processing events, the regulation of splicing occurs in a co-transcriptional manner and is heavily reliant on the function of RBPs at each stage of pre-mRNA synthesis (Bentley, 2014). Indeed, splicing factors such as the SR protein SRSF2, and those involved in capping are recruited by the elongating polymerase during transcription (Bentley, 2014).

RNA processing, including splicing, can also be influenced by dynamic post transcriptional modifications of RNA, termed the epitranscriptome (Jiang *et al.*, 2017). There is a huge diversity of RNA modifications that occur, one of the best studied being N^6-methyladenosine (m^6A). m^6A is present on between 0.2-0.6% of mRNAs and is deposited by the METTL family of methyltransferases (Roundtree *et al.*, 2017). Binding sites for these enzymes are enriched in introns, consistent with a role for m^6A in recruiting both the splicing machinery and factors for capping and polyadenylation of the mRNA transcript. Furthermore, m^6A can function to promote export of the mRNA in which it resides (Zhao *et al.*, 2016). In general, this mark serves to increase the processing of a transcript, including its translation in the cytoplasm (Zhao *et al.*, 2016).

m^6A was shown to have a specific role in splicing regulation, as its reader, YTHDC1, is also a splicing factor. Remarkably, YTHDC1 is capable of differentially manipulating two SR proteins to exert adverse functions. It recruits SRSF3 to promote exon inclusion, which prevents SFS10 binding to function as a splicing repressor (Xiao *et al.*, 2016). The m^6A modification also has important post-transcriptional implications to regulate developmental processes and tissue-specific RNA processing events, for example splicing of the FTO gene during adipogenesis via SRSF2 binding (Zhao *et al.*, 2014).

1.3.1 Alternative splicing

Alternative splicing occurs in approximately 95% of genes containing multiple exons, to create an unprecedented degree of proteome diversity and regulation (Figure 1.3). Such complexity is further increased by differential alternative splicing programmes that have evolved often due to changes in splicing factors present or their site usage in distinct tissues and cell types (Pan *et al.*, 2008). For example in neural lineages alternative splicing is extensively used to shape neurogenesis, in particular the splicing factors nSR100/SRRM4 and PTBP1 are capable of antagonising one another (Vuong *et al.*, 2016). nSR100/SRRM4 is a brain specific SR-like protein involved in splicing of microexons; highly conserved 3-27nt

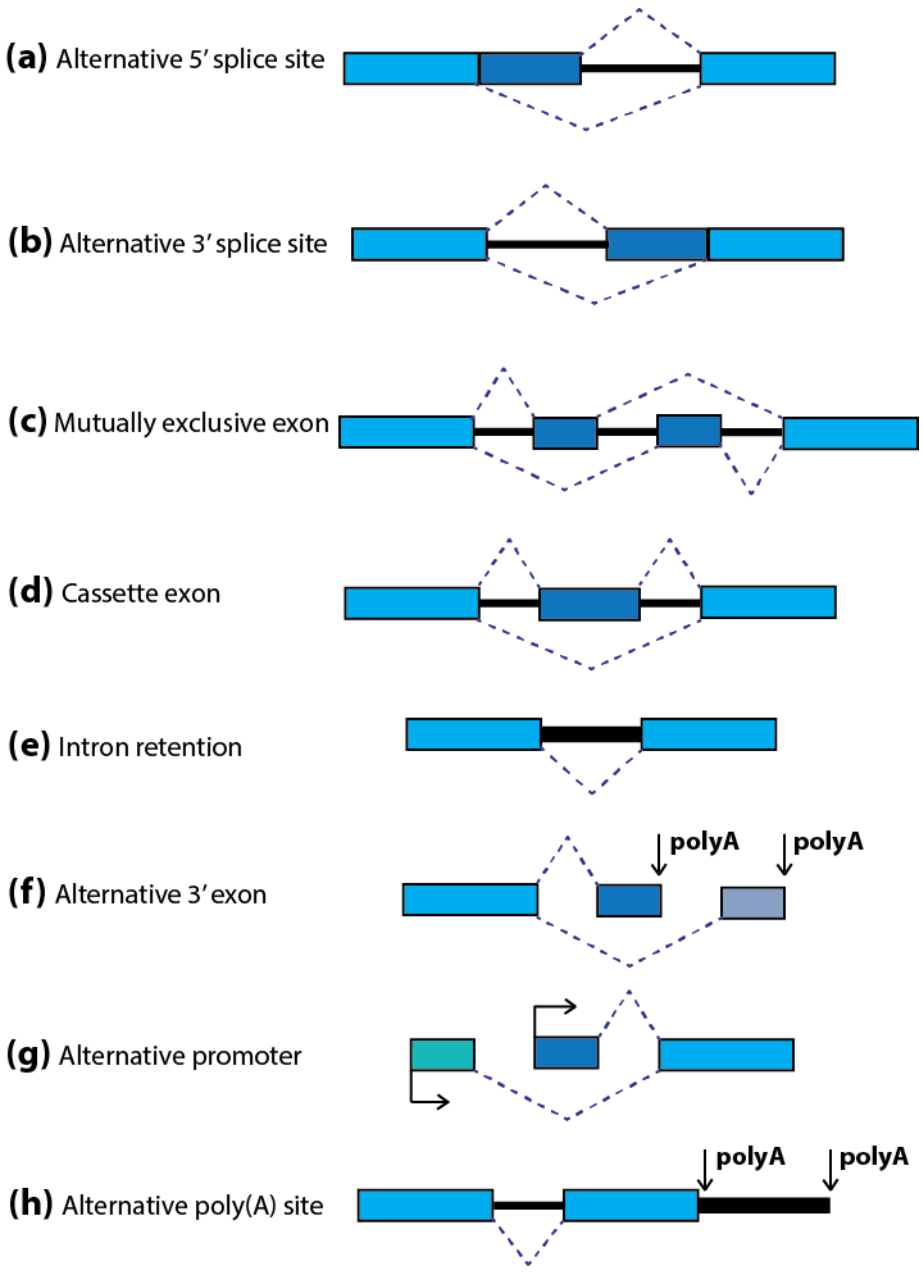


Figure 1.3: Modes of alternative splicing. The pre-mRNA transcript can be customised during splicing to promote the formation of countless alternative isoforms. Schematics of common alternative splicing patterns are shown (a)-(h). Large boxes (various shades of blue) represent exons, introns are represented by thick black straight lines between exons. Dark blue dashed lines indicate alternative routes that exons can be joined in the spliced mRNA. *Adapted from* (Vuong *et al*. 2016).

cassette exons, enriched in neural lineages. Modest downregulation of this factor causes an autism-like phenotype in mice models (Quesnel-Vallières *et al.*, 2016).. Furthermore, alternative splicing factors themselves can be alternatively spliced to causes changes in their

function. For instance, the PTBP1 transcript has an alternative exon 9, which can be selectively used or skipped to alter the regulatory capacity of PTBP1 protein. (Gueroussov *et al.*, 2015). This not only demonstrates the subtle nature of splicing regulation and splicing factor dosage, but also how perturbing these processes can have wide-ranging phenotypic consequences despite being dispensable for viability

Alternative splicing produces numerous mRNA isoforms, however there is debate as to whether this equates to protein isoform diversity (Tress *et al.*, 2017). On the other hand Genome Wide Association Studies (GWAS) have found that the non-coding regions of the genome, including splice sites, are highly heritable and can be linked to specific genetic traits. For example, a study conducted to investigate polymorphic transcript variation in human disease identified that splice patterns are often associated with SNPs that, particularly in B-cells, correlate with high prevalence of autoimmune diseases due to improper alternative splicing or polyadenylation (Fraser & Xie, 2009). This alludes to the importance regulatory and intronic sequence regardless of its protein-coding capacity (Manning & Cooper 2017).

1.4 SR Proteins

SR proteins act to regulate each step of gene expression and have concerted nuclear function in transcription, splicing and alternative splicing. Intron removal during both alternative and canonical splicing is coordinated by SR and hnRNP proteins which bind to regulatory motifs known as intronic or exonic splicing enhancers or silencers (ISE, ESE, ISS or ESS respectively) to antagonise one another's actions (Ast, 2004; Goren *et al.*, 2006). Once bound SR and hnRNP proteins mediate bridging interactions of required factors for initiation of that splicing reaction.

In intron definition, SR proteins bind to an intronic sequence to facilitate U1 binding to the upstream donor splice site and U2 to the downstream site. In exon definition, SR proteins bind to an exonic sequence to recruit U1 and U2/U2AF, and coordinate these factors across the exon (Rogozin *et al.*, 2012) (Figure 1.5). SR-like proteins are also required for both canonical and alternative splicing (Wahl *et al.*, 2009). hnRNP proteins represent the most diverse class of RNA binding proteins, and often act to antagonise the activity of SR proteins (Long and Cáceres, 2009). They were first identified as dynamic cellular components that consisted of multiple proteins and later found to crosslink to poly(A) containing mRNAs (Choi & Dreyfuss, 1984; Choi *et al.*, 1986).

1.4.1 Structure and nomenclature

SR proteins are ubiquitous in higher organisms and present in many other diverse species, such as plants and fungi, reflecting an early emergence in evolutionary history (Twyffels *et al.*, 2011). Budding yeast is devoid of true SR proteins but has three SR-like proteins: Npl3, Gbp2 and Hrb1, likely to represent the ancestral basis from which SR proteins subsequently evolved. These have numerous functions in gene expression; a feature conserved to their metazoan counterparts, indicating the level of importance of such regulation for cells (Martínez-Lumbreras *et al.*, 2016). In *Caenorhabditis elegans* there are seven SR homologues, depletion of which results in numerous phenotypic outcomes including lethality for the SRSF1 homologue (Longman *et al.*, 2000).

Before its classification as an SR protein, SRSF1 (then SF2) was identified within a group of multiple factors, that were capable of complementing mRNA splicing in splicing deficient extracts *in vitro* (Krainer & Maniatis, 1985). SR proteins were characterised as a family using, among other methods, the monoclonal antibody 104 (m104) which binds to the phosphoepitope of phosphorylated SR proteins (Roth *et al.*, 1990) and later demonstrated to be essential for splicing and alternative splicing (Krainer *et al.*, 1990; Zahler *et al.*, 1993).

SR proteins are defined by the presence of one or two RNA recognition motifs (RRMs) and an RS domain composed of serine/arginine repeats. RRM domains are widespread among RNA binding proteins and their flexibility permits diverse modes of binding between proteins (Maris *et al.*, 2005). The SR family nomenclature is historically diverse leading to the agreement of a consensus terminology to identify them (Figure 1.4) (Manley & Krainer, 2010; Twyffels *et al.*, 2011). RS repeats must comprise 40% of at least a 50 amino acid region in the RS domain for classification as an SR protein (Twyffels *et al.*, 2011). For many SR proteins including SRSF1 and SRSF2, the RS domain acts as a nuclear localisation signal and is required for nuclear import through a specific importin-β family receptor named transportin-SR (transportin-3, TNPO3), to which it binds directly (Cáceres *et al.*, 1997; Kataoka *et al.*, 1999). Transition into the nucleus through the nuclear pore in this mechanism is dependent on RS domain phosphorylation, which is also sufficient to act as a nuclear localisation signal for SRSF1 (Lai *et al.*, 2001; Hamelberg *et al.*, 2007).

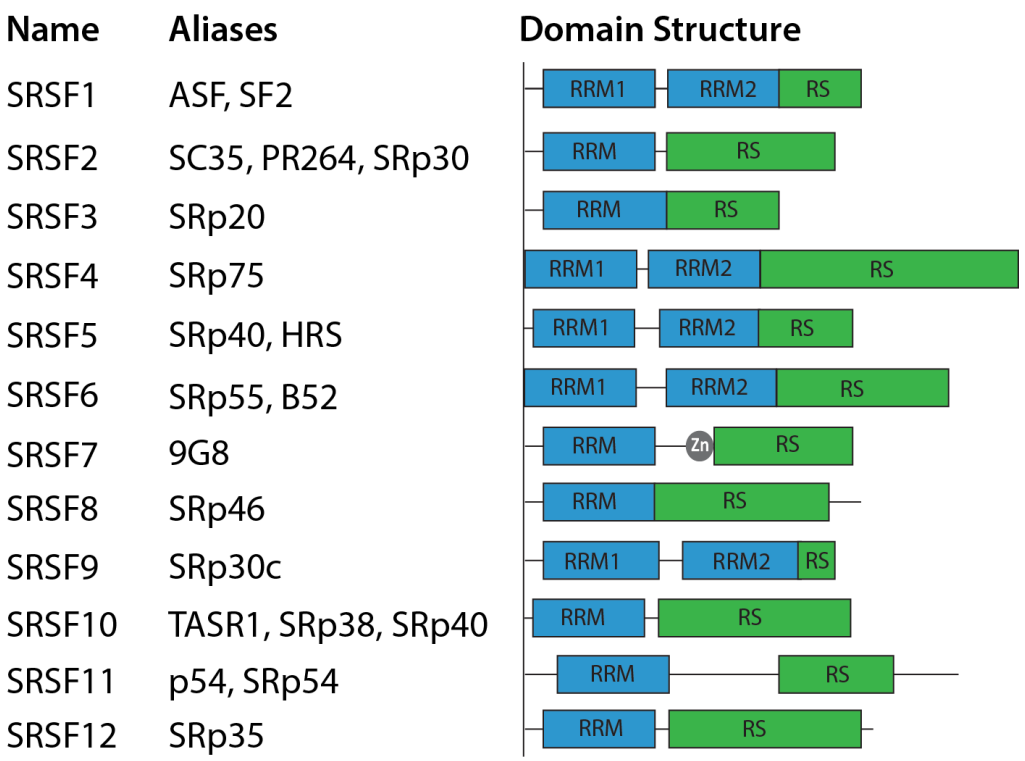


Figure 1.4: SR protein structure and nomenclature. SR proteins have a modular domain structure defined by one or two RRM domains at the N terminus (blue) and a C-terminal RS domain (green). Thin black lines indicate linker sequence that connects each domain. Current nomenclature is stated in the far-left column, followed by the previously used nomenclature for each SR protein. *Adapted from* (Manley & Krainer, 2010; Twyffels *et al.*, 2011).

SR-like proteins also exist, such as Tra2α/β, which contain an RS domain but lack the highly repetitive characteristic of true SR proteins or an RRM motif, although many of these are able to bind RNA via different domains and mechanisms (Long and Cáceres, 2009). Genome-wide analyses revealed that RS-domain containing proteins participate in numerous aspects of gene expression regulation, including those involved in the cell cycle (Boucher *et al.*, 2001).

1.4.2 Shuttling SR proteins

In humans, of the twelve canonical SR proteins, a subset exhibits nucleocytoplasmic shuttling and have numerous other post splicing and cytoplasmic roles (Long and Cáceres, 2009). The shuttling SR proteins were originally defined as: SRSF1, SRSF3, SRSF4, SRSF7, and SRSF10 in HeLa cells (Cáceres *et al.*, 1998). SRSF1 is the prototype for shuttling SR proteins and exhibits dynamic movement between cellular and sub-cellular compartments

(Cáceres *et al.*, 1998). This was determined by heterokaryon assays, which are the gold standard for determining shuttling capacity of a protein (Piñol-Roma & Dreyfuss, 1992).

More recently, other SR proteins were demonstrated to shuttle, such as SRSF9 (Sapra *et al.*, 2009; Fu *et al.*, 2013). In contrast, SRSF2 was consistently found to be a non-shuttling SR protein due to the presence of a nuclear retention sequence (NRS) at its C-terminus (Cazalla *et al.*, 2002). SR proteins exhibit differential shuttling abilities in differentiated and pluripotent cells. For example, heterokaryon assays to detect endogenous fluorescence demonstrated that all SR proteins are able to shuttle in pluripotent mouse P19 cells, whereas only SRSF1 always shuttles regardless of developmental stage or tissue type (Botti *et al.*, 2017). At a steady state in HeLa cells, SRSF1 is predominantly found in the nucleus and nuclear speckles, reflecting its roles as an essential factor for splicing (Misteli *et al.*, 1997). SRSF1 is highly relevant to this work and discussed in detail in section 1.6.

1.4.3 SR proteins in splicing and alternative splicing

SR proteins are the master regulators of both canonical and alternative splicing and they are essential for both processes (Figure 1.5). They are required both for initial spliceosome assembly and in downstream processes such as cleavage and selection of the 5' splice site (Krainer *et al.*, 1990). Early work also demonstrated that SR proteins are required in alternative splicing for definition 5'splice sites and could influence splice site choice (Ge *et al*. 1991, Krainer *et al*. 1991). Consistent with this, SR proteins are present in purified extracts of the spliceosome and are required for splicing *in vivo* (Zhou *et al.*, 2002; Ellis *et al.*, 2008) and act to mediate bridging interactions between the 5' and 3' splice sites (Long and Cáceres, 2009). In vertebrate genomes that contain a vast quantity of intronic sequence, this is particularly important to prevent potential cryptic splicing or mis-splicing events (Sibley *et al.*, 2016).

In canonical splicing SR proteins are required to recruit the U1 and U2 snRNPs to the required splice sites and initiate splicing of the transcript. Once bound, the RS domain of the SR protein transmits a multifaceted signal by contacting the RS domain in U2AF2, the branchpoint sequence and the 5' splice sites in the target mRNA (Shen & Green, 2004). Within nuclear speckles, SRSF1 has been shown to simultaneously bind U2AF1 and U1 thus bridging the 5' and 3' splice sites (Ellis *et al.*, 2008). SR proteins also act later in the splicing cycle to recruit the U4/U6.U5 tri-snRNP for incorporation into the spliceosome (Roscigno & Garcia-Blanco, 1995). In addition, SR proteins are able to mediate splicing of minor U12-introns by

functioning in an analogous manner to recruit the machineries of the minor spliceosome through interactions with the U5 snRNP (Shen & Green, 2007).

In alternative splicing, SR proteins act in trans to promote both exon inclusion and exclusion on a similar measure (Pandit *et al.*, 2013). They bind ESE/ESS and ISE/ISS cis-regulatory sequences in the target mRNA, which are between 4 and 18 nucleotides in length. The positioning of these sequences within the transcript is crucial in determining the regulation they exert, which offers some explanation to SR proteins diverse splicing roles (Goren *et al.*, 2006).

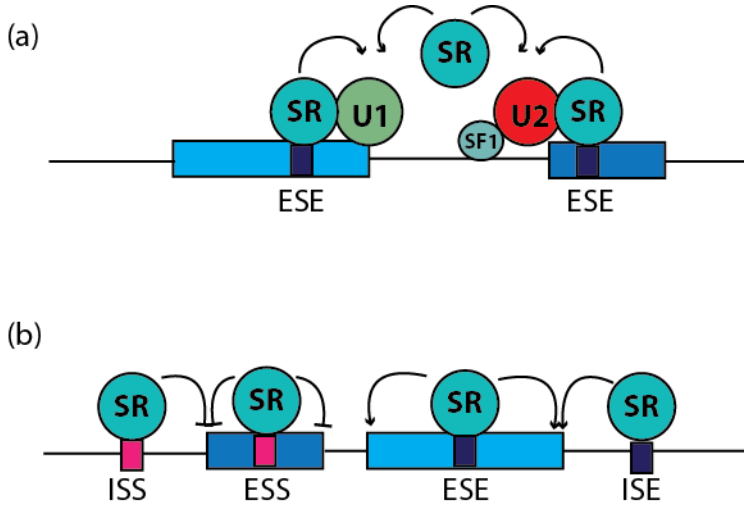


Figure 1.5: SR proteins in splicing. (a) SR proteins bind to regulatory enhancer sequences in the pre-mRNA transcript such as ESE/ESS (b) SR proteins can also act to antagonise exon or intron inclusion through binding silencing sites such as ISS/ESS.

Different SR family members are capable of degenerate binding, so have a natural degree of functional redundancy *in vitro*. Substrate binding is predominantly dictated by the RRM motifs of the SR protein, discussed later in the context of SRSF1. In canonical splicing, the RS domain is not required for substrate specificity, although its ablation can prevent alternative splicing function. Consistent with this, SR protein definition mediated by the RS domain is required for spliceosome recruitment and correct splicing at weak 3' splice sites (Zhu & Krainer, 2000). SRSF10 is distinct from the other family members as it can act as a splicing repressor on dephosphorylation by preventing proper U1 snRNP binding; an action predominant during mitosis and heat shock (Shin *et al.*, 2005).

Alternative splicing is important for accurate control of the cell cycle and occurs in a tightly regulated fashion to regulate specific transcripts periodically. SR protein

phosphorylation is altered during the cell cycle and after DNA damage. The SR protein kinase CLK1 is a key factor in this and its own expression oscillates during the cell cycle in a negative feedback loop, among other transcripts encoding for mitotic and associated factors (Dominguez *et al*. 2016). SRSF1 also controls a cell cycle specific alternative splicing programme, which on its knockdown, causes apoptosis and cell cycle dysregulation. Interestingly, co-depletion of SRSF1 and the mitotic regulatory protein aurora kinase A have synergistic effects in mediating splicing outcome of certain transcripts (Moore *et al.*, 2010).

The EJC is known to physically interact with SR proteins, particularly shuttling SR proteins such as SRSF1, and the two exist in stoichiometry within the mRNPs (Singh *et al.*, 2012). In addition to this, CLIP-seq revealed that the EJC binding consensus is highly similar to that of SR proteins, alluding to their spatial proximity on a transcript (Saulière *et al.*, 2012).

The splicing factor machineries, including SR proteins, are enriched at sites of active transcription although transcription is not required for their presence. SR proteins are also abundant in nuclear speckles which supply factors for transcription; when transcription is blocked, SR proteins relocate to nuclear speckles (Ellis *et al.*, 2008). Speckle components are distributed to the correct sites on demand allowing the cell to be constantly primed for on-demand splicing requirements. They can act to stabilise the nascent transcript and recruit the splicing machinery to facilitate pre-mRNA processing (Misteli *et al.*, 1997).

The SR protein SRSF2 has a direct and unique role in transcriptional elongation, and its depletion in liver tissue alters the elongation rate of RNA Polymerase II, causing it to stall at the gene body due to defective downstream transcript processing (Lin *et al.*, 2008). SRSF1 also able to couple transcriptional and alternative splicing regulation of the same transcript. For example, in T cells, SRSF1 binds to the promoter of CD3ζ and acts to increase its transcription in a dose-dependent manner as well as regulating alternative splicing of the CD3ζ mRNA through binding the 3'UTR (Moulton *et al.*, 2015). This highlights the intrinsic coupling of splicing to transcription and indicates the multiple layers of control opportunities for SR protein mediated regulation.

1.4.3.1 SRSF1 as splicing factor

SRSF1 acts to control splicing and alternative splicing as described for SR proteins in general. Indeed, many studies that defined such general mechanisms for the SR family are based on SRSF1 datasets. Further detailed information on the SR protein family is described in the context of SRSF1 in Section 1.6 of this thesis.

SRSF1 function in splicing is highly dependent on the dynamic phosphorylation of its RS domain, which modulates interactions with both the U1 snRNP and the ESE site. When SRSF1 is phosphorylated, its RRM can bind to the RRM of U1-70K to mediate spliceosomal E complex formation. However as the SRSF1 RS domain is progressively dephosphorylated, these RRM-RRM interactions are disrupted as SRSF1 preferentially binds to itself (Cho *et al.*, 2011).

CLIP-seq and functional SELEX data determined the SRSF1 RNA binding motif in exons as the GA-rich sequence GGAGA (Pandit *et al.*, 2013). SRSF1 is capable of binding to exons, introns and other non-coding RNA sequences in numerous cellular compartments. However, the consensus binding motifs for non-coding sequence, including intronic sequence, cannot be defined, at least from CLIP-seq datasets. Interestingly, only a small subset of CLIP-seq identified RNA binding targets were annotated as being products of SRSF1-mediated alternative splicing, indicating that SRSF1 binding has regulatory functions other than splicing (Sanford *et al.*, 2008). Despite this, genome wide studies to identify the effect of single nucleotide variant (SNV) genetic regulation on allele-specific alternative splicing events identified SRSF1 as one of the most influential factors, in part due to a strong consensus binding motif (Hsiao *et al* 2016). This study married *in vivo* and ENCODE data to demonstrate that SRSF1 ESE binding sites are both highly conserved and most frequently associated with SNVs that result in altered splicing. A proportion of these were in linkage disequilibrium with disease-associated loci, which eludes to their significance in maintaining proper splicing.

1.4.4 RNA-binding capacities of SR proteins

Different SR proteins are coordinated to bind transcripts in specific, mRNA dependent combinations (Björk *et al.*, 2009). All SR proteins are RNA and mRNA binding proteins as shown by PAR-CLIP coupled to quantitative mass spectrometry, which also determined the regions on the mRNAs to which the protein binds (Baltz *et al.*, 2012). There is some functional redundancy between SR proteins in splicing; however, they also play unique and essential roles in substrate specificity (Sanford *et al.*, 2009).

The RNA-binding landscape between SR proteins is highly complex and differs for each SR-family member, most likely in a context-dependent manner (Pandit *et al.*, 2013). There are various CLIP-seq datasets for SRSF1, SRSF2, SRSF3 and SRSF4, which provide a wealth of knowledge regarding their RNA-binding landscapes including consensus binding motifs (Sanford *et al.*, 2009; Änkö *et al.*, 2012). Overall, these data suggest that SR protein binding occurs over the entire length of the transcript including the UTRs and intronic

sequences. Although, in accordance with binding to ESE/ESS during splicing, SR proteins show a slight preference for binding to exons. The binding consensus motifs for each SR protein are different, which is consistent with the lack of complete functional redundancy between family members. For example, SRSF3 and SRSF4, the smallest and largest SR proteins respectively, have distinct mRNA binding motifs (CU-rich and GA-rich respectively). Accordingly, they target different RNA transcripts within distinct mRNP particles, indicating a lack of binding redundancy *in vivo* (Änkö *et al.*, 2010, 2012). Both proteins bind to the branch point sequence indicating a specific binding role at this region of the transcript during splicing (Änkö *et al.*, 2012). However, SRSF4 binds significantly fewer target RNAs than SRSF3, suggesting that not only do SR proteins have distinct targets, but also the degree to which they bind them is also disparate. SRSF1 and SRSF2 binding patterns are similar as they both preferentially bind exons, which is not wholly reflective of their function, as binding per se does not dictate alternative splicing patterns. SRSF1 and SRSF2 also have distinct binding profiles which are not functionally redundant, and perhaps most interestingly, they are able to co-regulate and coordinate other classes of splicing events (Pandit *et al.*, 2013).

1.4.5 SR proteins during development and differentiation

SR protein expression *in vivo* varies between tissues in a pattern highly analogous to the antagonistic hnRNP proteins and the expression ratio between them is developmentally important (Hanamura *et al.*, 1998). In *C.elegans* development, the homologue of SRSF1 is the only SR protein whose depletion by RNAi results in early embryonic development, whereas depletion of other SR proteins are viable (Longman *et al.*, 2000). Phosphorylation of SR proteins is a crucial requirement for development, as RNAi of SRPK1 is lethal (Longman *et al.*, 2000). Consistent with this, SR protein phosphorylation is also crucial for *Ascaris lumbricoides* (nematode roundworm) development. In very early embryos SR proteins are hyperphosphorylated but during the four to eight cell transition they become increasingly dephosphorylated. This coincides with zygotic gene activation (ZGA); marked by the onset of zygotic transcription, splicing and nuclear import of SR proteins. Dephosphorylation of the RS domain is required for such movement and necessary for *A. lumbricoides* development (Sanford & Bruzik, 2001).

In DT40 chicken cells SRSF1 is essential for viability and exerts unique function as its knockdown cannot be rescued by expression of other SR proteins, despite such close structural homology. RS domain substitution mutants demonstrated that individual SR protein

function in viability is not affected when this domain is disrupted (Wang *et al.*, 1998). In *C.elegans*, SR proteins appear to show some degree of functional redundancy, where only SRSF1 is essential for viability (Longman *et al.*, 2000). Consistent with this, SR protein genes are essential in vertebrates and ablation of SRSF1, SRSF2 or SRSF3 in early development causes embryonic lethality in mice (Wang *et al.*, 1998; Jumaa *et al.*, 1999; Xu *et al.*, 2005). In mouse models, homozygous deletion of SRSF1 results in embryonic lethality prior to E7.5. Interestingly, the conditional knockout in mature mouse cardiomyocytes is viable, indicating that SRSF1 is not required in these cells, perhaps through compensation of function by another SR protein (Xu *et al.*, 2005).

The homozygous SRSF2 mutation is lethal before E7.5, although heterozygotes are viable at the expected Mendelian ratio (Ding *et al.*, 2004). SRSF3 is required for viability after E3.5; although prior to this point maternal protein/mRNA is responsible for development. It is likely that during development, SRSF3 is absolutely required in a non-redundant manner (Jumaa *et al.*, 1999; Möröy & Heyd, 2007). In conditional knockout tissues of SRSF1 and SRSF2, loss of specific splicing events can cause lethality as seen with the conditional SRSF1 knockout in the heart, that causes mice to die two months after birth due to heart failure, as SRSF1 is responsible for propagation of the correct CaMKIIdδ splice isoform (Xu *et al.*, 2005).

In contrast, the SRSF2 heart specific knockout is viable although the mice develop a deleterious phenotype reminiscent of dilated cardio myopathy (Ding *et al.*, 2004). Interestingly it is not required in terminally differentiated postmitotic cells such as cardiomyocytes (Xiao *et al.*, 2007). In addition, SRSF2 is also required for T cell development, during which it regulates the crucial alternative splicing of the receptor tyrosine kinase CD45 and in pituitary gland development during which it regulates cell proliferation (Wang *et al.*, 2001; Xiao *et al.*, 2007). In the liver, SRSF2 is required to maintain metabolic homoeostasis through both transcriptional and alternative splicing regulation of numerous tissue-specific transcription factors. The liver specific conditional knockout of SRSF2 causes liver failure resulting in early death (Cheng *et al.*, 2016). SRSF3 is also required for alternative splicing in liver homeostasis and although conditional knockout mice display gross hepatic abnormalities, they are viable (Sen *et al.*, 2013). Interestingly, SRSF3 regulates distinct targets from SRSF2 in this tissue such as cell cycle regulators. In stark contrast, the reciprocal deletion for SRSF1 has no effect on liver function or lifespan. (Cheng *et al.*, 2016). Taken together, these studies are particularly interesting as they highlight the functional individuality of SR proteins both in splicing and other aspects of gene expression.

During neural differentiation of multipotent P19 mouse cells, SR protein abundance is uniform, except for SRSF3 and SRSF4 that show increased expression. In this system, the mRNAs with which each SR protein associates are also distinct. SRSF3 and SRSF4 regulate mRNA transcripts that encode different functional groups raising the intriguing possibility that each SR protein regulates functional subsets of mRNA targets upon differentiation (Änkö *et al.*, 2010).

Mouse and conditional knockout models indicate that each SR protein has a very specific repertoire of splicing targets in different tissues during development, which are exclusive to the cell type and developmental stage. In addition, gross canonical and alternative splicing in these models remain unaffected (Möröy & Heyd, 2007). This raises interesting questions regarding SR proteins during development and the relevance of their splicing function *in vivo*. Considering that several SR proteins have numerous functions other than those in splicing, it is tempting to speculate that these could have an important physiological role during development that was not investigated in previous studies.

1.5 mRNA Translation

Once mRNP particles exit the nucleus via the nuclear pore, their cargo mRNA is translated into a polypeptide chain. mRNA translation is a highly complex process, which is tightly regulated during events such as lineage differentiation or certain stress responses, both on a global and transcript-specific level (Gebauer & Hentze, 2004; Sonenberg & Hinnebusch, 2009). RBPs play a crucial role in all aspects of mRNA translation including binding to and regulating the assembly of ribosomal subunits (Wurth & Gebauer, 2014).

Throughout evolution of higher organisms, ribosome size has increased to provide an interactive landing site for auxiliary factors and facilitate customised mRNA translational landscapes (Simsek *et al.*, 2017). Many of these factors are not part of the canonical translation machinery for example cell cycle, metabolic and RNA processing proteins. Furthermore, they can be categorised as direct or indirect binders of ribosomes, including those that require RNA for the interaction and those that are specific for each ribosomal subunit (Simsek *et al.*, 2017). The entire translation process is dependent on the presence of sufficient cofactors and an active tRNA pool (Schmidt *et al.*, 2016). Due to the immense size and dynamics of the ribosome and its cofactor repertoire, gaining structural insight into ribosome assembly is not a small undertaking. However numerous structures of the ribosome particle in various conformations

have been solved which provides insight into the molecular mechanisms of its function (reviewed in (Aitken & Lorsch, 2012; Hinnebusch, 2017)).

In short, mRNA translation can be summarised into three phases: initiation, elongation and termination, all of which rely on the presence of the 80S ribosome, which, after several processing operations, is fully assembled at the start codon and poised for peptide synthesis. At each stage of the process, with an emphasis on initiation, there is a high degree of both internal and external regulation by various RNA binding proteins.

1.5.1 Translation initiation, elongation and termination

Cap-dependent translation initiation commences after two rate limiting steps: the association of the small 40S ribosome subunit and multiple cofactors, to form the 43S preinitiation particle and the formation of the eIF4F complex (comprised of eIf4A, eIF4E and eIF4G) on the mRNA cap (Chu *et al.*, 2016). For the formation of the 43S particle, the small subunit must be charged with the initiator methionyl tRNA (met-tRNAi), which is held in place by GTP-bound eIF2 (Hinnebusch, 2017). eIF4E is sequestered by hypophosphorylated 4E-BP1 (eIF4E binding protein 1), until the latter is phosphorylated to release eIF4E which binds the mRNA cap. The scaffolding protein eIF4G forms interactions with both the mRNA cap and PABP proteins bound to the poly(A) tail of the mRNA to cause transcript circularisation, which protects it from degradation (Wells *et al.*, 1998). In mammals, eIF4G also contains binding sites for eIF3, which restrains the 40S particle in an open conformation for mRNA attachment (Schutz *et al.*, 2008; Aitken & Lorsch, 2012). eIF4A is an ATP-dependent DEAD-box RNA helicase which has three isoforms in mammals (eIF4AI-III). These have some functional redundancy, but also tissue and development-specific expression patterns; suggesting functional relevance for each isoform (Chu *et al.*, 2016).

Once the 43S pre-initiation complex is bound, it scans the 5'UTR of the bound mRNA for the start codon (AUG), facilitated by the binding of additional initiation factors eIF1, eIF1A, eIF2, eIF3 and eIF5 (Schmidt *et al.*, 2016; Hinnebusch, 2017). eIF3 is the largest initiation factor complex and is required for optimal loading of the mRNA onto the 43S particle. It then scans for the initiation codon before stabilising the PIC-mRNA interaction (Aitken *et al.*, 2016). At the start codon, the PIC-mRNA complex base pairs with the initiator tRNA in the ribosomal P-site. The initiation factors are released in an energy-dependent manner and the complex is remodelled into the elongation competent 80S particle upon joining of the 60S subunit (Gebauer & Hentze, 2004). Subunit assembly is catalysed by eIF5B

and premature subunit association in the absence of an mRNA substrate is prevented by eIF6 which sequesters the 60S ribosome (Sonenberg & Hinnebusch, 2009; Schmidt *et al.*, 2016).

Internal ribosome entry site or IRES-dependent translation initiation can also occur which relies on secondary mRNA structures to drive initiation by recruitment of auxiliary factors and the pre-initiation complex. This process is less dependent on levels of eIFs, particularly those involved in scanning, therefore it is often used during cellular stress when initiation factors may be limiting (Sonenberg & Hinnebusch, 2009).

For translation elongation to proceed, the ribosome uses its three tRNA binding sites; the P site, to which the initiator codon binds, the acceptor (A) site and the exit (E) site which is the final site for tRNA binding. After initiation, eIF1A directs tRNA to the next codon of the open reading frame in a GTP dependent mechanism. On successful codon recognition, the tRNA is bound and the A site and GTP is hydrolysed. This triggers the release of eIF1A which can be recycled ready for the next round of elongation (Dever & Green, 2015). Once a peptide bond is formed between the tRNA in the A and the P site, translocation into the P and E sites can occur by binding eEF2. Peptide bond formation causes the small and large ribosomal subunits to rotate relative to each other to allow translocation along the transcript (Lareau *et al.*, 2014). As soon as the bound tRNAs are released from the E site, the next mRNA codon binds its cognate amino-acyl tRNA and the subsequent round of elongation beings (Zaher & Green, 2009; Dever & Green, 2015). During the pioneer round of translation, the EJC complex is removed from the mRNA. If this does not occur and it persists to encounter an elongating ribosome, it triggers nonsense mediated decay of the transcript (Hug *et al.*, 2015).

In eukaryotes, translation termination relies on the recruitment of the release factors eRF1 and eRF3. eRF1 mediates highly accurate recognition of a stop codon in the ribosomal A site (Zaher & Green, 2009; Dever & Green, 2015). Both eRF1 and eRF3 are also involved in mRNA quality control pathways to regulate the production of mRNAs including aberrant transcripts. These factors represent the two classes of release factors that are essential and widely conserved in eukaryotes (Atkinson *et al.*, 2008). eRF1 and eRF3 interact to form a distinct complex with the ribosome that is dependent on GTP. eRF3 increases the speed of nascent peptide release from the ribosome in such interactions, to cause GTP hydrolysis which allows eRF1 to be delivered to the A site (Dever & Green, 2015). Once the peptide chain has been released from the ribosome, the subunits are either recycled or can be reinitiated on the mRNA for subsequent rounds of translation by partially dissociating with the transcript (Dever & Green, 2015).

Translational readthrough, by bypassing the stop codon, into the 3'UTR sequence can produce less protein but does not trigger non-stop decay as there is no connection with poly(A) sequence. The 3'UTR sequence could be used to control protein levels of such transcripts where potentially aberrant readthrough occurs (Arribere *et al.*, 2016).

1.5.2 Translation regulation and downstream effectors

Initiation is the most comprehensively regulated stage of translation that requires numerous RNA binding proteins. For example, during periods of cellular stress the availability of core initiation factors is used to regulate translation (Sonenberg & Hinnebusch, 2009). Unsurprisingly, mutation of such RBPs involved in these processes can result in malignancies and tumorigenesis.

The eIF4F complex and components are crucial for cap-dependent translation initiation via interactions with the kinase mTOR, which is responsible for the phosphorylation of downstream factors that promote transcript-specific translation such as S6K1 (Saxton & Sabatini, 2012). This has important implications in cancer as mTOR is a platform for numerous signalling pathways that are key for oncogenesis (Chu *et al.*, 2016). mTOR is a central kinase in the control of cap-dependent translation and exists in two multiprotein complexes; mTORC1 and mTORC2, wherein it binds the cofactors Raptor and Rictor respectively (Hay & Sonenberg, 2004). mTOR is essential for viability and its disruption results in early embryonic lethality and impaired cellular differentiation (Gangloff *et al.*, 2004; Murakami & Ichisaka, 2004). Raptor null embryos have a phenotype similar to mTOR knockouts and die shortly after implantation, consistent with the necessity of a functional mTORC1 complex in embryogenesis (Guertin *et al.*, 2006). mTOR signalling can act to regulate translation throughout the cell-cycle in a transcript-specific nature. Interestingly, the RICTOR mRNA transcript is sufficient to mediate this effect and its translation is driven during the onset of S phase (Stumpf *et al.*, 2013)

The best characterised direct downstream effectors and phosphorylation targets of mTOR kinase are S6K1 and 4E-BP1, although the mTORC2 complex also targets Akt/PKB for phosphorylation (Guertin *et al.*, 2006). Interestingly, Raptor can also bind S6K1 and 4E-BP1, which once phosphorylated, releases eIF4E to drive cap-dependent translation initiation (Hay & Sonenberg 2004). Both mTORC1 and 2 are inhibited by rapamycin treatment, although mTORC1 is more acutely sensitive, and complete inhibition of both can be achieved through the use of inhibitors such as Torin1 that bind the mTOR active site (Fonseca *et al.*, 2015). mTOR possesses intrinsic serine/threonine kinase activity, thus is capable of

autophosphorylation (Hay & Sonenberg 2004). However, mTOR can also be activated by other factors bound to an mRNA. RNA binding proteins are crucial in these processes. SRSF1, for example, was shown to recruit mTOR to facilitate translation of a subset of mRNA targets, as discussed in detail later (Michlewski *et al.*, 2008). In line with this, mass spectrometry independently identified SRSF1 as a Raptor interacting partner (Fonseca *et al.*, 2015).

Other kinases such as cyclin dependent kinase 1 (CDK1) can also activate 4E-BP proteins. This mechanism is prevalent during mitotic cap-dependent translation, during which it phosphorylates the δ isoform of 4E-BP. Although, this is in contrast to studies which claim that cap-dependent translation is suppressed during mitosis, it does illustrate the extensive translational control mechanisms during the cell cycle (Shuda *et al.*, 2015).

Regulation and rates of translation can vary depending on cell type and differentiation state. For example, in embryonic stem (ES) cells, translation is minimal and more reliant on fine-tuned transcript-specific responses but during differentiation there is a global increase in polysome associated mRNAs and protein synthesis (Sampath *et al.*, 2008; Tahmasebi *et al.*, 2016). In this scenario, regulation is not uniform and transcripts can be categorised into varying groups, for example those which increase in abundance and/or translational efficiency during differentiation (Sampath *et al.*, 2008). Given the abundance of free ribosomes in undifferentiated cells, it is likely that they remain translationally poised for differentiation. When context cues are received, these are able to rapidly activate both global and transcript-specific translation; a process concurrent with an increase in 4E-BP phosphorylation (Sampath *et al.*, 2008). This is not limited to cap-dependent translation. For example, in human ES cells the eIF4G1 homologue DAP5/eIF4G2, which is unable to stimulate cap-dependent translation, regulates a subset of mRNA targets in a cap-independent manner. During differentiation it selectively regulates the translation of the HMGN3 transcript which encodes a crucial chromatin binder important in differentiation (Yoffe *et al.*, 2016).

4E-BPs are essential for the translation of key pluripotency factors, including the TET (ten eleven translocase) enzymes and Nanog, in mouse ES cells, and their ablation prevents reprogramming ability (Tahmasebi *et al.*, 2016). They are also crucial to repress the translation of factors that induce differentiation such as the transcription factor YY2. Remarkably, differential translation regulation of this transcript is coupled to the alternative splicing of an intron in the 5'UTR; providing an elegant example of the extensive coordination of layers of gene expression regulation (Tahmasebi *et al.*, 2016). In mouse ES cells, differential translation rates are frequently observed for splice variants, particularly when the splicing event involves the 5'UTR of the transcript (Wong *et al.*, 2016). Consistent with an essential role in

development, translational control is also exerted during the onset of meiosis in mammalian oocytes after nuclear envelope breakdown occurs, when mTOR/4E-BP/eIF4F-mediated regulation increases, particularly to regulate transcripts that contain a 5'terminal oligopyrimidine motif (5'TOP mRNAs). This is in contrast to the rapid dephosphorylation of 4E-BP (and hence inactivation of translation) that occurs upon fertilisation (Susor *et al.*, 2015). Downregulation of the mTOR complex during meiosis in oocytes results in a high frequency of chromosome and spindle abnormalities and downregulation of specific transcripts without affecting global levels of translation or blocking meiotic progression (Susor *et al.*, 2015).

Translational elongation is also subject to regulation and its malfunction results in abnormalities, for example defects in neurogenesis and the onset of neurological disorders. This can be attributed partially to the highly complex nature of the nervous system and associated stringent translation requirements for example highly localised translation (Richter & Coller, 2015). Ribosome stalling can also result in a wide range of malignancies from protein misfolding to oncogenesis, with the best characterised example being FMRP in intellectual disability (Richter & Coller, 2015).

Another major pathway in eukaryotic translational regulation is via micro RNAs (miRNAs). In short, miRNAs (~22nt long), whose biogenesis involves multiple proteins, most importantly Drosha and Dicer, bind to their target mRNA to trigger its degradation which requires Argonaute family proteins. miRNA binding ultimately results in a translational block of its target mRNA. There is extensive literature regarding this aspect of translational regulation, which will not be discussed in detail here.

1.5.3 Methods to study mRNA Translation – Ribosome profiling

Methods to study mRNA translation, particularly those which use a transcriptome-wide approach have triggered a revolution of knowledge in the field. Pioneering ribosome profiling or footprinting techniques have facilitated understanding of temporal, transcript-specific translational regulation (Ingolia, 2016). In such techniques, ribosomes are frozen on the mRNA, using the drug cyclohexamide, which halts translation through binding to the E site in the 60S subunits of elongating ribosomes. This ensures the ribosome remains bound to the target mRNA at the time of treatment and allows purification of ribosome/mRNA complexes. Ribosomes are then dissociated from their associated mRNAs and this mRNA "footprint" is subject to deep sequencing (Ingolia *et al.*, 2012).

Such methods have been revolutionary to gain an unprecedented understanding of targeted mRNA translation in numerous different cell types, stages of differentiation and growth conditions (Ingolia, 2014; Andreev *et al.*, 2017). One study coupled ribosome profiling with inhibition of mTORC1 by the mTOR inhibitor Torin1 in mouse embryonic fibroblasts (MEFs) (Thoreen *et al.*, 2012). Transcripts resistant to Torin1 encompassed both IRES containing transcripts and histone mRNAs, whereas those sensitive were highly enriched with 5'TOP (5'terminal oligopyrimidine) motifs or had long and complex 5'UTR sequences. Regulation of the latter by mTOR was highly dependent on 4E-BP regulation, perhaps as eIF4G1 is highly necessary to bind eIF4E to the cap in 5'TOP mRNAs (Thoreen *et al.*, 2012). Such motifs are enriched in ribosomal mRNA transcripts, and their regulation through mTOR-dependent mechanisms has also been demonstrated in the context of the RNA binding protein La-related protein 1 (LARP1). LARP1 is capable of binding 5'TOP containing transcripts to repress their translation in an mTOR-dependent manner by holding the mRNA in a highly stable closed loop conformation in which eIF4E is bound by 4E-BP1 (Fonseca *et al.*, 2015).

To investigate translation initiation, cells can be treated with drugs such as harringtonine or lactimidomycin, which prevent new 80S complexes from initiating on the mRNA, while allowing those already engaged in elongation to progress along the mRNA (Ingolia *et al.*, 2011; Lee *et al.*, 2012). Additional methods such as translation complex profile sequencing (TCP-seq) serve to investigate translation termination and ribosome recycling – information that is lacking in the ribosome profiling datasets (Archer *et al.*, 2016; Shirokikh *et al.*, 2017). In TCP-seq, formaldehyde crosslinking, rather than cycloheximide, is used to fix both the ribosome and its associated complex on the bound mRNA (Archer *et al.*, 2016). This facilitates the analysis of ribosome dynamics at each stage of translation as it does not discriminate against ribosomes composed of single subunits or transitionary complexes; for example, during the scanning phases of translation initiation when 40S subunits exist as the pre-initiation complex (Archer *et al.*, 2016; Shirokikh *et al.*, 2017).

1.6 SRSF1: The founding member of the SR protein family

SRSF1 has been extensively studied with respect to other members of the SR family to uncover overarching and highly regulated roles in each aspect of gene expression (Das & Krainer, 2014).

1.6.1 SRSF1 Structure and nomenclature

Shortly after the discovery of pre-mRNA splicing and the spliceosome, the minimal components of the reaction were defined *in vitro* (Krainer & Maniatis, 1985). Among those required for splicing initiation was the SR protein SF2 and the concurrently discovered ASF, later renamed SRSF1, which was micrococcal-nuclease sensitive and essential for *in vitro* splicing using S100 splicing deficient extracts.

The first evidence for human SRSF1 was described when cDNA corresponding to a 248 amino acid sequence was isolated from human cells (Ge *et al.*, 1991; Krainer *et al.*, 1991). Later gene mapping demonstrated that the corresponding gene was located at 17q21.3-q22 in humans and the analogous region of chromosome 11 in mice (Bermingham *et al.*, 1995). Similar to other SR proteins, SRSF1 consists of two RNA-binding domains, RRM1 and RRM2 and C-terminal RS domain (figure). It was noted in early work, that SRSF1 resembled other splicing factors in such as *Drosophila* tra2, which also contains an RS domain (Long and Cáceres, 2009). Moreover, the budding yeast SR-like protein, Npl3 shares functional homology with SRSF1 in mRNA splicing as it has been shown to promote pre-mRNA splicing of target transcripts in a mechanism coupled to transcription (Kress *et al.*, 2008).

SRSF1 mRNA is itself subject to alternative splicing and the transcript encodes several highly conserved isoforms (Sun *et al.*, 2010). The canonical isoform of SRSF1 contains four exons, which encodes a protein of 248 amino acids in length. This transcript is the predominantly expressed isoform and is highly conserved between human and mouse on both the mRNA and protein level. Other splice isoforms of SRSF1 are not evolutionarily conserved and several are predicted to undergo NMD. Interestingly three of these are retained in the nucleus and isoform ratio changes to increase these on overexpression of the canonical 248aa isoform (Sun *et al* 2010).

SRSF1 binds RNA with high affinity via two dynamic N-terminal RRM domains, essential in splicing to provide a high degree of substrate specificity. The SRSF1 RRM1 is a canonical RRM domain, but the second is a pseudo RRM domain, characterised by a highly conserved heptapeptide. RRM2 has a distinct binding mode and acts to regulate splicing events through antagonising other splicing factors such as hnRNPA1. These actions are blocked or occur with a lower affinity when the conserved heptapeptide is mutated (Cléry *et al.*, 2013). Conservation of this domain differs between SR protein family members and amino acid changes can be attributed to differential RNA binding. For example, the SRSF5 RRM domains are less conserved and bind RNA with a lower affinity in a distinct mode from SRSF1

(Cléry *et al.*, 2013). The SR protein family have harnessed these distinct modes of RRM substrate recognition and utilised them to coordinate and delegate differential splicing events.

1.6.2 SRSF1 Regulation and cellular localisation

SRSF1, as other SR proteins, is extensively post-translationally modified; to date it has been shown to undergo methylation, particularly arginine methylation, phosphorylation, ubiquitination and acetylation (Sinha *et al.*, 2010; Risso *et al.*, 2012). Modifications also differ between cell types, for example, in activated T cells, SRSF1 is ubiqutinated and degraded by the proteasome to compensate for mRNA upregulation (Moulton *et al.*, 2014).

1.6.2.1 SRSF1 Phosphorylation

Phosphorylation is the best understood and mechanistically characterised modification as it is key to the proper function and regulation of SRSF1 (Howard & Sanford, 2014). The characteristic RS domain of SR proteins is composed of serine/arginine repeats, which undergo continual cyclic phosphorylation to dictate both cellular localisation and function (Figure 1.6). Deletion of the RS domain is lethal and phosphorylation of this domain is required for nuclear escape of all shuttling SR proteins. The SRSF1 RS domain can be considered as two parts separated by a short spacer domain: RS1, consisting of an unbroken stretch of eight SR repeats, and RS2, which contains additional but interrupted SR repeats. NMR studies show that this domain is highly disordered but undergoes a conformational change on phosphorylation to become a partially rigid, arched structure termed an "arginine claw" (Hamelberg *et al.*, 2007; Xiang *et al.*, 2013). RS1 forms contacts with RRM2 which ensures the balance of phosphorylation and if disrupted results in increased translocation of SRSF1 to the cytoplasm (Serrano *et al.*, 2016). This demonstrates the essential and intrinsic coupling of phosphorylation and function of SR proteins.

There are several mechanisms, via different kinases, that result in SRSF1 phosphorylation *in vivo*. The main kinases responsible for RS domain phosphorylation are the nucleocytoplasmic shuttling SR protein kinases (SRPKs) and the exclusively nuclear Cdc2-like kinase (CLKs) family (Ko *et al.*, 2001; Ghosh & Adams, 2011).

SRPK1 is the prototypical SRPK that specifically phosphorylates serine residues in the SR dipeptide in a highly processive fashion once tightly bound to a substrate (Aubol *et al.*, 2003). This action is restrictive and only phosphorylates approximately half (up to 12) of the available target sites in the RS domain of SRSF1 (Xiang *et al.*, 2013). SRPK1 distribution is

largely cytoplasmic and regulated by a spacer element which facilitates this; alteration of this sequence results in aberrant nuclear translocation (Ding *et al.*, 2006). Data from the yeast homologue of SRPK1, Sky1p, demonstrated that the spacer domain and hence cellular localisation is essential for cell viability (Ding *et al.*, 2006). SRPK1 has a docking platform which mediates high-affinity interactions with a specific docking motif in SRSF1 (Ngo *et al.*, 2005). This motif is present in RRM2 which is more structured than the subsequent RS domain, thus can provide stability for docking and mediate directionality, although RS1 is also required (Serrano *et al.*, 2016). Once bound, SRPK1 feeds the RS domain directionally through the docking groove to its active site. Progressive phosphorylation causes a concomitant accumulation of negative charge which decreases SRPK1 binding affinity and encourages dissociation of the kinase (Ngo *et al.*, 2008). Cytoplasmic phosphorylation by SRPKs primes SRSF1 for nuclear import and speckle incorporation (Ngo *et al.*, 2005). This mechanism of phosphorylation is consistent with the observation that SR proteins are hypophosphorylated in nuclear speckles and require further phosphorylation for speckle exit and to participate in splicing (Cáceres *et al.*, 1997).

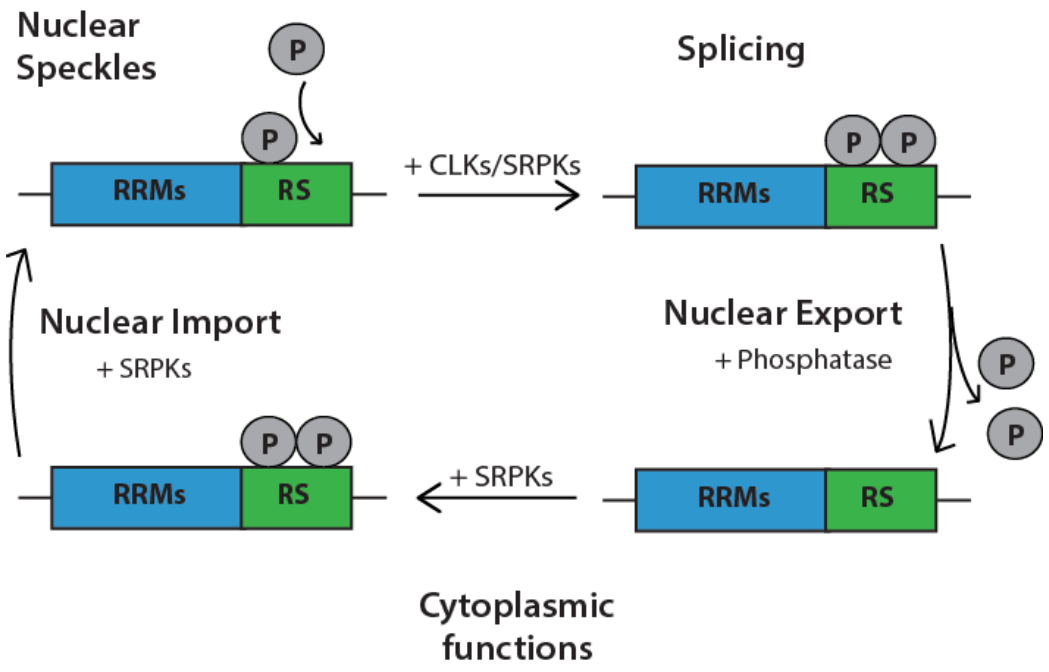


Figure 1.6: Phosphorylation cycle of SRSF1. SRSF1 is highly dynamic and shuttles between the nucleus and the cytoplasm as dictated by the phosphorylation status of its RS domain. The cellular location of SRSF1 governs the regulatory functions that it exerts on its bound mRNA. Note the two RRMs (blue) and the RS domain (green) are depicted as a continuum for visual ease. The top represented the nucleus and the lower part the cytoplasm.

SR proteins are hyperphosphorylated by CLKs, which govern nuclear distribution and facilitate splicing activity (Ngo *et al.*, 2005; Ghosh & Adams, 2011). There are four members of the CLK family (CLK1-4), all of which phosphorylate Ser/Pro dipeptides in addition to Ser/Arg dipeptides. CLK1 lacks the docking groove characteristic of SRPK1 but binds SRSF1 with a high affinity, via a distinct non-regiospecific mechanism which modifies up to 18 serine residues in the RS2 domain. The unique flexible N-terminus of CLK1, incidentally also an RS domain and nuclear localisation signal, is essential for this reaction as it directly binds the RS domain of SRSF1 (Aubol *et al.*, 2014). Also required for the reaction is the formation of a cooperative CLK1-SRPK1 complex in which the latter acts a release factor for the tightly bound CLK1 substrate (Aubol *et al.*, 2016). Specific Ser/Pro phosphorylation by CLK1 is required for SRSF1 exit from nuclear speckles and promotes RNA binding supporting SRSF1 function in splicing (Keshwani *et al.*, 2016).

SRSF1 requires dephosphorylation mediated by protein phosphatase 1 (PP1) for nuclear export and post-splicing functions. SRSF1 Ser/Pro phosphorylation causes a structural change that enhances RS domain dephosphorylation by PP1 highlighting the balance between kinase and phosphatase action (Keshwani *et al.*, 2016). The RS1 domain undergoes directional (N to C terminal) dephosphorylation by PP1 in a relatively inefficient reaction (Ma *et al.*, 2010).The RRM2 of SRSF1 directly binds PP1 and RRM1 allosterically sequesters the PP1 to dampen its catalytic action. Dephosphorylation of SRSF1 is balanced and highly regulated to maintain proper protein localisation. PP1 directs this to mediate the recently proposed model of "goldilocks" dosing of SRSF1 phosphorylation (Aubol *et al.*, 2017).

1.6.2.2 Autoregulation

All SR proteins and some other splicing factors autoregulate the level of their own expression through alternative splicing coupled to nonsense mediated decay (AS-NMD or unproductive splicing) (Lareau *et al.*, 2007; Ni *et al.*, 2007). SRSF1, for example, promotes inclusion of an intron after the canonical stop codon, which marks it as a premature termination codon (PTC) to elicit NMD of its own mRNA. For SRSF1, this exon is encoded in the ultraconserved sequence element in the 3'UTR although the location in the host transcript varies between SR family members (Lareau *et al.*, 2007) . An ultraconserved region is classified as >200bp perfect conservation in mouse, rat and human genomes (Ni *et al.*, 2007). The unproductive splicing events themselves are also highly conserved between evolutionary kingdoms and actively selected for, being even present in fungi. Furthermore, this mechanism of regulation has been shown to be extremely plastic in its evolution, arising

as independent events and multiple occurrences after loss in some lineages (Lareau & Brenner, 2015). Overall, this indicates the crucial nature of the process in maintaining optimum SR protein expression levels. Between cell types, the extent to which the PTC-containing isoforms prevail differs significantly, despite predictions that they account for approximately 7% of transcripts. One study found that in HCT116 cell lines the PTC containing variant is equivalent to the canonical SRSF1 isoforms. Furthermore, in these cells, the PTC-containing isoform is regulated by micro RNA (miRNA) binding and reducing mRNA stability rather than by NMD per se, consistent with earlier speculations regarding miRNA regulation (Sun *et al.*, 2010; Akaike *et al.*, 2011). Interestingly there is degree of cross-regulation between SR family members as SRSF3 binds to the retained intron cassette within other SR protein mRNA transcripts (Anko *et al* 2012). SRSF1 also antagonistically regulates SRSF3 autoregulation through splicing repression of the unproductive SRSF3 isoform (Jumaa & Nielsen, 1997)

The epithelial-to-mesenchymal transition (EMT) is a crucial process in early development and in the context of epithelial tumorigenesis. In this process, cells reorganise themselves from an epithelial state to a mesenchymal one through implementing a defined gene regulatory network to trigger specific signalling pathways particularly via the ERK1/2 kinase (Valacca *et al.*, 2010). During EMT, SRSF1 levels are differentially regulated by AS-NMD which is attenuated through treatment with an ERK1/2 inhibitor. The splicing factor and ERK1/2 kinase substrate Sam68 is responsible for this regulation as it binds to the ultraconserved element in the 3'UTR of SRSF1 thereby influencing unproductive splicing of the transcript. This mechanism of regulation is able to drive EMT or the reverse process MET in SW480 cells. (Valacca *et al.*, 2010).

1.6.3 Roles in oncogenesis

Considering the ubiquitous and essential nature of alternative splicing, it is unsurprising that its misregulation can have aberrant consequences, not least in oncogenesis. As master regulators of alternative splicing, SR and hnRNP proteins are obvious candidates for malfunction and indeed there are countless examples of their splicing dysregulation in a host of human malignancies. It was recognised that SR protein mediated dysregulation of alternative splicing could promote malignancy in early studies on breast cancer (Stickeler *et al.*, 1999). SRSF1 itself is a protooncogene and is upregulated in numerous human cancers, as are several other SR proteins. Overexpression of SRSF1 is sufficient to drive tumorigenesis through alteration of alternative splicing programs and also in regulation of apoptotic pathways that play a role in cancer progression (Karni *et al.*, 2007). Specifically, the RRM1

of SRSF1 was shown to be mandatory to drive tumour formation in immortalised cells. SRSF6 is also a proto-oncogene upregulated in several cancers such as lung and colon cancer, in which it promotes the formation of potent oncogene splice isoforms of transcripts such as MNK2. Interestingly, in colon cancer samples it is far more amplified than any other SR protein including SRSF1 (Cohen-Eliav *et al.*, 2013). SRSF6 is also implicated in basal cell carcinomas and melanomas (Jensen *et al.*, 2014)

SRSF1 modulates the splicing pattern of the tyrosine kinase and protooncogene Ron through binding to either an enhancer or silencer within exon 12 of the transcript to mediate the inclusion or exclusion of exon 11 (Ghigna *et al.*, 2005). Skipping of this exon results in a Ron isoform whose product is constitutively active and highly upregulated in many types of tumours. In this scenario, the amount of the malignant Ron isoform produced is dependent on SRSF1 levels. When SRSF1 is overexpressed, thus promoting exon 11 skipping in the Ron transcript, it promotes the loss of an epithelial phenotype and drives towards an aggressive mesenchymal state, in particular increasing cellular motility and tumour invasiveness (Ghigna *et al.*, 2005).

In breast cancer, SRSF1 is able to alter the splicing profiles for numerous genes including the kinases S6K1 and MNK2 by generating isoforms that promote transformation (Anczuków *et al.*, 2012, 2015). In breast tissue, tumours formed are highly malignant and invasive. Overexpression of SRSF1 preferentially generates the MNK2b isoform, which results in eIF4E phosphorylation to drive oncogenic translation pathways. In MCF-10A cells, which model mammary acini structure, SRSF1 overexpression results in enlarged acini and increased proliferation in an mTOR signalling dependent mechanism (Anczuków *et al.*, 2012). In addition, among other examples, increased levels of SRSF1 cause inactivation of the tumour suppressor gene BIN1 by promoting synthesis of an isoform unable to supress the oncogenic activity of its downstream target c-Myc (Karni *et al.*, 2007). In breast cancer tissue versus normal breast tissue from the same patients, the SRSF1 genomic locus was highly amplified specifically in tumours. This indicates that in part increased SRSF1 can correlate with gene copy number, however it can also be significantly upregulated in tumours without gene amplification (Karni *et al.*, 2007). Consistent with dysregulation of splicing, upregulation of SRSF1 nuclear functions is sufficient to drive tumorigenesis and increase in acinar size in the breast cancer model (Anczuków *et al.*, 2012).

SR proteins, particularly SRSF1 are also dysregulated in non-small cell lung cancer (NSCLC). SRSF1 upregulation has been proposed as a prognostic biomarker in NSCLC, correlated with poor survival as it is a major cancer driver (Jiang *et al.*, 2016). SRSF1

upregulation can enhance the expression of the anti-apoptotic protein survivin through alternative splicing changes in tumours. However, in contrast to breast cancer models, in NSCLC, SRSF1 can directly increase the production of survivin protein through mRNA binding and consequently upregulation of its translation in an mTOR dependent mechanism (Ezponda *et al.*, 2010). This indicates that the cytoplasmic roles of SRSF1 can influence cancer progression in a context dependent manner, concurrent with the fact that mTOR, specifically mTORC1 activation, is necessary for SRSF1-mediated transformation. Overexpression of SRSF1 causes increased phosphorylation of 4E-BP1 and S6K1 (both substrates for mTOR) to drive translation (Karni *et al.*, 2008). In contrast to wt SRSF1, overexpression of the nuclear retained fusion protein SRSF1-NRS, not cause tumour formation in cellular transformation assays in p53 null immortalised liver progenitor cells, despite a constant proliferation rate between them (Shimoni-Sebag *et al.*, 2013).

For the most part, splicing factor mediated splicing changes are implicated in cancer but there are reports that alterations of their downstream roles are also pathogenic. For example, hnRNP A1 is increasingly localised in the cytoplasm in breast cancer and can alter transcript expression via translational regulation to drive tumorigenesis. In this context it enhances the translation and hence protein expression of RON by directly binding to the 5'UTR of the mRNA which forms a G-quadruplex structure (Cammas *et al.*, 2016).

1.6.4 Functions of SRSF1 outside splicing

The nuclear roles of SRSF1 in splicing are well understood, however, as a shuttling protein it also has several ill-defined cytoplasmic functions. Endogenous mRNA transcripts are associated with SRSF1 in both the nucleus and the cytoplasm whereas SRSF2 exclusively associates with nuclear mRNAs. In fact, a significant proportion of SRSF1 RNA binding capacity is for cytoplasmic RNAs and those in polysomes (Sanford *et al.*, 2009; Sterne-Weiler *et al.*, 2013). Many of the RNA binding targets of SRSF1 also overlap between cellular compartments, corroborating that SRSF1 moves with target transcripts throughout their processing lifetime (Sanford *et al.*, 2008). Although the cytoplasmic roles of SRSF1 can be decoupled from those in splicing, it is believed that certain mRNA transcripts remain bound to SRSF1 throughout their processing until they are translated (Sanford *et al.*, 2005). In such a model, SRSF1 is exported from the nucleus with its target mRNA as part of an mRNP particle. In HeLa cells, such mRNPs contain a high copy number of SR proteins which is in correlation with exon-junction complex number (Singh *et al.*, 2013).

1.6.4.1 SRSF1 in mRNA translation

A role for SRSF1 in mRNA translation was established by Western blotting of cytoplasmic extracts separated across sucrose gradients in which it was found in polysomal fractions to be directly associated with 80S ribosomal particles (Sanford *et al.*, 2004). In addition to this, SRSF1 enhanced the translation of an mRNA reporter that harboured an ESE known to promote its binding. Interestingly, the presence of multiple ESEs, and presumably reciprocal increase in amount of SRSF1 bound to the transcript, caused an additive effect on translational stimulation which was unaffected by the presence of an intron (Sanford *et al.*, 2004). The correct phosphorylation state of the SRSF1 SR domain dictates shuttling propensity and SRSF1 must be dephosphorylated in the cytoplasm to interact with the translation machinery. The RS domain is required for mRNA binding in the cytoplasm but not for translational function, the latter requires RRM2. It was shown that mutating three residues of the conserved heptapeptide (WQD → AAA) of RRM2 abrogates SRSF1 function in translation, consistent with its requirement for SRSF1 splicing function (Sanford *et al.*, 2005).

SRSF1 activity can be modulated in response to growth factors which trigger a PI 3-kinase and Akt signalling cascade. Using the aforementioned EDA reporter system, it was shown that Akt can target SRSF1 to alter cellular localisation and hence translation. In addition, overexpression of the other SR protein kinases SRPK1 and CLK1 cause SRSF1 to accumulate in the nucleus in a hyperphosphorylated state, indicating that phosphorylation-mediated cellular location of SRSF1 is crucial for its varying functions (Blaustein *et al.*, 2005).

SRSF1-mediated translation occurs in an RNA-dependent mechanism that requires the translation initiation factor eIF4E and is increased by the presence of the m7G mRNA cap. SRSF1 is incapable of driving translation of known IRES-containing reporter constructs (Michlewski *et al* 2008). SRSF1 is able to stimulate mRNA translation *in vivo* via an mTOR dependent mechanism (Figure 1.7) (Michlewski *et al* 2008). Interestingly, this is not through direct interaction of the eIF4F complex or 4E-BP proteins but by modulating the phosphorylation status of the latter in a direct interaction with the mTOR kinase. Addition of SRSF1 to HeLa cell extracts decreased the rate 4E-BP1 dephosphorylation in an mTOR-dependent manner. As shown using Co-IP experiments, SRSF1 interaction with mTOR is direct and RNA independent and triggers the phosphorylation of 4E-BP1. SRSF1 is also capable of direct binding to PP2A phosphatase; however, this may be indirect, perhaps mediated by an auxiliary factor, as PP2A was not pulled down by SRSF1. PP2A is the phosphatase responsible for dephosphorylation of 4E-BP proteins hence acting antagonistically to mTOR. Intriguingly, the aforementioned SRSF1 RRM2-AAA mutant

(inactive in translation) was incapable of mTOR binding suggesting that this region of the RRM2 is essential and hence mTOR is required for SRSF1 function in translation to releases eIF4E for incorporation into the eIF4F complex and drive cap-dependent translation of the transcript.

It is interesting to note that the yeast SRSF1 homologue, Npl3 is also capable of nucleocytoplasmic shuttling and is able to regulate mRNA translation, for example it has a role in ensuring correct 80S monosome formation for successful translation initiation and termination (Estrella *et al.*, 2009; Baierlein *et al.*, 2013). Analogous roles for SRSF1 to Npl3 in initiation or termination of translation have yet to be determined. However, similarly to SRSF1, Npl3 is modified by arginine methylation which dictates its association with the U1 snRNP to facilitate a role in splicing (Muddukrishna *et al.*, 2017). Together, this indicates that the high conservation of the shuttling and translational roles of SR proteins are highly important for cellular physiology.

The mRNA transcripts translationally regulated by SRSF1 have been previously characterised in human cells in which it was transiently overexpressed (Maslon *et al.*, 2014). Polysome profiling analysis after cell fractionation over sucrose gradients, coupled with RNA deep sequencing identified transcripts that transitioned from the sub to the heavy polysomal fractions on SRSF1 overexpression. These transcripts represent those whose translation was enhanced by SRSF1 overexpression. The number direct translational targets of SRSF1 were inferred when these data were overlapped with previous CLIP-seq datasets to total 505 mRNA transcripts. Remarkably, transcripts identified were enriched for those encoding cell, cycle and mitotic proteins including those encoding RNA processing and splicing factors (Maslon *et al.*, 2014). Furthermore, many of these were regulated by SRSF1 at both the splicing and post-splicing level as they overlapped CLIP-seq data for SRSF1 (Sanford *et al.*, 2008; Maslon *et al.*, 2014). This provides strong indication that SRSF1 is capable of coupling splicing and translational regulation of its targets. Hence it is feasible to imagine an intricate regulatory network in which certain transcripts are bound by SRSF1 from alternative splicing to translation (such as the Clk1 mRNA), whereas others are regulated by SRSF1 at a single step alone.

SRSF1 function in translation is required for proper cell cycle and mitotic progression and its depletion from the cytoplasm results in formation of a multipolar mitotic spindle phenotype and cell cycle arrest. This can be attributed to the requirement of SRSF1 in the translational regulation of important related proteins such as NDC80 and SMC4 which are required for kinetochore assembly and accurate chromosome segregation (Maslon *et al.*,

2014). Concurrent with this, an independent genome-wide siRNA based study to establish mitotic regulators using high-throughput time-lapse microscopy of mitosis, identified SRSF1 as a candidate which displayed aberrant phenotypes on depletion including mitotic delay (Neumann *et al.*, 2010). In addition, specific mRNA transcripts are differentially regulated throughout the cell cycle, many of which are shown to be regulated by SRSF1. SRSF1 (and SRSF3) bind to interphase and post-mitotic chromatin in phosphorylation-dependent interactions with histone H3 but are released from mitotic chromosomes when both are phosphorylated by SRPK1 and Aurora B kinase respectively. SRSF1 is required in this process to remove HP1 proteins from chromosome to permit G0/G1 progression (Loomis *et al.*, 2009). Such data offer an explanation into the mitotic defects and delay in G0/G1 entry that occur on SRSF1 knockdown (Loomis *et al.*, 2009; Stumpf *et al.*, 2013).

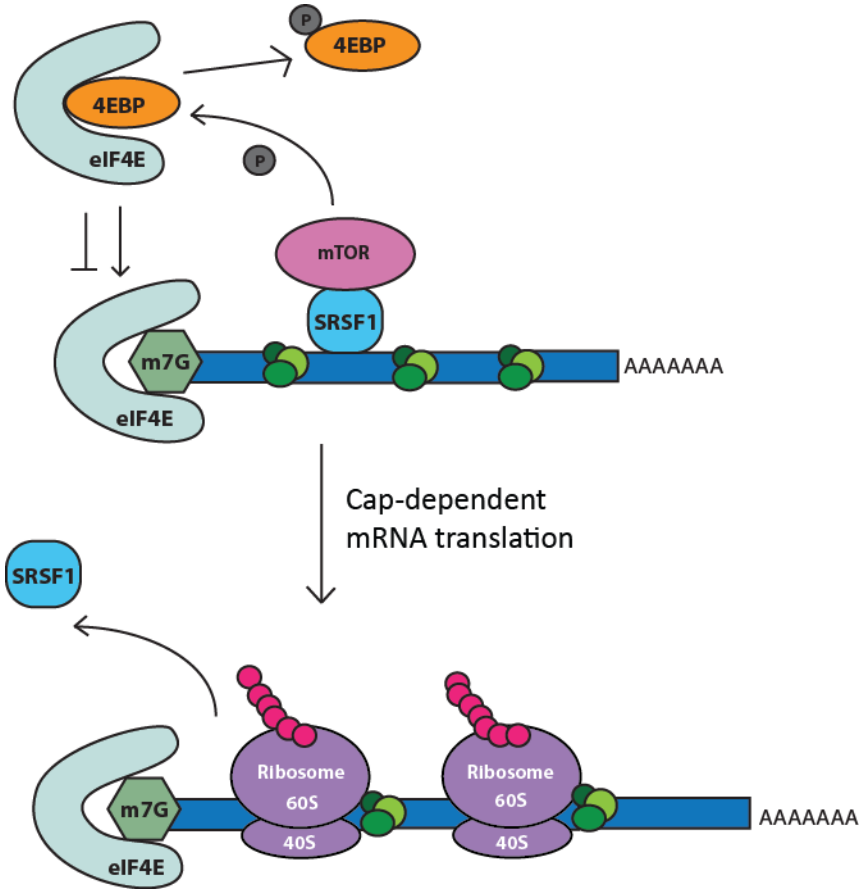


Figure 1.7: SRSF1 in mRNA translation. SRSF1 acts to enhance the translation of specific target mRNA transcripts in an mTOR-dependent mechanism. SRSF1 is bound to cytoplasmic targets which facilitates mTOR recruitment to drive phosphorylation of 4E-BP. This releases eIF4E which binds to the capped mRNA to initiate translation of the SRSF1-bound substrate (Michlewski *et al.*, 2008).

The physiological relevance of SRSF1-mediated translational in cell cycle regulation has yet to be dissected. Strikingly, it has recently been shown by proteomic analysis of polysome during the cell cycle, that splicing factors are recruited to and highly enriched on elongating ribosomes during mitosis (Aviner *et al.*, 2017). In particular, this study exemplified hnRNPC as showing significant increase in association with polysomes (between G1 and M phase); though SRSF1 and other SR proteins were also enriched in the dataset. The authors show that this association during mitosis is not reciprocal for or dependent on the spliceosome suggesting a clear distinction of functions. Furthermore, hnRNPC was able to influence the translation of ribosomal proteins and translation factors (Aviner *et al.*, 2017). hnRNPA1 was also shown to be involved in the translational regulation of specific transcripts (Roy *et al.*, 2014; Kim *et al.*, 2017)

Other shuttling SR proteins have also been implicated in translational control. For instance, SRSF3 is capable of interacting with PDCD4 (programmed cell death factor 4), which represses translation through inhibiting eIF4e action. PCDC4 is a putative tumour suppressor frequently downregulated in tumours. In cancer models, SRSF3 was shown to antagonise PCDC4 expression at numerous post-transcriptional stages including splicing, mRNA export and translation. Interestingly, the type of SRSF3-mediated regulation is isoform-specific; it decreases the translational efficiency of one specific PDCD4 isoform, but influences the splicing and export of another (Park & Jeong, 2016). The translational activity in this context was previously shown to be dependent on binding of the 5'UTR of the transcript (Kim *et al.*, 2013).

SRSF9 is also a shuttling SR protein that can regulate mRNA translation. For example it can upregulate Wnt/β-catenin signalling pathways by increasing the levels of the β-catenin protein (Fu *et al.*, 2013). This was shown to be dependent on mTOR, suggesting that the mechanism of translational activation outlined for SRSF1 could be shared with other members of the shuttling SR family to potentiate translation.

1.6.4.2 SRSF1 in genome instability and the cell cycle

In addition to the cell cycle defects caused by the ablation of SRSF1 function in mRNA translation, SRSF1 roles in splicing are also vital for proper cell cycle progression. SRSF1 ablation in DT40 cells which express a single cDNA copy of tetracycline inducible SRSF1 results in cell cycle arrest at G2 and concurrent apoptotic cell death, although the latter occurred in the absence of DNA laddering (Li *et al.*, 2005). SRSF1 regulates the alternative splicing of ICAD long and short (ICAD-L and ICAD-S respectively) isoforms which promote

apoptotic DNA laddering. ICAD isoform ratio must be properly maintained, but SRSF1 depletion causes a shift in the ratio of ICAD isoforms towards less ICAD-L which inhibits DNA fragmentation despite apoptosis initiation (Li & Manley, 2005).

SRSF1 is also required for the maintenance of genome stability and preventing R-loop formation by associating with mRNA transcripts, concurrent with roles in transcription and splicing. This not only illustrates the crucial nature for the coupling of gene expression regulatory pathways but explains why SRSF1 depletion causes genomic rearrangements and increased mutagenic frequency in DT40 and HeLa cells (Li & Manley, 2005). Interestingly, transcription induced R-loop formation on SRSF1 knockdown can be counteracted by exogenous overexpression of RNase H (the enzyme that resolves these structures in vivo) but cannot prevent cell death. It was speculated by the authors that this may be due to ablation of other essential processes which SRSF1 orchestrates (Li & Manley, 2005).

SRSF2 is also involved in the G2/M cell cycle transition and maintaining genome stability (R loops also formed on knockdown as for SRSF1). Its role in cell proliferation is dependent on p53, and loss of p53 is sufficient to ablate the cell cycle block induced on SRSF2 deletion. This has implications for the role of SR proteins in cancer, especially if p53 is mutated (Xiao *et al.*, 2007). SRSF1 is also capable of regulating p53 by preventing its proteasome-mediated degradation, which increases when SRSF1 is overexpressed. On induction of oncogenic or ribosomal stress, SRSF1 forms a crucial component of the RPL5-MDM2 complex which acts to mediate the oncogene-induced senescence pathway via p53 activation (Fregoso *et al.*, 2013).

1.6.4.3 SRSF1 role in mRNA export

SRSF1, and other SR proteins including SRSF3 and SRSF7, act as adapters for the export of mRNP particles travelling to cytoplasm, by recruiting the export factor NXF1/TAP (Huang *et al.*, 2003). This interaction is preferential for dephosphorylated SR proteins; hence such complexes assemble more readily after splicing. The mRNP is then able to directly interact with the nuclear pore complex to promote export of the mature mRNP to the cytoplasm (Lai & Tarn, 2004). Interestingly, this is not intron-dependent as SR proteins are capable of promoting the export of intronless transcripts (Huang & Steitz, 2001). For SRSF3 and SRSF7, NXF1/TAP binding is mediated by interactions with an arginine-rich motif at the C-terminus of the single RRM domain that does not contact RNA (Hargous *et al.*, 2006). However SRSF1, which contains two RRM domains, utilises the arginine-rich linker sequence between them to bind NXF1/TAP distinctly (Tintaru *et al.*, 2007). Although these SR proteins

can bind NXF1/TAP independently of RNA, an mRNA transcript is required for a stable association *in vivo*. Indeed, the 3'end of SR-protein bound target mRNA is thought to be necessary for NXF1/TAP binding.

Extensive iCLIP experiments and RNA-sequencing of cytoplasmic poly(+) mRNA have shown that SR proteins target specific transcripts for export and there is little overlap between them (Muller-McNicoll *et al.*, 2016). However, individual SR protein depletion had little effect on these numbers suggesting that there is some functional redundancy at least for export. The only outlier is SRSF3 whose roles in export cannot be compensated for by other SR proteins, indicating it is the most crucial for mRNA export. Interestingly, many transcripts targeted for export by all SR proteins are alternatively spliced, particularly with differential terminal exons and alternative polyadenylation marks, which may have different requirements for NXF1/TAP binding. This illustrates the roles that SR proteins play in the tight coupling of gene regulatory mechanisms (Muller-McNicoll *et al*. 2016).

1.6.4.4 SRSF1 in Nonsense-Mediated Decay and mRNA stability

Nonsense-mediated decay (NMD) is a translation-dependent mRNA degradation pathway that removes aberrant mRNA transcripts, which harbour a premature termination codon, thereby acting as a surveillance mechanism for faulty transcripts. It serves to regulate both a proportion of normal transcripts and those which are faulty in disease (Hug *et al.*, 2015). SRSF1 has a role in enhancing nonsense mediated decay (NMD) response in targeting reporter constructs that contain a premature termination codon. This was observed on SRSF1 overexpression and it requires a functional RS domain but not nucleocytoplasmic shuttling, although the exact mechanisms remain unclear. In addition to SRSF1, other SR proteins were also to elicit an NMD response to a lesser extent with the same reporter constructs (Zhang & Krainer, 2004).

SRSF1 also plays a role in determining the stability of mRNA transcripts to which it is bound. In DT40 chicken cells that expressed a single copy of tetracycline inducible SRSF1 cDNA, it was shown to destabilise a polyadenylated reporter construct by binding to the 3'UTR of the transcript. This behaviour was shown to be an independent SRSF1 function that is not determined by its role in splicing, however the exact mechanism of action remains unclear (Lemaire *et al.*, 2002). Therefore, although SRSF1 can travel with a newly transcribed mRNA until its translation, it is also capable of regulating targets at specific stages of gene expression.

1.6.4.5 SRSF1 in Protein sumoylation

Further to its roles in numerous aspects of mRNA processing, SRSF1 also has a role in protein sumoylation; the process by which the post-translational modification SUMO (small ubiquitin-related modifier) is added to proteins (Pelisch *et al.*, 2010). SRSF1 is capable of regulating cellular levels of sumoylation in a dose dependent manner specifically through interactions of its RRM2 domain with the SUMO E3 ligase PIAS1. This activity is independent of SRSF1 roles in splicing, despite the fact that sumoylation has a role during the splicing process (Pelisch *et al.*, 2010). Interestingly, this function of SRSF1 appears to be important to remove R-loops (RNA:DNA hybrids) and DNA breaks. To do so SRSF1 participates in a complex with PIAS1 and the DNA helicase RECQ5 to cause the SUMOylation of Topoisomerase I, which ultimately prevents genomic instability (Li *et al.*, 2015).

1.7 SRSF2: A non-shuttling SR protein

SRSF2 is a non-shuttling SR protein that resides constitutively in the nucleus and is a major component of and marker of nuclear speckles (Cáceres *et al.*, 1998; Saitoh *et al.*, 2004). Due to nuclear retention, SRSF2 is unable to partake in post-splicing roles analogous to other SR proteins. In contrast to shuttling SR proteins, SRSF2 does not bind NXF1/TAP, so does not play a direct role in export of its bound transcript mRNA. However, it may have an indirect role in export as SRSF2 knockdown results in changes in cytoplasmic mRNA fractions that are distinct from other SR proteins (Muller-McNicoll *et al*. 2016). SRSF2 binds distinct mRNA targets with more plasticity than its family counterparts, with its RNA binding consensus being SSNG (Fu & Ares, 2014); this could indicate a binding capacity for different types of transcripts in comparison to other SR proteins (Anko 2014).

SRSF8, previously known as SRp46, is the processed retropseudogene of SRSF2, which contains large sequence homology (Soret *et al.*, 1998). It is believed to not shuttle between the nucleus and the cytoplasm, but the propensity for this is untested. SRSF8 is functional as an SR protein and has distinct roles in splicing to SRSF2 and binds to different target mRNAs (Soret *et al.*, 1998). It is an orthologue of SRSF2 that arose after *C.elegans* in which there is no homologue (Longman *et al.*, 2000).

1.7.1 SRSF2 and Nuclear retention

SRSF2 remains sequestered in the nucleus due to a potent nuclear retention sequence (NRS) present at the very C-terminus of its RS domain (Cazalla *et al.*, 2002). The NRS is not only unique amongst SR proteins, but it has virtually no homologues in the human genome at peptide or nucleotide level. By nature, all RS domains have a high propensity for disorder, but the SRSF2 RS domain shows an even further increase in disorder due to the NRS. A high level of disorder in a sequence is used to increase the binding capacity and plasticity of the protein (Calabretta & Richard, 2015; Castello *et al.*, 2016). This distinct property of SRSF2 may allow it to be uniquely modified or bind certain factors that other SR proteins cannot. SRSF2 is also highly phosphatase resistant, suggesting that in comparison to shuttling SR proteins, it is less readily dephosphorylated. Shuttling SR proteins require dynamic dephosphorylation for nuclear export, thus nuclear retention of SRSF2 could be explained by differential or inefficient dephosphorylation (Cazalla *et al.*, 2002).

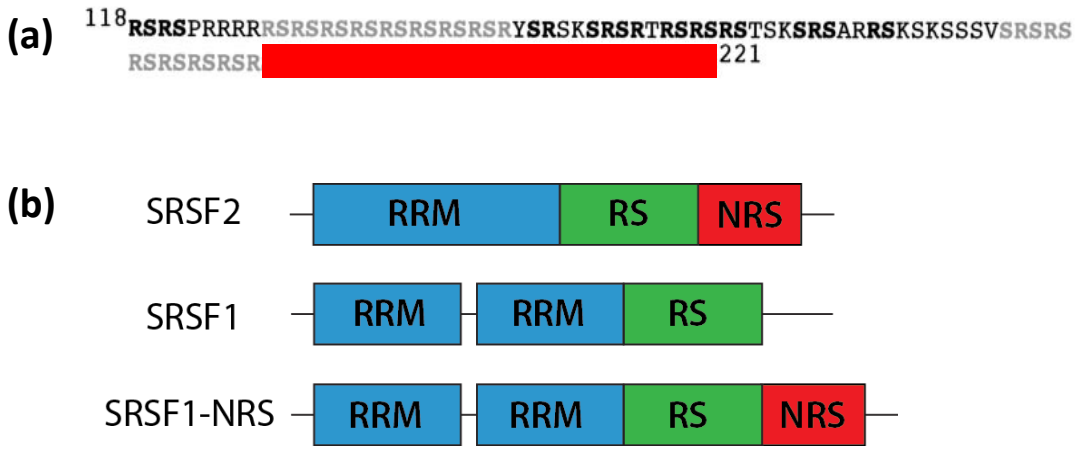


Figure 1.8: Nuclear retention of SRSF2 (a) Peptide sequence from the RS domain of SRSF2. The nuclear retention sequence (NRS) resides at the C-terminus of this domain, highlighted in red. (b) Domain structure of wt SRSF2 (non-shuttling), wt SRSF1 (shuttling) and the SRSF1-NRS fusion protein, (non-shuttling). In the SRSF1-NRS fusion protein the entire wt SRSF1 sequence is present. In each case, the NRS is highlighted and not drawn to scale. Figure adapted from (Cazalla *et al.*, 2002)

When fused to a shuttling protein, the NRS is sufficient to ablate its transition into the cytoplasm. This has been demonstrated extensively both within and outside SR family proteins (Cazalla *et al.*, 2002). The fusion protein SRSF1-NRS which does not shuttle to the nucleus but remains active in splicing, hence preserving cellular viability, is an important control to study the cytoplasmic functions of SRSF1 as it should behave in the antagonistic

fashion to shuttling SRSF1 (Sanford *et al.*, 2004). Indeed, it has been used in this capacity in numerous studies, including those that address SRSF1 function as an oncogene (Anczuków *et al.*, 2012). Although the functional capacities of the NRS for highly effective nuclear retention have been harnessed in numerous studies, the exact mechanism of its action are unknown.

SRSF1 containing the entire SRSF2 RS domain was functional and able to fully rescue the conditional knockout in DT40 cells. Furthermore, when SRSF1 is fused to just the nuclear retention sequence (NRS) from the RS domain of SRSF2, it is capable of rescuing the viability of conditional knockout mouse embryonic fibroblasts (Lin *et al.*, 2005). This strongly implies that both the RS domain and the shuttling roles of SRSF1 are not required for viability in this cell type.

1.8 Aims and objectives

The cellular roles of the SR protein family in splicing and alternative splicing regulation are exemplified by the prototype SR protein SRSF1, which continually shuttles between the nucleus and the cytoplasm due to continual phosphorylation of the RS domain. The nuclear functions of SRSF1, particularly in splicing, are well characterised and its mRNA targets have been previously described for this process in diverse systems. SRSF1 also has numerous cytoplasmic roles outside splicing, including the regulation of mRNA translation in an mTOR-dependent mechanism (Michlewski *et al* 2008). The direct translational mRNA targets of SRSF1 have been previously characterised for human cells in which it was transiently overexpressed (Maslon *et al.*, 2014). However, in comparison to the wealth of knowledge for SRSF1 function in the nucleus, little is understood regarding its cytoplasmic roles or the functional relevance of SR protein shuttling.

The overarching aim of this work is to further investigate the function of SRSF1 in the cytoplasm and examine the physiological consequences of its absence in both novel and highly relevant cell and animal model systems.

To do this, there are three key aims for the project:

(1) To create an endogenous model system for non-shuttling SRSF1 to further understand the cytoplasmic roles of SRSF1 in mouse ES cells.
(2) To determine the interactome of shuttling and non-shuttling SRSF1 protein
(3) To examine the physiological consequences of SRSF1 shuttling through creation of a mouse model for non-shuttling SRSF1.

Chapter 2

Materials and Methods

2.1 General Molecular Biology Methods

2.1.1 Plasmid preparation

To generate repair plasmids for CRISPR targeting in cells, the PCDNA3+ expression vector was used. Endogenous sequences were amplified from mouse genomic DNA, using 2x GoTaq Green Master Mix (Promega) and primers containing the appropriate restriction site overhangs (section 2.8). The thermal profile used was following the manufacturer's protocol with an optimised annealing temperature of 62^{o}C. Excess primers and PCR reaction components were removed from all PCR products using the PCR purification kit according to the manufacturer's instructions (Qiagen). After sequence verification, clean PCR products were ligated overnight 4^{o}C into the appropriately digested PCDNA3+ expression vector using DNA ligase (NEB). Ligation independent cloning was carried out using the same PCR amplification steps but insertion into the digested vector was via annealing of complementary primer overhangs.

Overexpression constructs for mass spectrometry were generated from existing Cáceres lab stocks. After bacterial amplification, plasmids were prepped using the Plasmid Maxi kit (Qiagen) as per the manufacturer's protocol. DNA was quantified by nanodrop and verified by Sanger sequencing.

2.1.2 Site Directed mutagenesis

PAM site mutations were generated on the CRISPR repair template using the QuickChange II site-directed mutagenesis kit (Agilent Genomics) and the appropriately designed primers (section 2.8). Plasmids were amplified using AccuPrime *Pfx* high fidelity DNA polymerase (ThermoFisher Scientific), with the thermal profile given in the manufacture's guidelines. Plasmids were subsequently digested with DpnI to remove the non-mutated plasmid. Reactions were transformed as detailed elsewhere and plasmids prepared using the Plasmid Mini kit (Qiagen). Successful mutations verified by Sanger sequencing, and plasmids purified to transfection grade using Plasmid Maxi kit (Qiagen).

2.1.3 Agarose gel electrophoresis

All agarose gels for electrophoresis were prepared using 1x TBE buffer (1 M Tris base, 1 M Boric acid, 0.02 M EDTA) supplemented with ethidium bromide (2 µl/100 ml). 1xTBE was used as the running buffer and the gel percentage was adjusted to suit application.

2.1.4 Bacterial DNA transformation

For all cloning applications, DNA was transformed into sub-cloning efficiency DH5α *E.coli* bacterial cells (Invitrogen). 50µl thawed cells were used per transformation. After addition of DNA, cells were incubated on ice for 30 minutes then subjected to heat shock at 42°C for 45 seconds, followed by 2 minutes on ice. 500 µl of SOC medium (2% (w/v) Tryptone, 0.5% (w/v) Yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM Glucose), was added and tubes were incubated at 37°C for 1 hour with shaking at 225 rpm. Transformants were plated on either Amplicillin or Kanamycin agar plates as appropriate.

2.1.5 Western Blotting

Cells were grown to 80-90% confluency in dishes and washed twice in ice cold PBS before being harvested on ice with a cell scraper in 1 ml PBS. Cells were pelleted by centrifugation at 1000 rpm for 4 minutes and the supernatant removed before lysis in the appropriate volume of lysis buffer (50 mM Tris-Cl pH 8, 150 mM NaCl, 1% NP40, protease inhibitors (complete mini EDTA-free, Roche)). For Western Blotting from mouse tissues, the sample was homogenised and lysed in the same lysis buffer used for cells. In both cases, lysis was performed for 15 minutes on ice and spun at 13000 rpm for 12 minutes. The cleared lysate was transferred to a clean Eppendorf on ice. Bradford assay to estimate protein concentration were carried out for each sample in duplicate using the BioRad reagent and homemade BSA standards. The latter were homemade by serial dilution from 5 mg/ml BSA (NEB). Protein concentration was equalised using lysis buffer and samples prepared for SDS-PAGE (10x reducing agent, 4x sample buffer, (x) µl lysate).

For standard western blotting, precast 4-12% Bis-Tris gels were used (NuPAGE, Invitrogen). The benchmark prestained protein ladder (Invitrogen, 10748010) was run alongside all samples to calibrate molecular mass. Gels were transferred onto appropriately sized, preassembled nitrocellulose membrane stacks using the iBlot 2 dry blotting system (Invitrogen) as in manufacturers protocol, following programme P3 (20 V, 7 minutes). The membrane was stained briefly with Ponceau S to visualise successful transfer of proteins before blocking in 5% nonfat fat milk made up in PBS tween (0.1% (v/v)) for 1 hour with shaking. Unless otherwise stated, all antibodies were prepared in blocking solution and membranes were incubated for 1 hour at room temperature or overnight (4°C). For antibody specifications and concentrations used see section 2.9.

2.1.6 Immunoprecpitation

Cells were plated in 10 cm dishes and grown until 90% confluent. For endogenous IPs cells were harvested 24 hours after plating. Transfected cells were harvested 48 hours post plating and 24 hours post transfection. Cells were washed twice with ice cold PBS and harvested on ice by scraping before lysis in 500 µl ice cold IP/lysis buffer (20 mM Tris-Cl pH 8, 150 mM NaCl, 1 mM EDTA, 1% NP40, 0.2% Sodium Deoxycholate, 1 mM DTT, protease inhibitor tablets (complete mini EDTA-free, Roche)). Lysates were incubated on ice for 20 minutes and subject to centrifugation at 13 000 rpm for 15 minutes. The supernatant was loaded on Protein G sepharose fast flow beads (GE healthcare). Prior to use, beads were washed three times in 1 ml PBS and resuspended to 50% slurry in IP/lysis buffer. Lysates were precleared using 20 µl Protein G beads per sample for 30 minutes at 4^{o}C with rotation.

After preclearing, protein concentration was determined using Bradford assay as described for Western Blotting. Protein concentrations were calibrated in an equal volume of IP/lysis buffer and incubated with 40 µl of T7-antibody coupled beads (Millipore) overnight rotating at 4^{o}C. Beads were washed three times in 1 ml cold PBS and prepared to 50% slurry in IP/lysis buffer prior to use. After incubation beads were centrifuged for 2 minutes at 2000 rpm to remove the unbound fraction. Beads were washed three times in IP/lysis buffer with rotation at 4^{o}C for 5 minutes and twice without incubation. For mass spectrometry experiments the final two washes were performed using 1x TBS (Tris-buffered saline, in place of IP/lysis buffer) to remove excess detergent in the samples. Finally, beads were resuspended in 10 µl 1x TBS before submission to the in house mass spectrometry facility. For other experiments, complexes were eluted from beads by heating at 95^{o}C in 4x LDS sample buffer supplemented with 10x reducing agent (both NuPAGE, LifeTech) and subjected to SDS-PAGE.

To visualise mass spectrometry samples prior to their analysis, a portion of IP beads were eluted in sample buffer and run on precast 4-12% Bis-Tris polyacrylamide gels as previously described. Gels were stained using either SilverQuest or Colloidal Blue Staining kits (both Invitrogen) following the manufacturer's protocols.

2.1.7 DNA Sequencing

All Sanger sequencing, including that for genotyping, was performed by in-house local facility on the 3130/3730 genetic analyser (Applied Biosystems). Sequencing primers used are indicated where appropriate and in Table 2.1. Sequence analysis was carried out using the SnapGene software package (GSL Biotech; available at snapgene.com).

2.1.8 Subcellular fractionation

To fractionate cells for mass spectrometry, HeLa cells were grown in 15 cm dishes until confluent, washed twice in PBS and harvested by scraping. Cells were pelleted by centrifugation at 1000 rpm for 4 minutes and resuspended in 150 µl nuclei buffer A (NB-A) (85 mM KCl, 5.5% Sucrose, 10 mM Tris pH 7.5, 0.5 mM Spermidine, 0.2 mM EDTA, 250 µM PMSF). 150 µl of nuclei buffer B (NB-B) (NB-A supplemented with 0.2% NP40) was added and samples mixed well by inversion. Lysates were incubated on ice for 5 minutes then spun at 2000 rpm for 3 minutes to pellet nuclei. The soluble supernatant was harvested as the cytoplasmic fraction. Nuclei were washed gently in 1 ml of nuclei release buffer (NB-R) (85 mM KCl, 5.5% Sucrose, 10 mM Tris pH 7.5, 1.5 mM $CaCl_2$, 3 mM $MgCl_2$, 250 µM PMSF) and centrifuged at 1200 rpm for 3.5 minutes. Nuclei were resuspended in 150 µl NB-A and sonicated for 5 cycles of 30 seconds on/30 seconds off using the high setting on the Biorupter sonicator. Extracted protein was subjected to Western blotting, to ensure the purity of each fraction.

2.2 Cell culture Techniques

2.2.1 Cell Lines

HeLa, NIH3T3 and mouse E14 embryonic stem (ES) cells were all obtained in house from lab stocks in liquid nitrogen storage. Murine Neural Stem Cells (ANS4), directly isolated from adult mouse brain tissue, were a gift from Steven Pollard's Laboratory at the University of Edinburgh.

2.2.2 Maintenance of cell lines

All cell lines used were grown under standard culture conditions at 37ºC with 5% CO_2. NIH3T3 cells and HeLa cells were maintained in Dulbecco's Modified Eagle Media (DMEM) (Gibco Thermo) supplemented with 10% fetal calf serum (FCS) and detached on passaging using 1x Trypsin-EDTA (diluted from 10x stocks in PBS, Sigma).

Mouse E14 ES cells, were maintained in Glasgow Minimum Essential Media (GMEM) supplemented with 10% (v/v) FCS, 1% (v/v) Non-essential amino acids (LifeTech), 1% (v/v) Sodium Pyruvate (Sigma), 0.1% (v/v) 2-mercaptoethanol (50 nM) (ThermoFisher), 1% (v/v) L-Glutamine, 1% (v/v) Pen/Strep and LIF (homemade). mESCs were grown on

plates pre-coated with gelatin (0.1% (w/v) sterile stock from powder (Sigma)) and all cell washes were performed using sterile PBS (Sigma).

Neural Stem Cells were grown on plates pre-coasted with 0.1% gelatin in DMEM/HAMS-F12 +HEPES + Sodium Bicarbonate +L-Glutamine (Sigma), supplemented with 7.25ml Glucose (Sigma), 1x MEM NEAA (Gibco), 1% Pen-Strep (Gibco), 0.16% (v/v) BSA Solution 7.5% (Gibco), 0.1% 2-mercaptoethanol (50 nM) (ThermoFisher), 1% (v/v) B27 Supplement 50x (LifeTech/Gibco), 0.5% (v/v) N2 Supplement 100x (LifeTech/Gibco 17502-048). Before use media was further supplemented with mouse EGF (final concentration 10 ng/ml) (Peprotech), human FGF (final concentration 10 ng/ml) (Peprotech) and Laminin (final concentration 1 μg/ml) (Sigma L2020-1MG/ML). Cells were washed in sterile PBS (Sigma) and detached using Accutase solution (eBioSciences).

2.2.3 DNA Transfections

All DNA transfections were carried out using Lipofectamine 2000 (ThermoFisher Scientific) following the manufacturer's protocol. Briefly, DNA was diluted in the appropriate volume of opti-MEM serum free media (Gibco). In a separate tube, the appropriate volume of lipofectamine was also diluted in opti-MEM media. The two mixes were combined at a 1:1 ratio and incubated at room temperature for 20-30 minutes. For standard transfections, DNA/lipofectamine mixes were added dropwise to cells. For reverse transfections, cell suspension was added to the plate with the transfection mixes and cells allowed to adhere to the place. All media was replaced after 6-12 hours.

2.2.4 Propidium Iodide staining

For Propidium Iodide staining prior to FACs analysis, mouse E14 cells were grown to confluency in a T75 flask and harvested by trypsinisation before collection by centrifugation at 1000rpm for 4 minutes. Cells were washed once in PBS before fixation in ice-cold 100% ethanol and incubation at 4^{o}C for 24 hours. Cells were spun at 1000 rpm for 4 minutes and washed once in PBS, before cell pellets were resuspended in PBS. RNase A was then used at 0.5 μg/ml (Roche) for 30 minutes at 37^{o}C, before incubation with 50 μg/ml propidium iodide (Sigma) for 15 minutes on ice. The cell cycle distribution of the stained cells was then analysed by FACs.

2.3 CRISPR Methodology in Cells

2.3.1 Guide RNA Design

The 20 nucleotide Guide RNA sequences required for Cas9 specificity were chosen using the online tool at http://crispr.mit.edu/ (Zhang Lab) and further design undertaken as in Ran *et al* 2013. Up to 200 bp of genomic DNA sequence at the appropriate locus was used (Ensembl genome browser) at each selection. Guide RNAs with the highest "score", as determined by the program, were preferentially chosen. This score takes into account both base pair match number to the target region and any potential off-target binding. Complementary sequence to each guide was determined and Bbs1 restriction overhangs added to each end to facilitate ligation into the targeting vector (Ran, Hsu, Wright, *et al.*, 2013).

2.3.2 gRNA Cloning

All gRNAs were ordered as custom single stranded DNA oligos (Sigma) with the appropriate reverse sequence and resuspended to 100 µM in dH_2O. A full table of gRNA sequences used can be in section 2.8. Complementary oligos were annealed at 100 µM in 10x T4 DNA Ligase Buffer (NEB) by heating to 95ºC for 2 minutes in a heat block which was subsequently allowed to cool slowly to room temperature. Successful annealing was confirmed by gel electrophoresis and introduced into either the PX461 or PX458 cloning vectors using Bbs1 (NEB) restriction cloning.

2.3.3 Fluorescence activated cell sorting (FACS)

Transfected cells were analysed for GFP+ fluorescence and single cell sorted using the BD FACS Jazz flow cytometer (BD Biosciences). Suitable and stringent fluorescence gating was applied to cells which were directly sorted into either 96-well plates or falcon tubes containing appropriate cell medium. Cells sorted into the latter vessel were subsequently plated at a density of 100 or 200 cells/ml in 10 cm dishes for colony expansion. For cell cycle analysis of clones, the FACs Fortessa LRII was used.

2.3.4 Isolation of Clones

Clonal cell populations were isolated from 10 cm dishes after 8-10 days of colony expansion. Individual colonies were isolated under a dissection microscope using a P20 Gilson pipette and deposited into single wells of a 96-well plate containing 30 µl warm trypsin. After

incubation for 5 minutes at 37°C, trypsin was quenched using 70 µl of appropriate warm media. Dispersed clones were transferred to fresh 96-well plates containing 100 µl of media. After 24 hours growth, growth media was replaced to remove any residual trypsin.

2.3.5 Genomic DNA extraction

After replica plating, cells were grown in 96-well plates for 2-3 days until overconfluent. To extract genomic DNA, cells were washed twice with PBS and lysed either in 50 µl DirectPCR Lysis Reagent (Cell) (Viagen Biotech), supplemented with 0.5 µg/µl proteinase K, as per the manufacturer's instructions or in 40 µl of homemade Proteinase K buffer (50 mM Tris pH 8.9, 1.5 mM $MgCl_2$, 0.5% (v/v) Tween and 0.5 ul/ul Proteinase K. Cells were incubated overnight in a homemade humidity chamber at 55°C. Genomic DNA precipitated the following day using 100 µl 100% ethanol supplemented with 150 mM NaCl for a minimum of 30 minutes at -20°C. Plates were spun using at a low speed and resultant genomic DNA pellets washed 3 times in 70% ethanol before resuspension in 30 µl TE buffer or dH_2O.

2.3.6 PCR-based genotyping

Genotyping PCR screening reactions for mice and cells were carried out using 2x GoTaq Green Master Mix (Promega) following the manufacturers guidelines. 1 or 2 µl of the cell lysate was used for amplification. The forward SRSF1 Ex4 screening primer and reverse SRSF1 3'UTR screening primer were used at a final concentration of 1.0 µM. All primer sequences can be found in section 2.8. The thermal profile used was optimised to: initial denaturing 95°C 2 minutes, then 35 cycles of: second denaturing 95°C 30 seconds, annealing 62°C 30 seconds, extension 72°C 1 minute, before final extension 72°C 5 minute. PCR products were separated by standard gel electrophoresis on 1 or 1.5% (w/v) TBE agarose gels.

2.3.7 Sub-cloning PCR products

To identify heterozygous CRISPR clones, PCR products were subcloned first then multiple transformants sequenced to identify the allele frequency. PCR products generated by a Taq polymerase, which have a 3' A overhang, were cloned into the PGEM T-easy vector (Promega), following the manufacturers protocol. For cloning of sequencing libraries, blunt-ended PCR products were cloned into the pCRII-TOPO blunt end vector system following the manufacturer's protocol. The T7 primer was used for sequence verification in both cases.

2.4 Targeting methodologies for CRISPR mice

2.4.1 Microinjection Reagents

For blastocyst injections, Cas9 RNA was used in combination with *in vitro* transcribed gRNAs and a DNA oligo repair template. Mix components were synthesised and assembled as follows. gRNA sequence from PX458 was amplified using oligos complementary to guide sequence and containing the T7 promoter sequence at the 5' end of the forward primer (section 2.8) PCR was performed using GoTaq 2x Green Master Mix in a 50 µl reaction volume with 200 ng template DNA. PCR thermal programme used was follows: initial denaturing 95ºC 2 minutes, then 35x cycles of: second denaturing 95ºC 30 seconds, annealing 56ºC 30 seconds, extension 72ºC 20 seconds, before final extension 72ºC 5 minutes.

PCR product was purified using the PCR purification kit (Qiagen) as per the manufacturer's instructions. A second PCR amplification step was performed using the entire elute (30 µl) from this cleanup of the first PCR reaction using the same thermal cycling profile.

RNA synthesis was carried out from the amplified DNA using the MEGAscript T7 Transcription kit (Invitrogen, AM1334) following the manufacturer's instructions. The entire purified PCR product (approximately 1 mg) was used with incubation at 37ºC overnight. Transcribed RNA was purified using the RNAeasy kit (Qiagen) following the "cleanup" part of the manufacturer's protocol. RNA was quantified using the Nanodrop, with dilution as appropriate. Injection mixes were prepared in a total volume of 20 µl as follows: Cas9 RNA, 50 ng/µl, guide RNAs, 25 ng/µl, repair template oligo, 75 ng/µl. Mixes were centrifuged at maximum speed, 4ºC for 30 minutes to remove impurities, before the supernatant was transferred to a fresh tube.

2.5 High throughput screening of clonal populations

2.5.1. Generation of CRISPR clone pools

Mouse E14 ES cells were grown and transfected as previously described for CRISPR experiments. Both wt cells and were targeted in parallel experiments (see Chapter 3, section 3.3.6 for details of clone 1). For wt mESCs, the original CRISPR strategy was repeated (same repair oligo and gRNA); SRSF1-NRS-T7 (clone 1) cells were retargeted with a new repair oligo and the allele specific gRNA (section 2.8).

To determine the base allele frequency either wt or SRSF1-NRS-T7 (clone 1) cell populations, t=0, $1x10^6$ cells were harvested at the time of transfection, by centrifugation at 1000 rpm for 4 minutes and pellets snap frozen on dry ice. The rest of the cells were reverse transfected as previously and FACs sorted 48 hours later. Prior to FACs sorting, $1x10^6$ cells of both wt and SRSF1-NRS-T7 (clone 1) cells were harvested for t=1. FACs sorted cells were divided into two pools; one for passaging in flasks, and the other pool plated in 10 cm dishes at a low density for picking clones. Clones were picked into 96-well plates as previously described. Cells in flasks were passaged and 10^6 cells collected for time points two to seven over 18 days. Collected cell pellets were lysed and the genomic DNA extracted using the DNeasy blood and tissue kit (Qiagen, 69504) as the manufacturer's protocol. Genomic DNA was resuspended in 100 µl AE buffer provided with the kit and quantified using the Qubit high sensitivity dsDNA assay (ThermoFisher Scientific).

2.5.2 Miseq Library Preparation

Library Preparation for Miseq experiments was carried out following the Illumina protocol for 16S Metagenomic Sequencing Library Preparation (available at: https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html) with various adaptations. For all PCR steps Q5 high fidelity DNA polymerase (NEB) was used to amplify 2 µl of genomic DNA obtained as previously (section 2.4.1). Nextera adapter primers, containing indices were custom ordered from IDT (section 2.8). Libraries were quantified using both the Qubit high sensitivity double stranded DNA assay and the Bioanalyser Agilent High Sensitivity. 5 µl of each library at 5 nM was pooled and sequenced on the MiSeq platform (Edinburgh Genomics, Ashworth Building, Kings Buildings, University of Edinburgh). All data analyses from the high throughput Miseq experiments were carried out by Martijn Kelder of Andrew Wood's Lab at the IGMM, University of Edinburgh.

2.6 Imaging Techniques

2.6.1 Immunofluorescence

Coverslips for immunofluorescence (round, 16 mmFisher Scientific) were sterilised with 100% ethanol before use. Cells were plated onto coverslips in 12 well plates at a density of $1x10^4$ cells/ml, except for reverse transfection of the T7-SRSF1 plasmid into mESCs, when $1x10^5$ cells/ml were used. Cells were rinsed with PBS and fixed with 4% (w/v) paraformaldehyde (PFA) for 10 minutes at room temperature. After fixation, cells were

washed 3x for 5 minutes with PBS and permeabilised with 0.2% (v/v) Triton X-100 for 10 minutes at room temperature. Cells were washed 3x for 5 minutess with PBS. Coverslips were incubated in block (1% (w/v) BSA, 0.01% (v/v) Triton in PBS) for 1 hour before primary antibody application (used as stated in section 2.9). Cells were washed 3 times for 5 minutes in PBS with 0.01% (v/v) Triton X-100, and incubated with Alexafluor secondary antibodies 488 (FITC) or 594 (Texas Red) (Abcam) for 1 hour and washed 3 times for 5 minutes in PBS with 0.01% (v/v) Triton X-100. Cell nuclei were stained with DAPI (50 mg/ml) for 10 minutes and rinsed briefly in dH2O. Coverslips were mounted onto slides using Vectashield mounting medium (Vector Labs) and sealed with nail varnish before imaging.

2.6.2 Heterokaryon Assays

Heterokaryon assays were principally performed as described in Cáceres *et al* (1998), with a few modifications. Mouse E14 ES cells were plated on sterile, gelatin coated coverslips in 12 well plates at a density of $1x10^5$ cells/ml. Cells were grown to confluency in a T75 flask (Corning), harvested 48 hours after plating the and resuspended at $1x10^6$ cells/ml. Media was removed from the mouse E14 ES cells and resuspended HeLa cells were added at a 1:1 ratio in the presence of 50 µg/ml (in 100% ethanol) cyclohexamide (CHX) for 3 hours. The concentration of CHX was increased to 100 µg/ml for 30 minutes before media was removed and cells washed in sterile PBS. Coverslips were inverted onto a drop of prewarmed 50% PEG 1000 for 2 minutes, then washed twice in sterile PBS to remove the PEG. Cell fusions were placed in heterokaryon growth media (complete ES cell media supplemented with 100 µg/ml CHX) for 2 hours before fixation as for Immunofluorescence.

2.7 Mass Spectrometry

Samples sent for Mass spectrometry analysis were prepared as in section 2.1.6. Subsequent to immunoprecipitation, 10 µl of T7-antibody coupled beads were washed twice in 1x TBS before submission to the IGMM in house mass spectrometry facility, which performed analysis and data processing. All mass spectrometry experiments were carried out using the using the Q-Exactive hybrid quadrupole Orbitrap mass spectrometer (ThermoFisher). Prior to injection into the machine, molecules were separated by nano liquid chromatography using the Dionex UPLC (UltiMate 3000 RSLCnano system, ThermoFisher).

For peptide Mass Spectrometry experiments, Robot IPs were carried out prior to proteomic analysis by the IGMM facility. Peptides were couple Steptavadin breads

(Dynabeads) and the Kingfisher Duo (ThermoFisher Scientific) robot used to perform the IP. Protein complexes were washed in Lysis buffer (see 2.1.6), before elution and processing analogous to manually performed IPs.

Raw data were analysed using MaxQuant software and protein annotations were obtained from the latest release of the UniProt database both by Jimi Wills of the IGMM in house mass spectrometry facility. Gene Ontology (GO) analysis was performed using online tools available from the Gene Ontology Consortium, available at: http://www.geneontology.org/.

2.8 Oligonucleotides

CRISPR targeting vector PCR primers	**Sequence 5'-3'**
SRSF1 Exon 1 FWD	GCGGGTACCATGTCGGGAGGTGGTGTGAT
SRSF1 Exon 1 REV	TACGGATCCTGTACGAGAGCGAGATCTGC
SRSF1 3'UTR FWD	CAGATGGGTTAATCTAGAGATGATTGGTGACACTTTTT
SRSF1 3'UTR REV	ATAGGGCCCTCTAGACAATTTCATCTGTGACAATAGC
T7 FWD	AATAATGAATTCATGGCATCGATGACAGGTGGCCAACAGAT GGGTTAATCTAGAGCG
T7 REV	CGCTCTAGATTAACCCATCTGTTGGCCACCTGTCATCGATGC CATGAATTCATTATT
SRSF2 NRS FWD	AATGGATCCCCTCCGCCCGTGTCGAAGC
SRSF2 NRS REV	TCGTATGAATTCGGAAGAAACTGCTCCCTCTTC
CRISPR modified targeting vector PCR primers	**Sequence 5'-3'**
SRSF1-T7 linker FWD	GATCCCCCGGCGCCGGCGCCATGGCATCGATGACAGGTGG CCAACAGATGGGTTAAG
SRSF1-T7 linker REV	AATTCTTAACCCATCTGTTGGCCACCTGTCATCGATGCCATG GCGCCGGCGCCGGGG
SRSF1-T7x2 FWD	GATCCATGGCATCGATGACAGGTGGCCAACAGATGGGTAT GGCATCGATGACAGGTGGCCAACAGATGGGTTAAG
SRSF1-T7x2 REV	AATTCTTAACCCATCTGTTGGCCACCTGTCATCGATGCCATA CCCATCTGTTGGCCACCTGTCATCGATGCCATG

SRSF1-T7x3 FWD	GATCCATGGCATCGATGACAGGTGGCCAACAGATGGGTATGGCATCGATGACAGGTGGCCAACAGATGGGTATGGCATCGATGACAGGTGGCCAACAGATGGGTTAAG
SRSF1-T7x3 REV	AATTCTTAACCCATCTGTTGGCCACCTGTCATCGATGCCATACCCATCTGTTGGCCACCTGTCATCGATGCCATACCCATCTGTTGGCCACCTGTCATCGATGCCATG
Site directed mutagenesis	**Sequence 5'-3'**
mutPAM A SRSF1 FWD	CCCAAGCTTGGTACCATGTC
mutPAM B1 SRSF1 REV	TCTGCTTCTTCTTGGGGAGT
mutPAM B2 SRSF1 REV	CAGTTACTCCCCAAGAAGAA
mutPAM C1 SRSF1 FWD	ACTCCCCAAGAAGAAGCAGA
mutPAM C2 SRSF1 FWD	AGAAGAAGCAGAGGATCACC
mutPAM D SRSF1 REV	GTGTCACCAATCATCTCTAGATTA
gRNAs	**Sequence 5'-3'**
Cas9n A FWD	CACCGGCTATGACGGGGAGAATAGCG
Cas9n A REV	AAACCGCTATTCTCCCCGTCATAGCC
Cas9n B FWD	CACCGGTACGGCTTCTGCTACGACTA
Cas9n B REV	AAACTAGTCGTAGCAGAAGCCGTACC
Cas9 wt A FWD	CACCGGGGCTCTCGTACATAAGATGAT
Cas9 wt A REV	AAACATCATCTTATGTACGAGAGCCCC
Retargeting Clone 1 gRNAs	**Sequence 5'-3'**
Cas9 wt A FWD	CACCGGGGAGGGGATCCTGTACGAGAG
Cas9 wt A REV	AAACCTCTCGTACAGGATCCCCTCCCC
SRSF1 knockin screen PCR primers	**Sequence 5'-3'**
SRSF1 Exon 4 FWD	TTGATGGGCCCAGAAGTCC
SRSF1 3'UTR REV	ATAGGGCCCTCTAGACAATTTCATCTGTGACAATAGC
Mouse Microinjection gRNA PCR primers	**Sequence 5'-3'**
Cas9 wt A FWD	TGTAATACGACTCACTATAGGGGGGCTCTCGTACATAAGATGAT

Cas9 wt A REV	AAAAGCACCGACTCGGTGCC
CRISPR HDR template ssODNs	**Sequence 5'-3'**
SRSF1-T7(linker)	TTACTCCCCAAGGAGAAGCAGAGGATCACCACGCTATTCTC CCCGTCATAGCAGATCTCGCTCTCGTACAGGATCCCCCGGC GCCGGCGCCATGGCATCGATGACAGGTGGCCAACAGATGG GTTAAGATGATTGGTGACACTTTTTGTAGAACCCATGTTGTA TACAGTTTTCCTTTACTCAGTACAATCTTTTCA
SRSF1-NRS-T7 (BamHI)	AAGCAGAGGATCACCACGCTATTCTCCCCGTCATAGCAGAT CTCGCTCTCGTACAGGATCCCCTCCGCCCGTGTCGAAGCGA GAGTCCAAGTCTAGGTCGCGGTCCAAGAGCCCACCCAAGTC TCCAGAAGAAGAGGGAGCAGTTTCTTCCATGGCATCGATG ACAGGTGGCCAACAGATGGGTTAAGATGATTGGTGACACT TTTTGTAGAACCCATGTTGTATACAGTTTTCCTTTACTC
SRSF1-NRS-T7 (EcoRI) (retargeting Clone 1 mESCs)	ACCACGCTATTCTCCCCGTCATAGCAGATCTCGCTCTCGTAC AGAATCCCCTCCGCCCGTGTCGAAGCGAGAGTCCAAGTCTA GGTCGCGGTCCAAGAGCCCACCCAAGTCTCCAGAAGAAGA GGGAGCAGTTTCTTCCATGGCATCGATGACAGGTGGCCAAC AGATGGGTTAAGATGATTGGTGACACTTTTTGTAGA
SRSF1-wt (mouse injections)	CGTAGCAGAAGCAACAGCAGGAGTCGCAGTTACTCCCCAA GGAGAAGCAGAGGATCACCACGCTATTCTCCCCGTCATAGC AGATCTCGCTCTCGTACATAAGATGATTGATGACACTTTTTG TAGAACCCATGTTGTATACAGTTTTCCTTTACTCAGTACAAT CTTTTCATTTTTTTAATTCAAGCTGTTTTGTTCAG
Mouse UBE sexing PCR	**Sequence 5'-3'**
UBE1XA FWD	GGCAGCAGCCCATCATAATCCAGATC
UBE1XB REV	TGGTCTGGACCCAAACGCTGTCCACA
Amplicon PCR primers for Miseq	**Sequence 5'-3'**
SRSF1 FWD	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTGATGG GCCCAGAAGTCC
SRSF1 REV	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGAATG TAGATGTTAGGAGCAAGG
FWD Index PCR primers for Miseq (I5)	**Sequence 5'-3'**
S502 NRS t=0	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTC GGCAGCGTC

S503 NRS t=1	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC
S505 NRS t=2	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTC
S506 NRS t=3	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC
S507 NRS t=4	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTC
S508 NRS t=5	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTC
S510 NRS t=6	AATGATACGGCGACCACCGAGATCTACACCGTCTAATTCGTCGGCAGCGTC
S511 NRS t=7	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGTCGTCGGCAGCGTC
S502 WT t=0	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTC
S503 WT t=1	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC
S505 WT t=2	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTC
S506 WT t=3	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC
S507 WT t=4	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTC
S508 WT t=5	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTC
S510 WT t=6	AATGATACGGCGACCACCGAGATCTACACCGTCTAATTCGTCGGCAGCGTC
S511 WT t=7	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGTCGTCGGCAGCGTC
REV Index PCR primers for Miseq	**Sequence 5'-3'**
N701 NRS t=0	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGTCTCGTGGGCTCGG

N701 NRS t=1	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGTCTCGTGGGCTCGG
N701 NRS t=2	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGTCTCGTGGGCTCGG
N701 NRS t=3	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGTCTCGTGGGCTCGG
N701 NRS t=4	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGTCTCGTGGGCTCGG
N701 NRS t=5	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGTCTCGTGGGCTCGG
N701 NRS t=6	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGTCTCGTGGGCTCGG
N701 NRS t=7	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGTCTCGTGGGCTCGG
N702 WT t=0	CAAGCAGAAGACGGCATACGAGATCGTACTAGGTCTCGTGGGCTCGG
N702 WT t=1	CAAGCAGAAGACGGCATACGAGATCGTACTAGGTCTCGTGGGCTCGG
N702 WT t=2	CAAGCAGAAGACGGCATACGAGATCGTACTAGGTCTCGTGGGCTCGG
N702 WT t=3	CAAGCAGAAGACGGCATACGAGATCGTACTAGGTCTCGTGGGCTCGG
N702 WT t=4	CAAGCAGAAGACGGCATACGAGATCGTACTAGGTCTCGTGGGCTCGG
N702 WT t=5	CAAGCAGAAGACGGCATACGAGATCGTACTAGGTCTCGTGGGCTCGG
N702 WT t=6	CAAGCAGAAGACGGCATACGAGATCGTACTAGGTCTCGTGGGCTCGG
N702 WT t=7	CAAGCAGAAGACGGCATACGAGATCGTACTAGGTCTCGTGGGCTCGG

2.9 Antibodies

Primary Antibody	**Technical specifications**	**Concentration**
SRSF1	Mouse monoclonal 96 (Hanamura *et al*. 1998)	1:1000

T7 tag	Mouse monoclonal, Novagen (69522)	1:10000, 1:5000
Tubulin	Mouse monoclonal, Sigma (T8328)	1:1000
H3	Rabbit polyclonal, Abcam (1791)	1:50000
XRN2	Rabbit polyclonal, Bethyl Labs (A301-103A)	1:1000
HnRNPC	Mouse monoclonal, Santa Cruz, (sc32308)	1:3000
Staufen 1	Rabbit polyclonal, Proteintech (14225-1-AP)	1:500
Ncbp1	Rabbit polyclonal, Bethyl Labs (A301-794A)	1:2000
SafB	Mouse monoclonal, ThermoFisher (RG2236804)	1:1000
Secondary Antibody	**Technical specifications**	**Concentration**
Anti-mouse	BioRad	1:10000
Anti-rabbit	BioRad	1:10000
Alexafluor (488)	Invitrogen	1:10000
Alexafluor (594)	Invitrogen	1:10000

Chapter 3

Investigation of cytoplasmic functions of SRSF1 using CRISPR-mediated targeting in mouse ES cells

3.1 Introduction

3.1.1 CRISPR/Cas9 Gene Editing

To accurately understand the genetic mechanisms responsible for a given phenotype, an ability to test the cellular system in a controlled manner is required. Genomic mutagenesis and manipulation has enabled this from the advent of molecular biology, for example using chemical mutagens. However, this approach was essentially random with very little control over the mutagen target site. Traditional homology directed gene targeting strategies have been widely used to enable the production of knockout/knockin models. Although, not only are these methods are highly costly, time consuming and characterised by poor efficiency, but the targeted genomic loci often retained exogenous sequence as result, for example *loxP* recombination sites (Sander & Joung, 2014). Such problems were partially solved on a molecular level through the development of TAL effector nucleases (TALENs) and Zinc finger nucleases (ZFNs), which provided a tool to direct nuclease activity to specific loci; albeit in a fairly inefficient manner (Wang & Qi, 2016).

More recently, the development of the CRISPR/Cas9 gene editing tools have revolutionised both the editing field and molecular biology in general. In its endogenous function, CRISPR (clustered interspersed palindromic repeats)-Cas systems form the basis for an adaptable bacterial immune response against the invasion of foreign DNA (Figure 3.1). Fundamentally, all CRISPR/Cas systems consist of a site-specific RNA-directed DNA nuclease, although there are numerous types of CRISPR-Cas systems, that have evolved to serve diverse bacterial species. The first to be harnessed for gene editing applications and the most popular CRISPR/Cas system is a type-II system from *Streptococcus pyogenes* that utilises a Cas9 nuclease (Lewis & Ke, 2017).

In type-II CRISPR systems, the RNA guide (gRNA) exists within a duplex formed of tracrRNA:crRNA that forms base pairs with a target DNA sequence (Doudna & Charpentier, 2014). Once in contact with the target, the nuclease induces a double-stranded break (DSB) in the bound DNA. Importantly, for cleavage to occur the complex must recognise a short consensus region termed a Protospacer adjacent motif (PAM) site downstream of the base-paired DNA sequence to facilitate recognition of a correct target. Crystal structures of Cas9 in complex with its target DNA sequence have demonstrated that the PAM forms interactions with arginine residues in the C-terminal domain of Cas9. Mismatches in this site abolish cleavage as it is the necessary moiety for nuclease activity and DNA binding. Furthermore, PAM recognition stabilises the Cas9:DNA complex and encourages DNA melting to facilitate

gRNA binding (Doudna & Charpentier, 2014). For the *S.pyogenes* Cas9 nuclease, the PAM must have the consensus sequence -NGG, although there is some tolerance for the -NAG motif (Hsu *et al.*, 2013). Given the correct binding conditions, the nuclease induces a double-stranded break in the target DNA with highly predictable cleavage patterns due to the tight binding sequence constraints. More recently, spCas9 has been engineered to recognise alternate PAM sites thus facilitating targeting at non-NGG consensus regions (Komor *et al.*, 2016).

The Cas9 activity can be easily manipulated through the use of a sequence-specific gRNA template to direct the Cas9 complex to any genomic locus of choice (Ran *et al*. 2013). These qualities have revolutionised gene editing and hundreds of different applications of the system in diverse physiological contexts have already been demonstrated.

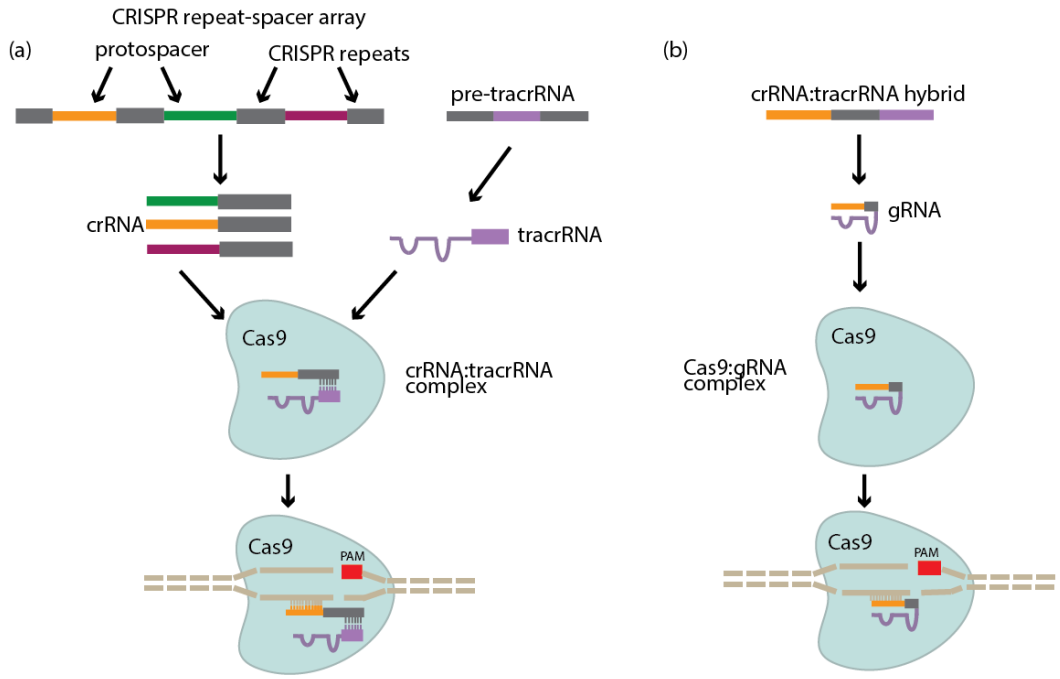


Figure 3.1: The *S.Pyogenes* CRISPR/Cas9 targeting system. (a) in the bacterial system, the crRNA is produced from the CRISPR-repeat array. A separate tracrRNA is required to bind each crRNA to facilitate Cas9 cleavage at the encoded sequence. (b) In the repurposed CRISPR editing system, the crRNA and the tracrRNA exist as a chimeric RNA that is incorporated into Cas9 to direct activity.

In the repurposed CRISPR/Cas9 toolkit, the bacterial crRNA and tracRNA are fused into a single chimeric gRNA that can direct nuclease activity; which is highly advantageous from both a technical and design standpoint (Cong *et al.*, 2013; Hsu *et al.*, 2013; Ran, Hsu, Lin, *et al.*, 2013). The guide RNA must be between 18 and 22 nucleotides in length with

perfect homology with the genomic target site. At the genomic locus, this site must be followed by the -NGG consensus PAM to facilitate cleavage by Cas9.

On the generation of a single stranded nick or a double-stranded break the genomic locus is repaired by the cell's DNA repair machineries. There are two mechanisms in which this could occur: non-homologous end joining (NHEJ) or homology directed repair (HDR). The route favoured is scenario-specific and can depend on many factors, but once committed to, is subject to tight regulation. For example, RAD51 is fundamental to repair pathway choice on generation of a DSB and tightly regulates HDR outcome (Ceccaldi *et al.*, 2016). Interestingly, RAD51 recruitment following a DSB can be influenced by the heterochromatic environment in which the break occurs. In pericentromeric heterochromatin, RAD51 is recruited only after DNA replication has occurred, whereas in centromeric heterochromatin its recruitment occurs throughout the duration of the cell cycle (Tsouroula *et al.*, 2016). This could influence gene editing studies as it implies that targeting efficiency may vary dependent on chromosomal location and epigenetic topography of the DNA at the target region.

In general, the NHEJ pathway, can occur via numerous mechanisms during which DNA lesions are repaired in an error prone manner (Bothmer *et al.*, 2017). It is a highly flexible process that can involve numerous factors which differ depending on the scenario of each DNA break. Furthermore, NHEJ is the predominant DNA repair mechanism when cells are not in S/G2 phase (Lieber, 2011). Usually, NHEJ is not the favoured outcome of a CRISPR targeting event as it can result in unwanted indels including premature termination codons, which may trigger mRNA degradation pathways such as NMD (Hendel *et al.*, 2015). Furthermore, NHEJ can also target each allele differently resulting in heterozygous indels (Shalem *et al.*, 2015). However, it can be useful when the objective is simply to ablate gene function.

In contrast to NHEJ, the HDR pathway proceeds with high fidelity in well characterised mechanisms; the major being route being homology repair (HR). HR requires the presence of a long region of homologous sequence to repair the lesion, hence its prevalence over NHEJ during S/G2 due to chromatid proximity (Lieber, 2011; Ceccaldi *et al.*, 2016). Indeed, cell synchronisation has been shown to increase HDR-mediated gene targeting in HEK293T cells (Lin *et al.*, 2014). HDR can be harnessed for gene editing to create customised insertions or base changes by providing the cell with a repair template that has sequence homology to the targeted genomic locus. There are several ways to supply the exogenous repair template be it DNA or RNA, double or single stranded and via transfection or viral

transduction; although it is unclear which method may be the most universally efficient (Graham & Root, 2015).

CRISPR/Cas9 targeting has been revolutionary in generating knockout models on both a gene specific and genome-wide level. Large libraries of gRNAs have been frequently used to implement genome-wide knockout screens. These have been particularly useful in contrast to shRNA library screening, since CRISPR approaches favour complete loss of function of a gene (Sander & Joung, 2014). This has been harnessed in diverse studies, particularly in cancer biology to identify novel drivers of carcinogenesis in numerous scenarios. The sequence-specific targeting ability of the CRISPR/Cas9 system has also been exploited to recruit other factors to the DNA, facilitated by catalytically dead Cas9 mutants (dCas9). These proteins do not possess nuclease activity but are still programmable with gRNA specificity. Most frequently, this has been used to modulate transcription whereby the dCas9 is fused to transcriptional activators/repressors that act on the target gene or surrounding sequence (Shalem *et al.*, 2015). Recently, Cas9 has been engineered to reduce off target effects (Slaymaker *et al.*, 2016).

3.1.2 Physiological roles of cytoplasmic SRSF1

The functions of SRSF1 in both canonical and alternative splicing have been extensively studied and are extremely well characterised. However, although SRSF1 is the prototype for shuttling SR proteins, the exact mechanisms and extent of cytoplasmic activity remain elusive in a physiological context.

SRSF1 is known to participate in several cytoplasmic roles, including mRNA translation (see Chapter 1). Briefly, SRSF1 enhances the mRNA translation of a specific subset of target transcripts including those encoding RNA processing factors, mitotic and cell-cycle associated factors. It does so through enhancing the phosphorylation of 4E-BP in an mTOR dependent manner, which causes the release of eIF4e to enable translation of its bound mRNA (Michlewski *et al.*, 2008; Maslon *et al.*, 2014).

A particularly useful tool to study the cytoplasmic SRSF1 functions is the chimeric SRSF1-NRS protein, which contains the potent nuclear retention sequence (NRS) from SRSF2 and is actively retained in the nucleus, representing a non-shuttling SRSF1. Complete ablation of SRSF1 function results in cell death and early embryonic lethality in animal models, due to a loss of the essential nuclear roles of SRSF1 (Xu *et al.*, 2005). The use of the SRSF1-NRS model has been crucial to delineate the cytoplasmic functions of SRSF1 as it

permits SRSF1 ablation in the cytoplasm while preserving nuclear SRSF1. The latter is sufficient to preserve cellular viability in conditional SRSF1 knockout mouse embryonic fibroblasts (MEFs) (Lin *et al.*, 2005; Xu *et al.*, 2005). In a physiological setting, wild type SRSF1 strictly controls its own homeostasis via an unproductive splicing feedback loop (Sun *et al.*, 2010). Forced overexpression of either SRSF1 or SRSF1-NRS may perturb this system and other roles of SRSF1, therefore its physiological relevance must be questioned.

In addition to using overexpression systems, previous work on SRSF1 in the cytoplasm has been limited to mostly human cell lines. Considering that SRSF1 exhibits a differential expression pattern across tissue types (Zahler *et al.*, 1993; Hanamura *et al.*, 1998), there are distinct possibilities that this has consequences for SRSF1 function. To date, little is understood regarding the tissue-specific roles of SRSF1 function in the cytoplasm or their physiological relevance, both on a cellular level and for the whole organism.

SRSF1 is essential for viability as demonstrated from both DT40 chicken cell and mouse models (Wang *et al.*, 1996; Xu *et al.*, 2005). However, there have been conditional SRSF1 knockout mouse models described that have differential effects. For instance, Cre-induced ablation of SRSF1 during the onset of cardiogenesis is not lethal and mice are born as normal, although it was noted by the authors that this phenotype is representative only of a partial knockout (Xu *et al.*, 2005).

3.2 Aims

To facilitate an understanding of SRSF1 function in the cytoplasm and its physiological relevance, work described here aims to create an endogenous system of non-shuttling SRSF1. To do this, CRISPR/Cas9 gene editing tools were designed to target the nuclear retention sequence (NRS) from the non-shuttling SR protein SRSF2 to the endogenous SRSF1 genomic locus. Mouse embryonic stem cells mESCs were targeted to generate cell lines that endogenously expressed either the T7-tagged SRSF1 or the nuclear retained version; the SRSF1-NRS-T7 fusion protein.

In summary, the main aims are as follows:

(1) To design and optimise a CRISPR/Cas9-based strategy to create an endogenous model of non-shuttling SRSF1 in mouse embryonic stem cells
(2) To investigate the phenotypic and physiologically relevant consequences of non-shuttling SRSF1 in the model cell system.

3.3 Results

The main aim of the following work was to target the nuclear retention sequence (NRS) from SRSF2, followed by a T7 tag to the endogenous SRSF1 locus using the CRISPR/Cas9 gene editing tools. The T7 tag is a small epitope tag consisting of 11 amino acids, which has been successfully used with SRSF1 for diverse biochemical applications and does not perturb canonical SRSF1 function (Cáceres *et al.*, 1997). Therefore, a T7 tag alone was targeted to mESCs to create an SRSF1-T7 knockin line. This acts as proof-of-concept for CRISPR targeting and efficiency of it at the SRSF1 genomic locus, but also serves as an invaluable tool to dissect SRSF1 function while maintaining endogenous levels of the protein. Complete SRSF1 knockout systems are inviable as they remove the essential, nuclear roles of SRSF1. The SRSF1-NRS-T7 knockin approach permits the study of the consequences of the absence of cytoplasmic SRSF1. Previous work conducted by others, suggest that the cytoplasmic functions of SRSF1 are not required for viability, at least in MEFs (Lin *et al.*, 2005).

To establish the SRSF1 knockin models, mouse ES cells were targeted as proof-of-concept prior to generation of the mouse model. Mouse ES cells are diploid with a stable karyotype compared to other cultured cell lines such as HeLa or HEK293T cells, making them suitable to generate physiological models of early development (approximately E3.5-E4.5). With addition of the correct growth factors, they can be maintained in culture in a pluripotent state and passaged indefinitely (Martello & Smith, 2014). Furthermore, mESCs harbour an open chromatin state, which may facilitate nuclease binding and increase chances of successful HDR on generation of DSBs (Chen, Rinsma, *et al.*, 2016). In this chapter, the generation of SRSF1-NRS-T7 and SRSF1-T7 systems will be discussed in detail.

3.3.1 CRISPR/Cas9 targeting strategy

To generate the SRSF1-NRS-T7 and SRSF1-T7 cell lines, several rounds of CRISPR targeting were undertaken with various modifications at each stage to improve targeting efficiency. A summary of the workflow used to generate the corresponding clones is shown in Figure 3.3 and targeting attempts undertaken are shown in Table 3.1.

The NRS-T7 or T7 insertions were targeted to the C-terminus of SRSF1, downstream of the final exon (exon 4) of the genomic sequence. Protein sequence between human and mouse is identical but the genomic regions are different, thus all targeting was designed specifically for the murine SRSF1 genomic locus (Hanamura *et al.*, 1998). To ensure

translational readthrough into the inserted sequence, targeting constructs were designed to remove the canonical stop codon of the SRSF1 transcript on a successful integration event. This strategy should not hamper mRNA processing of the transcript as use of the SRSF1-NRS fusion construct has been previously documented to produce a functional protein (Cazalla *et al.*, 2002).

The NRS sequence is naturally present at the C-terminus of SRSF2, so the described targeting approach maintains a similar endogenous positioning of this domain. It is unknown whether, for example, proximity to an SR domain is required for its retention function. Furthermore, considering that the RRM domains of SRSF1 are important for RNA binding, particularly in splicing, it is prudent to minimalise any potential interference that the NRS may have on these roles. Positioning the NRS sterically away from such domains may fulfil this requirement; although once established, the cellular models should be analysed by RNA-seq to ensure this is the case and splicing is unaffected. Given that the molecular basis of TAP-binding is mediated through the second RRM, it is unlikely that SRSF1-NRS-T7 affects mRNA export therefore, in the CRISPR model, mRNA export should be unaffected.

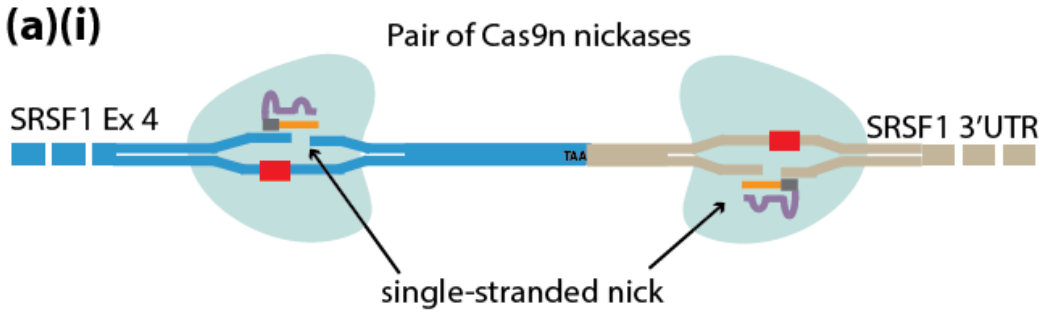
3.3.2 CRISPR/Cas9 design

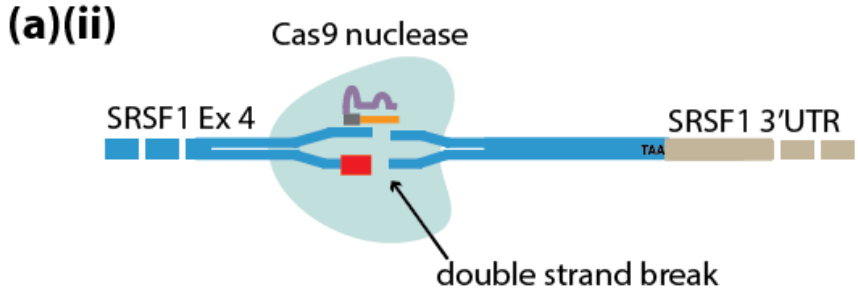
In the initial targeting strategy, the CRISPR/Cas9 nickase system was used in mouse ES cells (E14s). The nickase system, utilises a Cas9 catalytic mutant (Cas9n) that contains a single point mutation in the RuvC1 domain of the nuclease. Specifically, the D10A Cas9n was used although there are other catalytically inactive mutants available (Sander & Joung, 2014). Cas9n is unable to induce a DSB but facilitates single-stranded nicking of the target DNA (Ran *et al*. 2013). This approach requires a pair of Cas9n enzymes to generate two single stranded nicks on opposing strands of the genomic DNA. It has been demonstrated that the D10A nickase can generate breaks that are repaired by HDR (Bothmer *et al.*, 2017). The SRSF1 genomic locus was targeted through the action of a pair of single gRNAs designed to facilitate double nicking in close proximity to Exon 4 (Figure 3.2). All gRNAs used contained overhangs to facilitate Bbs1 directed restriction site cloning into the pX461 vector (Ran *et al*. 2013), which also contains GFP to enable FACs sorting of transfected cells. Each guide sequence was designed to be 18-22 nucleotides in length with an additional guanine residue upstream top facilitate transcription by the U6 promoter in the plasmid.

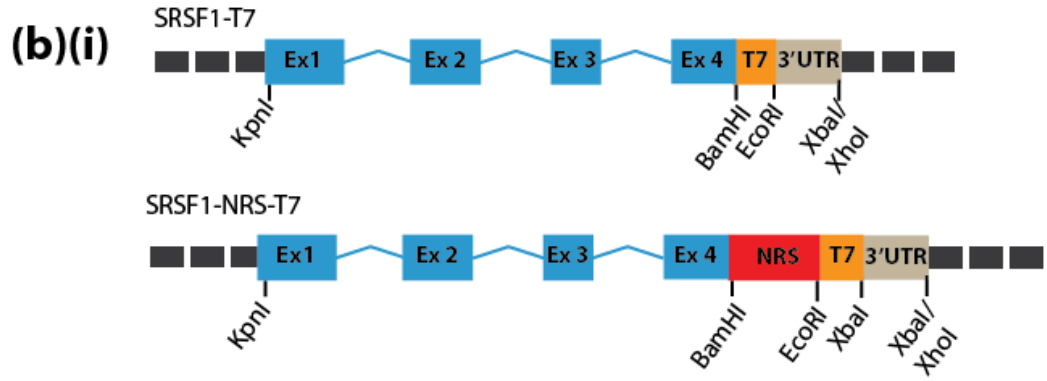
When this approach was initially designed, it was proposed that the Cas9 nickase has fewer off-target effects than its nuclease counterpart (Ran *et al*. 2013). Latterly, however, comprehensive studies have failed to document this as a global rule for Cas9n targeting events

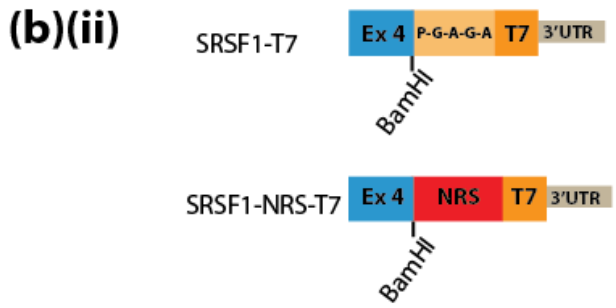
and it has been proposed that single paired nicks can result in a low frequency of HDR is some cell types (Komor *et al.*, 2016). Therefore, further targeting events, were directed by the Cas9 nuclease, which utilises a single gRNA to generate a double stranded break. This is useful when the genomic sequence of interest is not amenable to Cas9 gRNA design or guide performance is poor for this region. Logically, this system should also increase the likelihood of a successful HDR-mediated targeting event as only a single cleavage by Cas9 is required (Komor *et al.*, 2016).

To implement the Cas9 nuclease approach the gRNA was cloned into the pX458 targeting vector, which also contains GFP for downstream FACs sorting (Ran *et al*. 2013). The gRNA used in the nickase approach was predicted to be a highly robust sequence for targeting the desired region using the Zhang Lab gRNA design tool and had generated a knockin clone. Thus it was used again here with an additional G upstream as this has been shown to reduce off-target effects when using the nuclease approach (Cho *et al.*, 2014). After CRISPR/Cas9 cleavage to generate a double-stranded DNA break, the cell repairs the locus through either through non-homologous end joining (NHEJ) or homology directed repair (HDR). Integration of an insert of choice following cleavage requires successful HDR as it ensures a high-fidelity repair and reduces indel formation via NHEJ. The repair template used for CRISPR targeting experiments must contain arms of homologous sequence to the locus targeted to facilitate HDR, in this case the C-terminal and 3'UTR regions of SRSF1 (Figure 3.2). The length of the arms should be a minimum of 60bp, although this is proportional to the size of the insert; the longer the insert, the longer the homology arms for efficient insertion. This maximises the chance of the cell recognising the supplied repair template as self-DNA, which it should use to repair DSBs at homologous regions.


(a)(i)
Pair of Cas9n nickases
SRSF1 Ex 4
TAA
SRSF1 3′UTR
single-stranded nick


(a)(ii)
Cas9 nuclease
SRSF1 Ex 4
TAA
SRSF1 3′UTR
double strand break


(b)(i)
SRSF1-T7
Ex1
Ex 2
Ex 3
Ex 4
T7
3′UTR
KpnI
BamHI
EcoRI
XbaI/
XhoI
SRSF1-NRS-T7
Ex1
Ex 2
Ex 3
Ex 4
NRS
T7
3′UTR
KpnI
BamHI
EcoRI
XbaI
XbaI/
XhoI


(b)(ii)
SRSF1-T7
Ex 4
P-G-A-G-A
T7
3′UTR
BamHI
SRSF1-NRS-T7
Ex 4
NRS
T7
3′UTR
BamHI

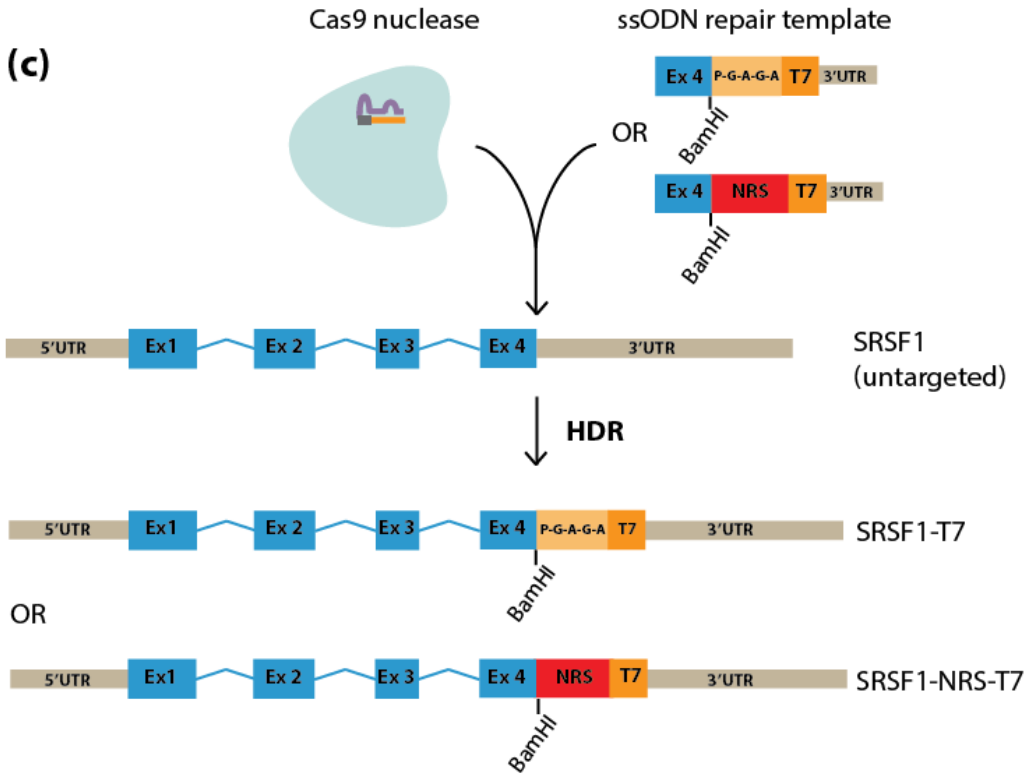


Figure 3.2: CRISPR/Cas9 strategy to target the SRSF1 genomic locus (a)(i) Initial approach undertaken using the Cas9 nickase and a pair to gRNAs. (ii) Optimised strategy using the Cas9 nuclease. Note the same sequence gRNA as in one of the nickase gRNAs was used here. (b) Repair template design (i) Initial targeting repair template as a dsDNA plasmid. Black dashed lines indicate the PCDNA3+ cloning vector sequence that contained the insert for targeting. The entire SRSF1 sequence including intros was used to create long arms of homology. (ii) Optimised ssODN design in latterly used targeting. Only a partial sequence of SRSF1 Exon 4 was included in shorter homology arms. Note the SRSF1-T7 repair template contains a small peptide linker (P-G-A-G-A) as rationalised in the main text (c) Summary schematic of optimised targeting design used to generate clones.

3.3.2.1 Repair Template Design

The repair templates used to target SRSF1 for initial CRISPR targeting experiments were double stranded DNA plasmids that contained large regions of homology with the genomic SRSF1 locus. To generate these, the SRSF1 gene was PCR amplified from mouse genomic DNA (primer sequences Table 2.1). Both intronic and 3'UTR sequence was included in this plasmid to maximise the length of the homology arms for targeting and to maintain proper post-transcriptional processing of the SRSF1 transcript if a larger proportion of the

vector was integrated. This is particularly important for SRSF1 so as not to perturb the autoregulatory feedback loop or the generation of alternatively spliced isoforms. The nuclear retention sequence from SRSF2 was also amplified from mouse genomic DNA. The T7 tag sequence is short and exogenous, thus annealed oligos were used for cloning in the targeting vector. Integration of PCR products into the PCDNA3+ plasmid was carried out using restriction site cloning or ligation independent cloning. The resultant plasmid contained SRSF1 followed by the NRS and the T7 epitope tag sequence or SRSF1 followed by the T7 tag alone. The sequence fidelity of targeting vectors was confirmed by Sanger sequencing (data not shown), and expression confirmed using western blotting following transient transfection (Figure 3.4).

3.3.2.2 Screening primer design

Screening primers were designed to amplify the insertion and additional sequence outside the region targeted to ensure that targeting occurred at the correct genomic locus. The resultant amplicon contains a large proportion of 3'UTR sequence, as this region is ultraconserved and must be intact to maintain the autoregulatory feedback loop for the SRSF1 transcript (Sun *et al.*, 2010). It is crucial that SRSF1 homeostasis is not perturbed in targeted cells as improper expression can have aberrant consequences such as cellular transformation (Karni *et al.*, 2007). Amplification of integration events using the primers designed are shown in Figure 3.3.

3.3.3 CRISPR targeting and Optimisation in mESCs

The first round of CRISPR targeting using the initial described strategy in mESCs resulted in the survival of few clones (32 and 28 for SRSF1-NRS-T7 and SRSF1-T7 respectively). It is possible that this is due to poor transfection efficiency or extensive cell death after FACs sorting. To overcome this, an alternative flow cytometer was used which exerts less pressure on the cells during the sorting process.

PCR-screening of clones obtained from the first round of targeting indicated that CRISPR targeting had occurred to generate what appeared to be both NRS-T7 and T7 integration events. However, these represented false positive events, as after clonal expansion insertions were no longer present. It is likely that the double stranded repair template persisted in the cells in the initial stages of clone growth, which was PCR amplified in the screening

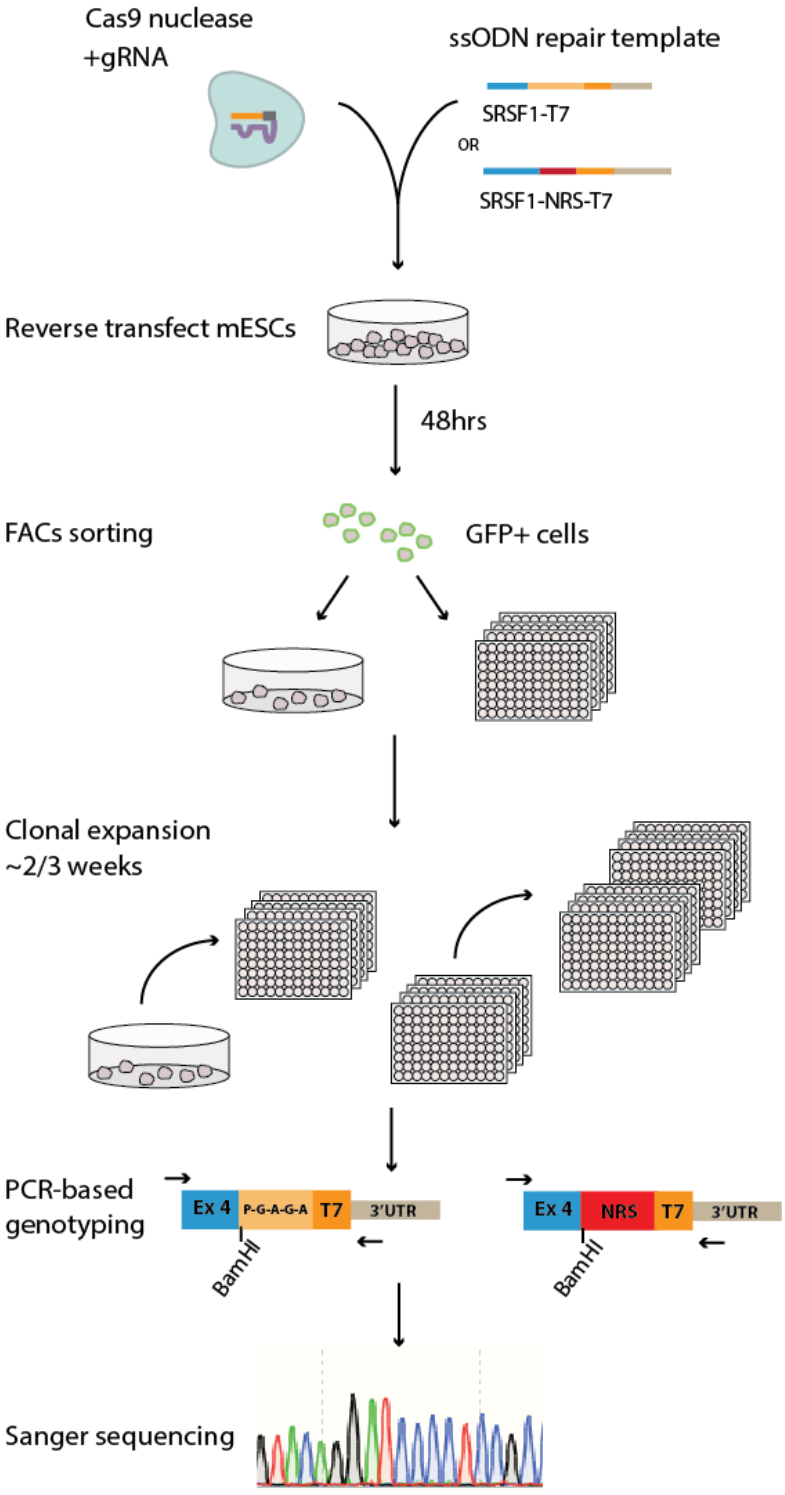


Figure 3.3: Overview of workflow used for CRISPR targeting strategy in mESCs. Successfully transfected cells were selected for using FACs sorting for GFP+ cells (the Cas9 plasmid contains GFP). Clones were expanded from single cells over a maximum of three weeks, depending on growth rate, before genomic DNA was screened for the insert using PCR and downstream sequencing analysis to verify clones.

process. Furthermore, sequencing revealed that Cas9 cleaved the genomic DNA and the transfected repair template extensively to leave indels at the PAM site of the gRNA consensus region. To prevent this in future rounds of targeting, a single nucleotide of the PAM site in the targeting vector was mutated using site directed mutagenesis, which is sufficient to ablate Cas9 cleavage (Ran *et al*. 2013). The directed mutations do not cause a change in amino acid sequence and mutated targeting vectors were expressed in cells as previously described (Figure 3.4 (b)).

Targeting was repeated using the modified repair templates and 96 clones for each SRSF1-NRS-T7 and SRSF1-T7 targeting were screened by PCR. Several of these were chosen for further investigation by western blotting (Figure 3.4 (c)). No positive clones were obtained for the SRSF1-NRS-T7 insertion. A single SRSF1-T7 clone was obtained (clone C3 42), however, it was unusable as the epitope tag, though expressed, was not detected by the T7 antibody (data not shown). This problem was unforeseen as the epitope does not behave this way when fused to the N-terminus of SRSF1. It is possible that at the C-terminus, the tertiary structure of the SRSF1-T7 fusion protein results in masking of the epitope tag as it becomes buried within the protein. However a larger SRSF1 protein could be identified by the SRSF1 mAb96 (Figure 3.4 (c)) which recognises the N-terminal region of SRSF1 in a phosphorylation independent manner (Hanamura *et al.*, 1998).

To ensure detection of the T7 tag upon integration at the SRSF1 genomic locus, several new targeting constructs were designed that contained variants of the T7 epitope. Either a small peptide linker (P-G-A-G-A) or multiple T7 (x2 or x3) tags were added prior to the tag sequence (Figure 3.4(d)(i)). Inclusion of a proline residue in the peptide linker sequence was an attempt to create steric changes in the peptide chain to reveal the epitope from burial in SRSF1. As proof of principle, these were cloned into the PCDNA3+ vector and expression in HEK293T cells was confirmed by western blotting (Figure 3.4(d)(ii)).

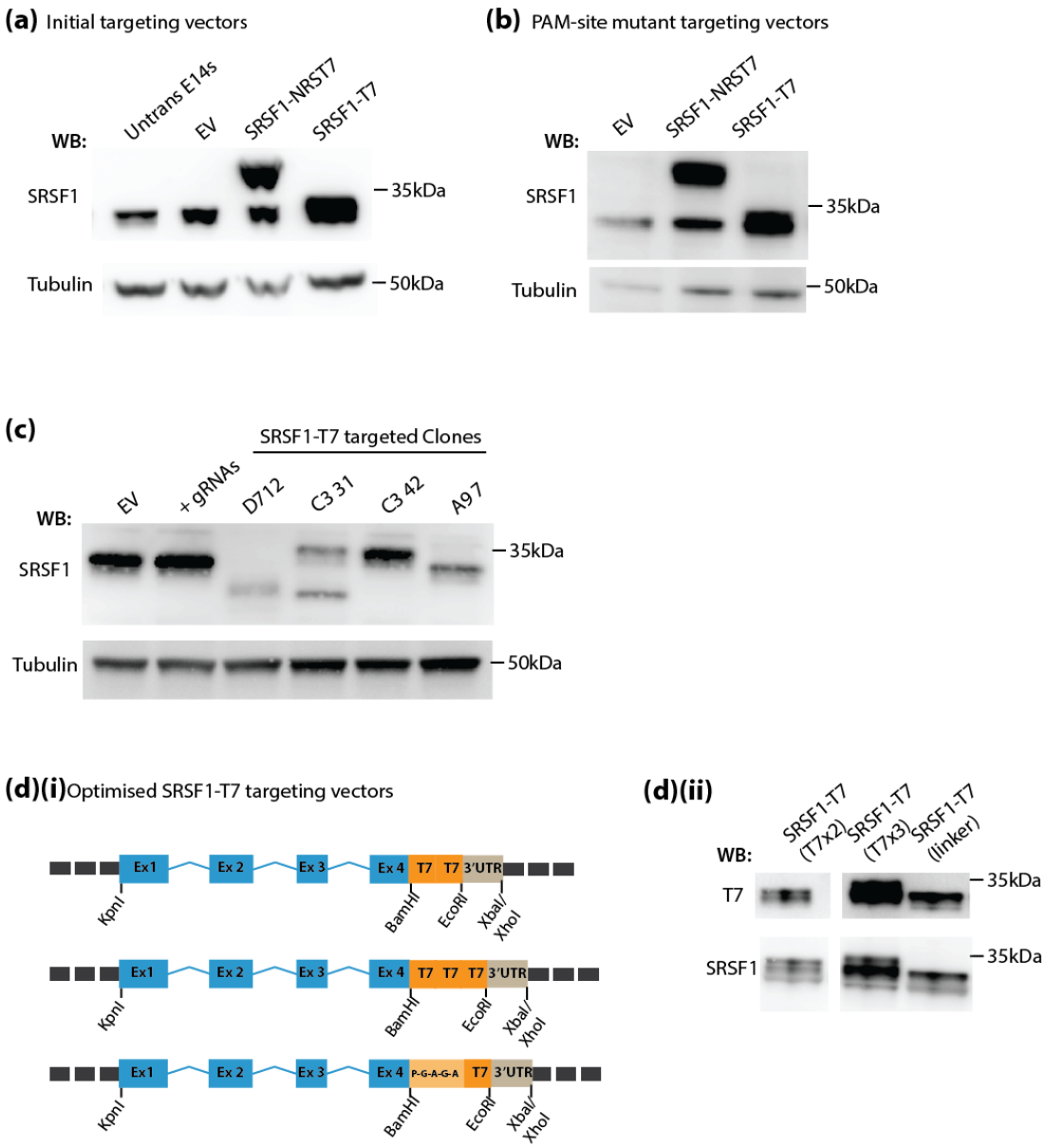
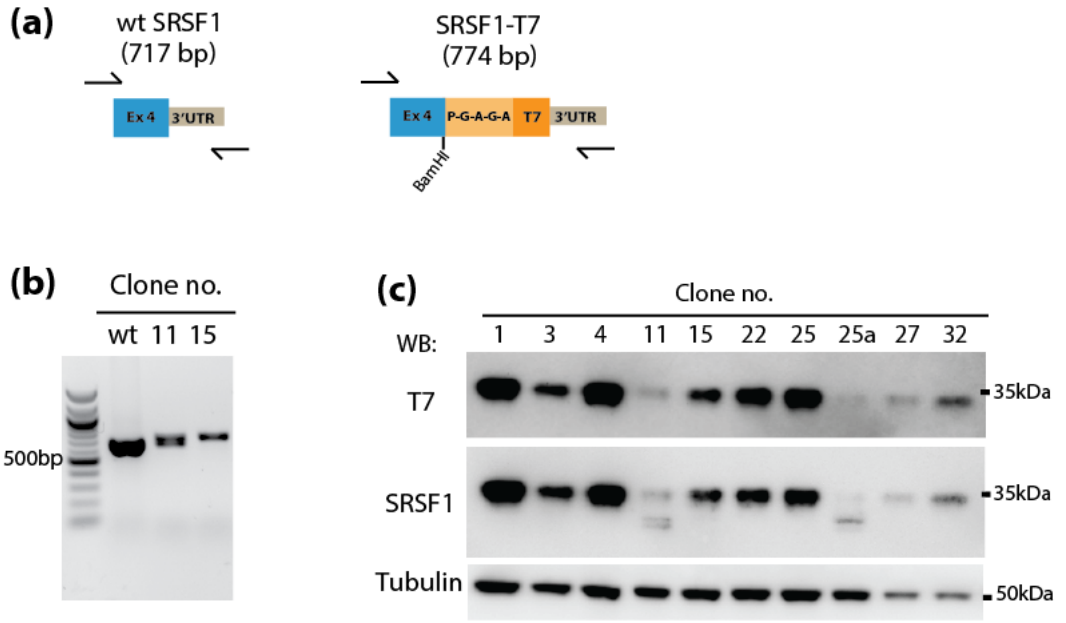


Figure 3.4: Optimisation of homology repair templates for CRISPR targeting. (a) Transient overexpression of initial repair templates generated. Western blotting with SRSF1 antibody to confirm expression in mESCs. Blotting with tubulin antibody is shown as a loading control (b) Western blotting to demonstrate expression of repair template subsequent to PAM-site optimisation in mESCs. Blotting with tubulin antibody is shown as a loading control (c) mESC clones generated with the genomic SRSF1-T7 insertion that is undetectable by the T7 antibody (blank blot for T7 antibody not shown). For example, clone C342 contains a homozygous SRSF1-T7 as a band larger than that of wt SRSF1 in the EV (empty vector) or gRNA only controls can be detected with the SRSF1 antibody. Blotting with tubulin antibody is shown as a loading control (d)(i) Schematic outlining repair template design to ensure the T7 tag is detectable with the T7 antibody. (ii) Expression of optimised targeting vectors in HEK293T cells as detected by Western blotting with the T7 antibody.

3.3.4 Generation of homozygous SRSF1-T7 mESC clones

To improve targeting efficiency, single stranded DNA oligonucleotides (ssODNs) were designed for use as the repair templates for both SRSF1-T7 and SRSF1-NRS-T7 targeting (Figure 3.2(b)(ii). These oligos contain the appropriate PAM site mutation. In addition, Cas9 nuclease was used to direct DNA breaks at the C-terminus of SRSF1. With this approach, it was possible to successfully generate several homozygous SRSF1-T7 knockin clones as confirmed by sequencing of the genomic SRSF1 locus (Figure 3.5).



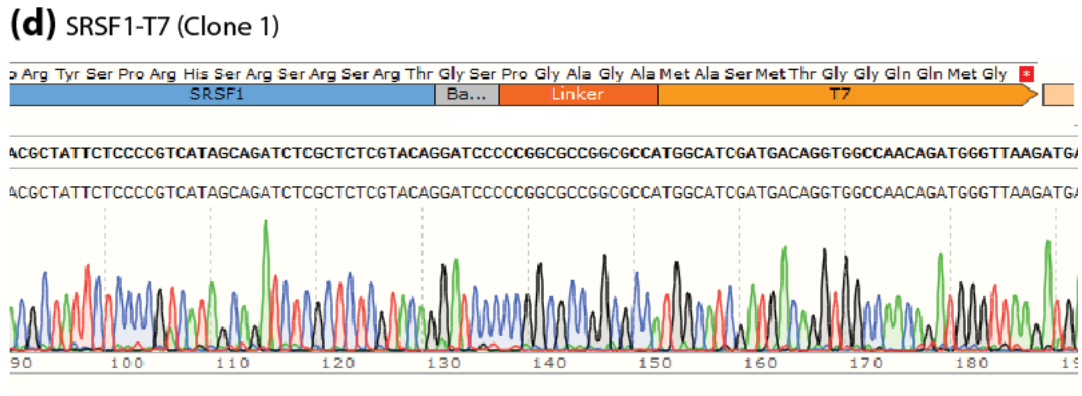


Figure 3.5: Screening of SRSF1-T7 knockin mESCs. (a) Outline of PCR screening strategy in the context of wt SRSF1 or SRSF1-T7 knockin cells. The approximate size of the amplicon is indicated in bp. (b) Example of gDNA screening of two knockin clones (11 and 15) alongside wt mESCs for comparison using the outlined PCR screen. (c) Western blotting to demonstrate that SRSF1-T7 knockin clone express the correctly sized protein that can be detected by SRSF1 and T7 antibodies. Homozygous and heterozygous clones (e.g. 1 and 11 respectively were generated. (d) Example of correct Sanger sequencing verification for knockin clone 1.

3.3.5 Generation of homozygous SRSF1-NRS-T7 mESC clones

The optimised CRISPR targeting strategy was implemented to generate SRSF1-T7 clones with good efficiency and relative ease. This demonstrates that the SRSF1 genomic locus is targetable at the given site in Exon 4. In a subsequent round of targeting, an increased number of clones were screened with the rationale that targeting efficiency and HDR-mediated repair of NRS-T7 is much lower than for the tag alone. This may be feasible, considering the insertion is 90 bp larger in size.

Using this approach, what appeared in genomic DNA screening as a single homozygote SRSF1-NRS-T7 clone was obtained (Figure 3.6). However, downstream analysis of the protein and sequence data revealed that this clone expressed an almost wild type SRSF1 protein. This clone, referred hereafter as Clone 1, is the closest to a homozygous knockin in mESCs that could be generated in this work. It harbours an NRS-T7 insertion on both alleles, illustrating that targeting of the insert is possible. However, one of the alleles contains two single base pair deletions that create a frameshift mutation resulting in a premature termination codon that prevents expression of the SRSF1-NRS-T7 fusion protein from this allele (Figure 3.6). Clone 1 expresses an intact SRSF1-NRS-T7 protein from a single allele, thus can be considered heterozygous for the insertion, which provides a useful tool to study SRSF1 cytoplasmic function.

Western blotting demonstrates that the truncated protein produced from allele 2 is not in stoichiometry with the reciprocal from allele 1. This is reproducible and is not an artefact of the blotting signal, but possibly due to reduced protein production from this allele. Concurrent with this, a late shift in open reading frame which disrupts the UTR sequence, such as that occurs from the indels in allele 1, has been shown to reduce protein production from the transcript (Arribere *et al*. 2016). Such translational readthrough, by bypassing the stop codon, can produce less protein but does not trigger non-stop decay as there is no connection with poly(A) sequence (Arribere *et al.*, 2016). Alternatively, as a premature

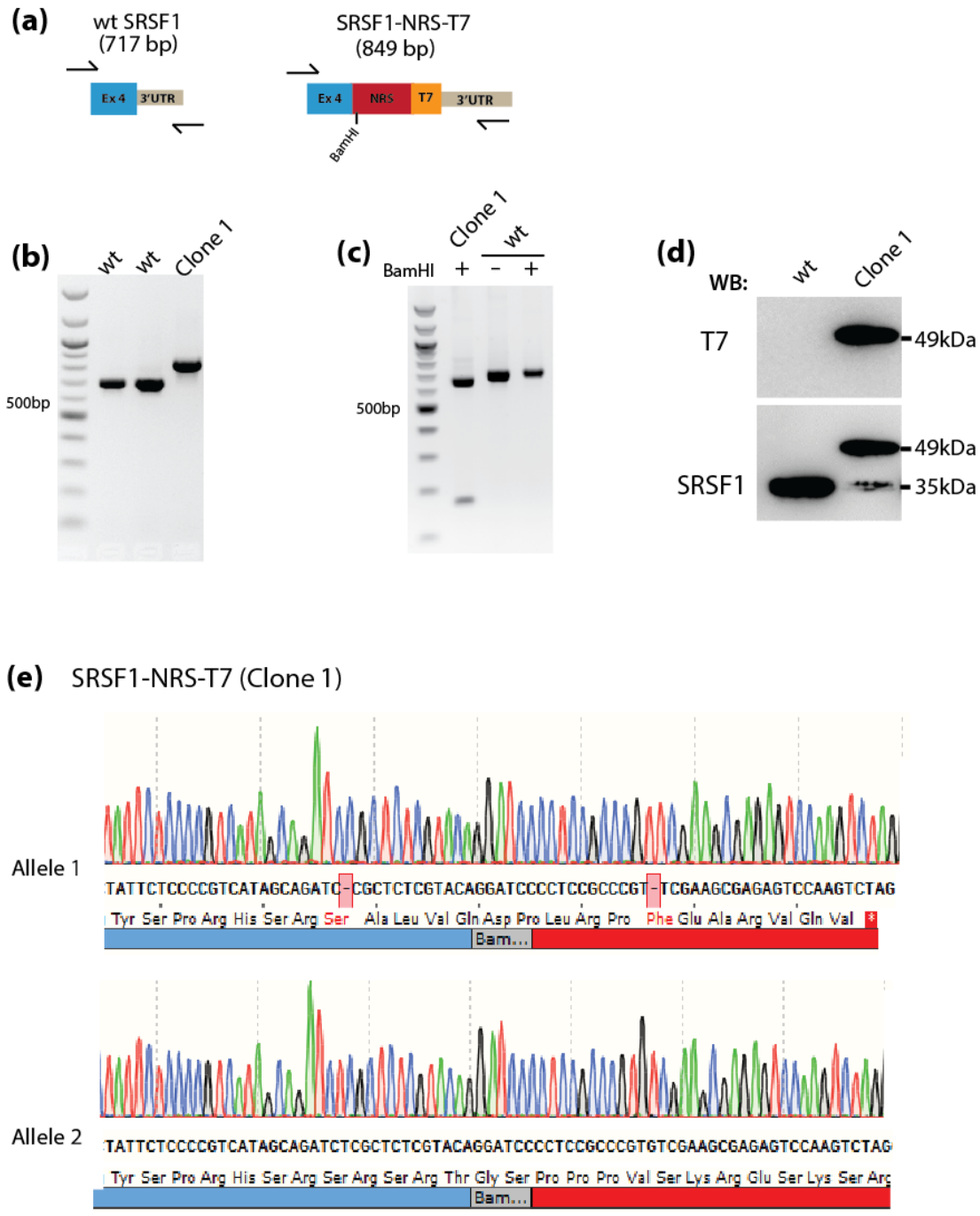


Figure 3.6: Screening of SRSF1-NRS-T7 knockin mESCs. (a) Outline of PCR screening strategy in the context of wt SRSF1 or SRSF1-NRS-T7 knockin cells. The approximate size of the amplicon is indicated in bp. (b) Example of gDNA screening of Clone 1 knockin clone alongside wt mESCs for comparison using the outlined PCR screen. (c) BamHI diagnostic digest of Clone 1 cells (producing products of 167 bp and 682 bp) or wt cells (do not contain BamHI site so are undigested). . (d) Western blotting to demonstrate that SRSF1-NRS-T7 knockin clone is not homozygous as indicated by gDNA screening, but expresses wt SRSF1 protein as detected by the SRSF1 antibody. (e) Sanger sequencing verification for knockin clone 1 of both alleles. PCR product subcloned prior to sequencing.

to degradation by NMD causing a reduction in protein expression (Hug *et al.*, 2015). Depletion of the core NMD factor UPF1 in these cells may help delineate this in the future.

3.3.6 Frequency of homozygous SRSF1-NRS-T7 alleles decline over time

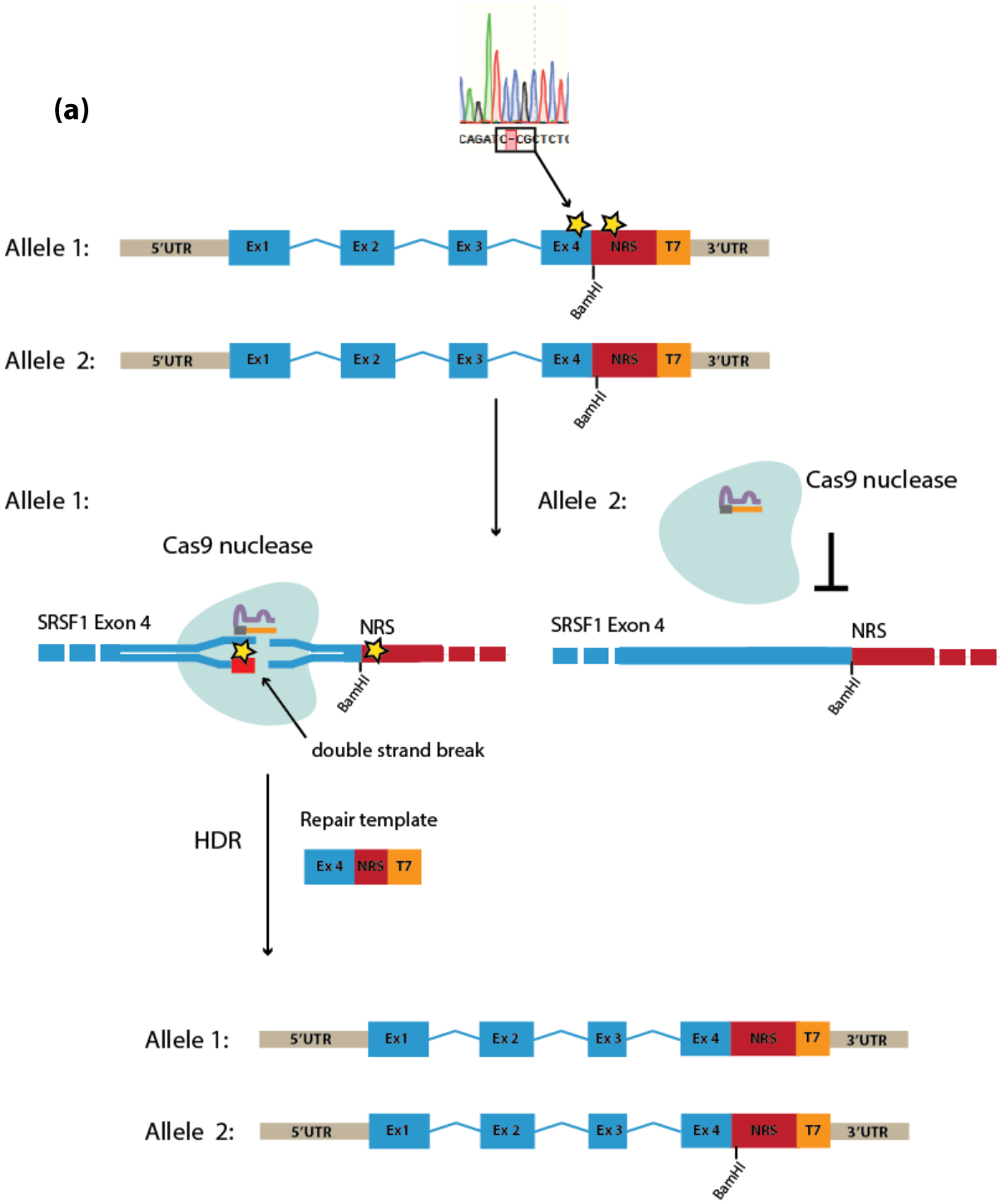
Targeting attempts in multiple cell types failed to generate SRSF1-NRS-T7 homozygous knockin lines, with the mESC Clone 1 cells being the closest achieved (Table 3.1). Considering this, one could conclude that there is significant negative selection against the generation of the desired knockin. If the NRS-T7 insertion is lethal in undifferentiated cells those that are successfully targeted initially, may die during the period of clonal expansion, (approximately 16-21 days). It is an attractive possibility that the cytoplasmic roles of SRSF1 are required in these cells for viability, perhaps for the translational regulation of essential mESC-specific mRNA transcripts. However, it is also feasible that the observed effect is due to extremely low targeting efficiency or due to technical reasons.

Homozygous SRSF1-NRS-T7 clones	SRSF1-T7 clones	Cas9 used	Problems	Modifications to Protocol
0/128	1/124	Nickase	T7 tag not recognised by T7 antibody	Insert peptide linker upstream of T7-tag. Switch to ssODN repair template.
1/441	10/126	Nuclease	Unsuccessful for SRSF1-NRS-T7	ssODN repair template for SRSF1-NRS-T7 targeting

Table 3.1: Overall summary of CRISPR targeting in mESCs. Numbers indicated are the total achieved encompassing all stages of the optimisation processes except changing from Cas9 nuclease to nickase as this is the most drastic protocol alteration implemented.

To investigate the possibility that targeted mESCs cells die during clonal expansion, SRSF1-NRS-T7 Clone 1 cells were retargeted in an attempt to repair the indels on the compromised SRSF1-NRS-T7 allele. Remarkably, the 5' proximal single bp deletion in allele 1 generates a novel Cas9 PAM recognition site that is not present in the second allele. Since the –NGG PAM consensus is the only sequence absolutely required for Cas9 cleavage, this was harnessed to permit design of a new gRNA sequence to specifically target the

compromised allele 1 but not the intact SRSF1-NRS-T7 allele 2. In addition, a new repair ssODN was used which contains an additional base pair change that destroys the BamHI site to aid the detection of repair in the population during downstream analyses. To monitor the frequency of repaired allele 1 in the population, genomic DNA was harvested throughout the period of clonal expansion and the region of interest subject to high throughput sequencing analysis (Figure 3.7).



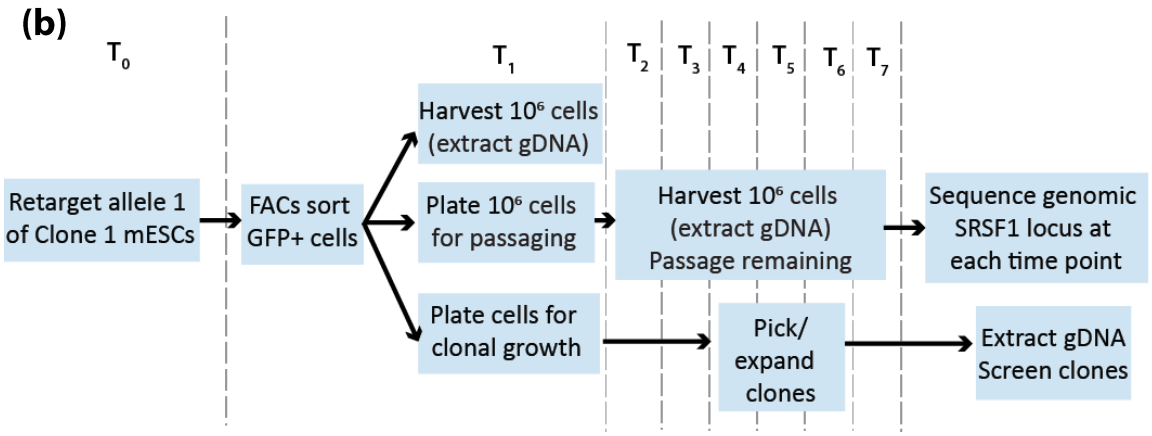


Figure 3.7: Retargeting Clone 1 SRSF1-NRS-T7 cells and the timeline used. (a) Clone 1 cells were retargeted using Cas9 nuclease and a gRNA specific for allele 1 targeting which will not cleave allele 2. The novel PAM site generated by the indels in allele 1 is indicated. A new ssODN repair template was used to mutate, and so remove, the existing BamHI site to remove to detect successful retargeting events. (b) Timeline of workflow used for retargeting. Genomic DNA was harvested from cells over the indicated time course. Concurrent with this, clones were expanded from single cells as for initial targeting strategies.

All data analysis for this experiment was carried out by Martijn Kelder of Andrew Wood's Lab at the IGMM. The allele frequency in the population was calculated by extracting the number of reads which contain the entire repaired sequence, including the destruction of the restriction site. The first time point, prior to transfection of the CRISPR machinery (t=0) represents the base line frequency of the alleles in the population, which is approximately 1:1. As expected, the proportion of reads generated from allele 2 is constant over the time course, as it is not targeted by Cas9 in this experiment (Figure 3.8 (a)). There are slight fluctuations over time; however, this is most likely due to PCR or pipetting-based error. In contrast, upon CRISPR targeting (t=1), there is a drastic reduction in the frequency of allele 1 reads in the population, indicating that allele-specific targeting using CRISPR-Cas9 was successful and led to repair. In theory, if targeting leads to repair of the locus, a reduction in allele 1 frequency should be accompanied by an increase frequency of reads containing the restriction site change; indicative of successful homology directed repair. Initially, this is the case and repair occurs at a frequency of around 1% but decreases over the time course to barely detectable levels in the population (Figure 3.8 (b)). Therefore, homozygous SRSF1-NRS-T7 cells are generated but most likely die during clonal expansion due to negative selection pressure, substantiating the data obtained by single clone analysis.

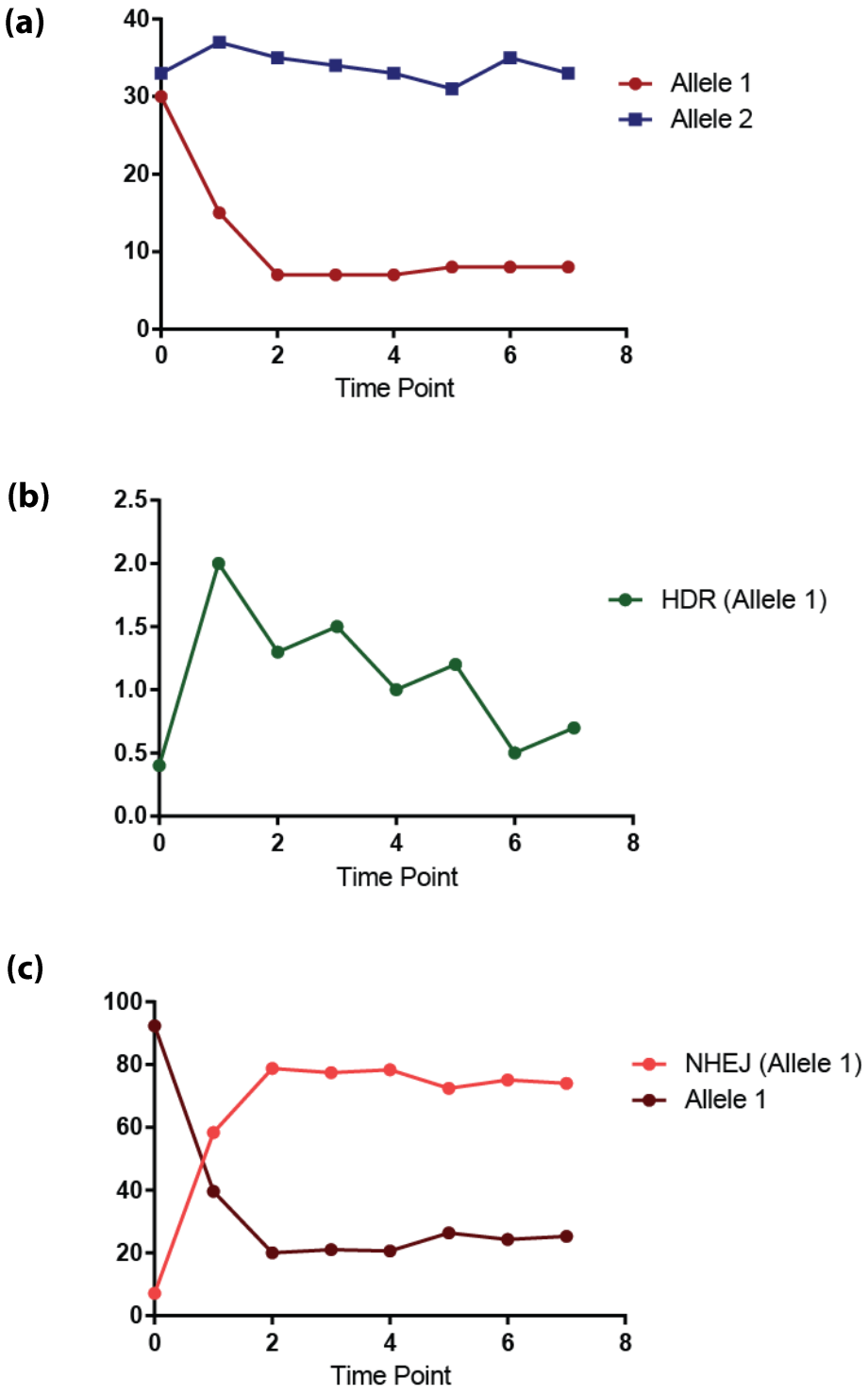


Figure 3.8: Repair of SRSF1-NRS-T7 is possible but declines rapidly over time (a) Over time the frequency of allele 1 decrease to a steady level of expression within the population. (b) Homology directed repair inferred by reads containing the repaired sequence. This peaks after targeting but declines to baseline frequency. (c) NHEJ events occur with high frequency after targeting and are maintained in the population. Reads were normalised to allele 2 in this instance which remains untargeted in the strategy used. *Data were generated (prior to plotting) by Martijn Kelder of Andrew Wood's Lab at the IGMM.*

It is possible to infer the frequency of NHEJ events in the population of cells after targeting, as represented by reads that do not match unrepaired or repaired allele 1 sequence when normalised to allele 2 (figure 3.8). Using this approach, the pattern of NHEJ events is the converse for HDR and rapidly increases to a sustained plateau. These cells most likely contain additional compensatory mutations to preserve expression of a wild type SRSF1 protein.

In addition to high throughput sequencing, the same pool of Clone 1 cells were subject to clonal growth and screening analogous to the approach used to generate Clone 1 cells. The rationale being that if targeting is extremely inefficient, screening more clones may eventually generate SRSF1-NRS-T7 homozygotes. From screening an additional 430 manually picked clones, preliminary analysis of the sequenced PCR amplicon indicates that a single clone was successfully repaired and persisted during the period of clonal expansion. This demonstrates that the SRSF-NRS-T7 homozygous insertion is viable in mESCs.

Overall, these data indicate that *de novo* generation of SRSF1-NRS-T7 would only be observed if over >1000 clones were initially screened, as on retargeting cells already contain one copy of the SRSF1-NRS-T7 knockin. Interestingly, there also are several clones that are partially repaired or those that are completely repaired over the region tested, but contain additional mutations that result in a new premature termination codon. In summary, there is a high degree of selection against generation of homozygous SRSF1-NRS-T7 mESCs using the described CRISPR/Cas9 targeting approach.

3.3.7 Targeting differentiated cell lines to generate SRSF1-NRS-T7 knockin clones

Data from the Clone 1 retargeting experiment and failure to target hundreds of manually screened clones demonstrate that there is strong negative selection towards generation of SRSF1-NRS-T7 in mESCs. The SRSF1-NRS-T7 mutation may be unfavourable when cells are maintained in an enforced pluripotent state. To determine if this is the case, the SRSF1-NRS-T7 knockin was targeted to two cell types that represent different stages during differentiation.

Firstly, mouse neural stem cells (mNSCs) were targeted using the optimised CRISPR design strategy to generate either SRSF1-T7 or SRSF1-NRS-T7 cells. These are further differentiated in comparison to embryonic stem cells as they are committed to the neural lineage, although they are not terminally differentiated and can still be maintained indefinitely

in culture (Conti *et al.*, 2005). Furthermore, they have been shown to be amenable to effective CRISPR/Cas9 targeting (Bressan *et al.*, 2017). In comparison to mESCs, neuronal stem cells exhibit a longer cell cycle to mESCs with an increased G1 phase (Pauklin & Vallier, 2013). Therefore, if targeting problems in mESCs are due to cell cycle problems this may circumvent that. For example, ablation of SRSF1 function in translation may exhibit different consequences for cell cycle regulation on transition from the pluripotent state. This is feasible considering that SRSF1 is capable of specifically regulating mRNAs involved in cell cycle and mitotic processes (Maslon *et al.*, 2014).

Following a series of experiments, successful targeting of the NSCs with the SRSF1-NRS-T7 insertion could not be achieved. This is likely due to several technical problems, including those with FACs-based sorting methods which prevented screening of large numbers of clones; the few colonies that were screened did not contain either the NRS-T7 or T7 insertion (data not shown).

Secondly, mouse NIH3T3 cells were similarly targeted with each insertion. NIH3T3 cells are fibroblasts and thus mimic most aspects of a fully differentiated cell type, although they are immortalised by spontaneous transformation on SV40 infection. These did not represent an ideal candidate for the creation of CRISPR/Cas9 knockins as they can have a variable karyotype. However, in theory, if the roles of cytoplasmic SRSF1 were required for the maintenance of stemness, then differentiated systems such as 3T3 cells should be amenable to targeting with the NRS-T7 insertion. Again, targeting data were inconclusive as numerous technical issues prevented large numbers of clones from being screened, such as poor survival rates on single-cell clonal expansion; but no knockins were obtained (data not shown).

3.3.8 Cellular localisation of knockin clones

SRSF1 is the prototype for shuttling SR proteins and moves continually between the nucleus and the cytoplasm at a steady state. The T7-tag at the N-terminus does not perturb this movement as demonstrated by heterokaryon assay when the T7-SRSF1 fusion protein is overexpressed (Cáceres *et al.*, 1998). To determine the cellular location of SRSF1 in the mESC lines generated, immunofluorescence was performed against the endogenous T7 tag (Figure 3.9). Both the SRSF1-T7 and the SRSF1-NRS-T7 (Clone 1) cells were used for comparison. Although Clone 1 cells do express wt (shuttling) SRSF1, this protein is not detected in this assay as the T7 tag is not expressed.

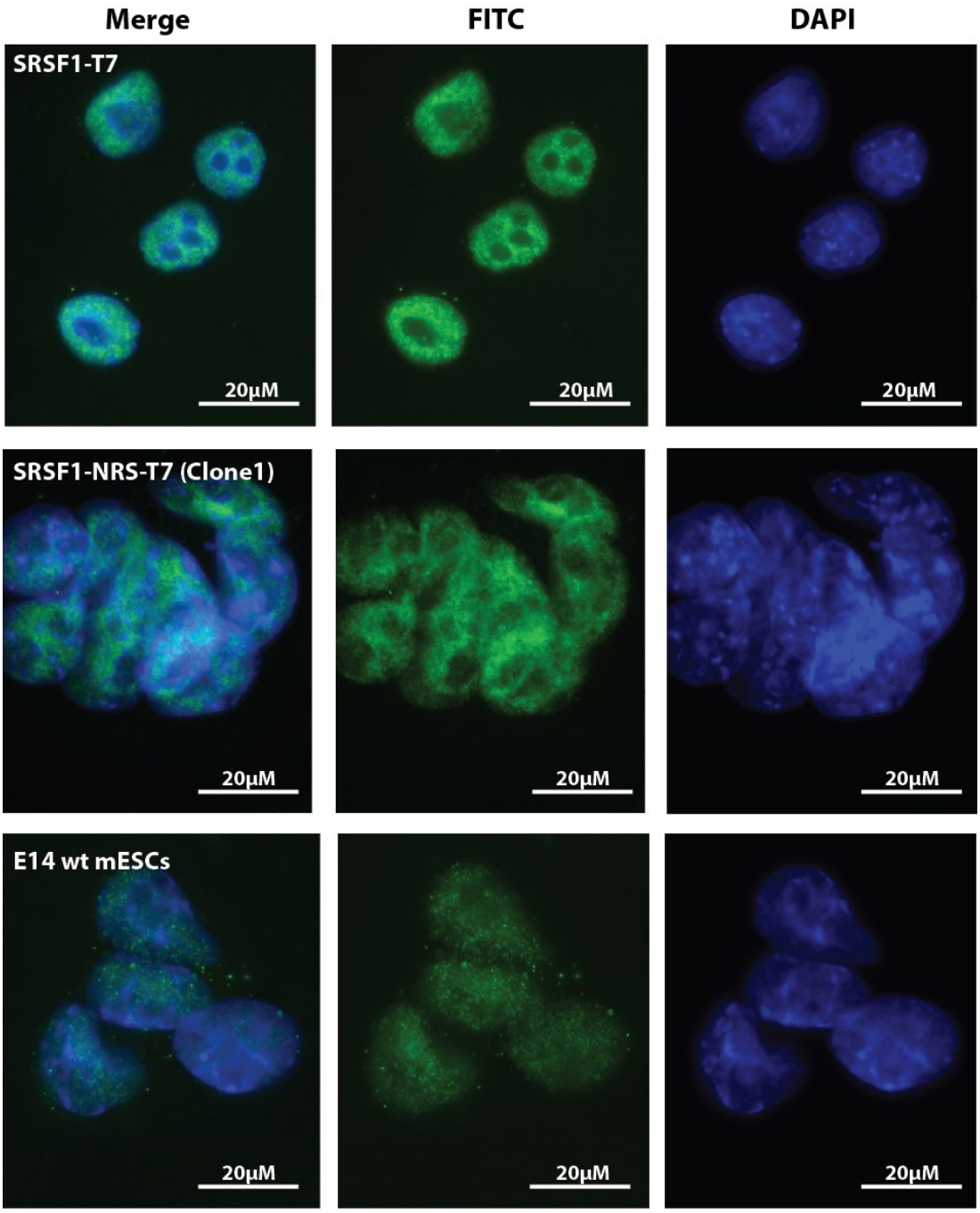


Figure 3.9: Immunofluorescence of endogenously tagged mESCs using the T7 antibody. SRSF1-T7 and SRSF1-NRS-T7 (clone 1) cells express endogenous SRSF1 that is nuclear at a steady state as detected by the T7 antibody (1:5000). Wt mESCs are given as a control for background detection of the T7 antibody. Nuclei are stained with DAPI. Scale bar represents 20uM.

In mESCs, tagged SRSF1 appears nuclear for both the shuttling and the non-shuttling SRSF1 (Figure 3.9). SRSF1-NRS-T7 is strictly nuclear and mimics the nuclear staining of the wild type shuttling counterpart, indicating that subnuclear localisation in Clone 1 cells is not

abberrantly affected by the NRS fusion, similar to that previously described (Cazalla *et al.*, 2002). Given that wild type SRSF1 function is predominantly in the nucleus, and shuttling is a transient dynamic property it is unsurprising that SRSF1-T7 protein is not detected in the cytoplasm using this assay. Furthermore, previous immunoflourescence data for endogenous SRSF1 have documented similar findings (Cáceres *et al.*, 1998). Treatment with inhibitors of transcription, such as actinomycin D, causes shuttling SR proteins including SRSF1 to accumulate in the cytoplasm (Cáceres *et al.*, 1998). This approach could have been undertaken here, however it does not confirm shuttling propensity. To truly determine if the tagged proteins are capable of nucleocytoplasmic shuttling, it would be necessary to perform interspecies heterokaryon assays. This is the gold standard assay for the detection of shuttling proteins (Piñol-Roma & Dreyfuss, 1992; Gama-Carvalho & Carmo-Fonseca, 2006).

For heterokaryon assays, tagged mESCs were fused with HeLa cells in the presence of cyclohexamide to inhibit mRNA translation, and immunoflourescence performed against the T7 tag. In this assay, if T7-tagged SRSF1 protein from in the knockin cells is capable of nucleocytoplasmic shuttling, it should be detected in the HeLa cells after cell fusion. This demonstrates shuttling as the HeLa cells do not express this protein so it must have migrated from the mouse cells if detected by the T7-antibody. In contrast if the protein cannot shuttle, such as in the SRSF1-NRS-T7 fusion protein, no signal should be observed from the T7-antibody staining in the fused HeLa cells.

The attempts at heterokaryon assays performed were inconclusive, as no shuttling of SRSF1-T7 was observed in the cells tested (data not shown). It is possible that this was due to inefficient cell fusion or suboptimal cell density ratio between the cell types. Moreover, due to the colony-type growth of mESCs, there are additional problems for imaging as these cells appear three dimensional whereas the HeLa are flat on the slide. Therefore, even if shuttling has occurred between species, it is unlikely to be observed using the methods here. In the future, the use of a high resolution confocal microscope would help to delineate this and scan through the focal plane of the mESCs to the HeLa cells.

3.3.9 Cell cycle analysis of knockin clones

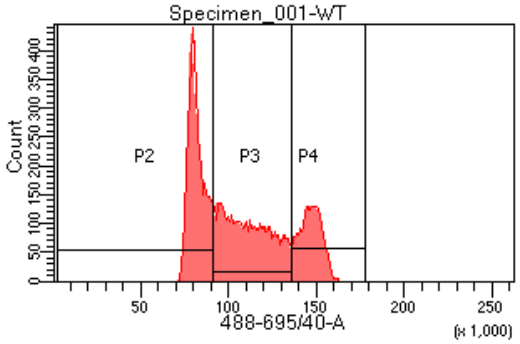
SRSF1 presence in the cytoplasm is required for the translational regulation of target mRNAs involved in cell cycle related processes such as the core kinetochore complex protein NDC80 (Maslon *et al.*, 2014). Consistent with this, siRNA mediated knock-down of SRSF1 results in mitotic arrest and development of a multipolar spindle phenotype, which can only

be rescued by expression of shuttling SRSF1 and not the nuclear retained counterpart (Maslon *et al.*, 2014).

The SRSF1 Clone 1 mESCs generated express a single copy of the SRSF1-NRS-T7 fusion protein. Although these cells exhibit a growth pattern analogous to the SRSF1-T7 cells with no gross changes, it is possible that there are subtle mitotic changes or changes in DNA content in the population. Given the previous data from human cells, and the fact that ESCs display a markedly different cell cycle and translational regulation thereof, this is an interesting possibility in relation to SRSF1 (Maslon *et al.*, 2014). To determine if this is the case, SRSF1-T7 and SRSF1-NRS-T7 cells were subject to treatment with propidium iodide prior to analysis by FACS to analyse the cell cycle and DNA content (Dominguez *et al.*, 2016).

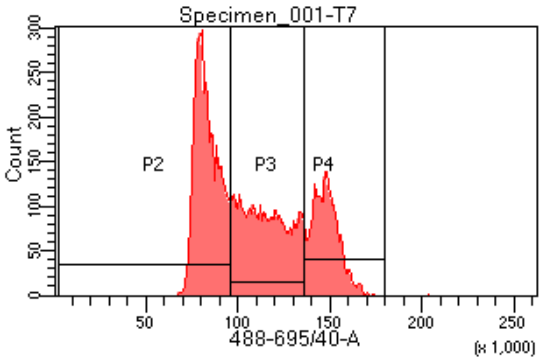
Consistent with normal growth patterns observed for the knockin cell lines, there is little difference in DNA content between clonal lines in comparison to untagged mESCs. Furthermore, there are similar proportions of the population engaged at each stage of the cell cycle, namely G1, S and G2/M phases. This suggests that the T7 tag does not perturb SRSF1 roles in cell division and these lines have normal DNA content, unaltered by CRISPR targeting. It is likely that for the SRSF1-NRS-T7 cells, expression of a single copy of the wild type, shuttling SRSF1 protein is sufficient to prevent cells from displaying aberrant mitotic phenotypes observed on SRSF1 knockdown. Indeed, expression of a single copy of SRSF1 in conditional knockout MEFs is sufficient to rescue cell death (Lin *et al.*, 2005). In the future, it would be interesting to knockdown the residual copy of shuttling SRSF1 in the Clone 1 cells and repeat PI staining and FACs analysis. Design of an siRNA specific for allele could permit this (Schwarz *et al.*, 2006).

E14 WT



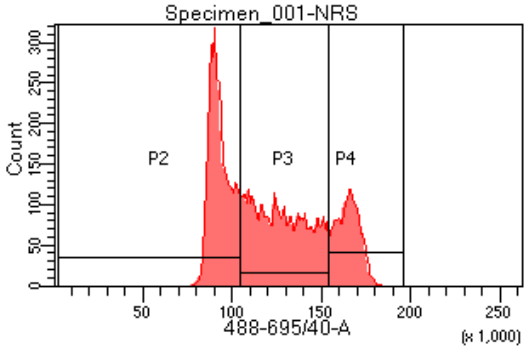
Population	#Events	%Parent	%Total
All Events	19,544	####	100.0
P1	10,522	53.8	53.8
P2	4,025	38.3	20.6
P3	4,293	40.8	22.0
P4	2,186	20.8	11.2

SRSF1-T7



Population	#Events	%Parent	%Total
All Events	21,464	####	100.0
P1	10,000	46.6	46.6
P2	4,199	42.0	19.6
P3	3,549	35.5	16.5
P4	2,234	22.3	10.4

SRSF1-NRS-T7 (Clone 1)



Population	#Events	%Parent	%Total
All Events	12,990	####	100.0
P1	10,000	77.0	77.0
P2	3,825	38.2	29.4
P3	4,268	42.7	32.9
P4	1,901	19.0	14.6

Figure 3.10: FACs analysis of SRSF1-T7 and SRSF1-NRS-T7 clone 1 mESCs. Cells were treated with propidium iodide prior to analysis. There are no gross changes in cell cycle observed, between SRSF1-T7 and SRSF1-NRS-T7 (clone 1 cells). Wt E14 cells are shown for comparison. Cell count is given on the Y-axis of each graph. P2, P3 and P4 represent cells in G1, S and G2 phase respectively.

3.4 Conclusions and discussion

SRSF1 is a highly dynamic shuttling SR protein that has numerous roles in both the nucleus and the cytoplasm. The nuclear functions of SRSF1, particularly in splicing, have been extremely well characterised; though in comparison little is understood regarding cytoplasmic function. SRSF1 is involved in the translational regulation of a specific subset of mRNA transcripts via an mTOR-dependent mechanism (Michlewski *et al.*, 2008; Maslon *et al.*, 2014).

Here, work aimed to create an endogenous model system in mouse embryonic stem cells to investigate the cytoplasmic roles of SRSF1, particularly those that have implications for mRNA translation. CRISPR/Cas9 was successfully designed to target the nuclear retention sequence from SRSF2 and a T7 epitope tag to the C-terminus of the SRSF1 at the endogenous genomic locus. In this system, the SRSF1-NRS-T7 fusion protein is restricted to the nucleus and cannot function in the cytoplasm. The SRSF1-T7 reciprocal cell line was designed as a control for targeting and to create a tool for downstream analysis.

Both homozygous and heterozygous SRSF1-T7 mESC lines were generated with relative ease once the optimal targeting protocol was in place (Figure 3.5). These cells exhibit a nuclear staining pattern by immunofluorescence analogous to wt SRSF1 however whether the tagged SRSF1 in these cells is capable of shuttling remains to be determined as heterokaryon assays were inconclusive due to technical problems.

In stark contrast, the generation of intact SRSF1-NRS-T7 homozygous cells was not achieved due to high rates of negative selection, as demonstrated by a decline in targeting frequency over time in the population (Figure 3.7). There are several potential reasons why this may be the case; either technical or biological.

Technically, the efficiency of generating a successful knockin may be extremely low using the targeting strategy implemented. Although the SRSF1-T7 knockins were generated with ease using the analogous strategy, in this context the inserted sequence is significantly smaller. Considering that the HDR template was delivered to cells as a ssODN of less than 300bp, by nature the homology arms are proportionally larger than in the SRSF1-NRS-T7 targeting design. This alone could provide inherent targeting advantages for generating SRSF1-T7 as there is a longer matched genomic sequence that is recognised for HDR to occur. In addition, when the targeting strategy was designed, CRISPR-targeting was at the pioneer stage as a gene editing tool. Therefore, the design of the strategy is similar to the first CRISPR targeting methods used in mammalian cells (Ran *et al*. 2013). Since then, numerous studies

have systematically demonstrated ways to manipulate this protocol to enhance the success of required targeting. For example, using cell synchronisation or other methods to increase HDR frequency may have helped in improving the efficiency of generating a successful knockin (Lin *et al.*, 2014). Changing the guide RNA used or the method for delivery of the CRISPR reagent may also have helped (Graham & Root, 2015). Also numerous additional CRISPR/Cas systems have been repurposed for mutagenesis, which have different consensus sequences for cleavage and thus may have been more suitable for targeting the SRSF1 locus as they do not require the -NGG PAM consensus, but have alternative cleavage requirements (Lewis & Ke, 2017; Mitsunobu *et al.*, 2017).

Mouse embryonic stems cells were used in the targeting approach described which exhibit a markedly different cell cycle profile than differentiated cells. On average, the cell cycle of undifferentiated cells is comparatively fast to differentiated cells, with mESCs undergoing mitosis every 11-16 hours, mainly due to an extremely short G1 phase of just 1.5 hours (Burdon *et al.*, 2002; Orford & Scadden, 2008). mESCs are grown in an artificial scenario due to the presence of Leukaemia inhibitory factor (LIF) in growth media. It is possible that this contributes to the negative selection observed in the generation of homozygous SRSF-NRST7 knockin mESCs.

LIF is a signalling molecule that acts to suppress differentiation through activation of the transcription factor Stat3 and other complex signalling pathways including ERK and PI3K (Martello & Smith, 2014). This facilitates growth of embryonic stem cells in culture in the absence of feeder cells. Continuous stimulation of these pathways results in a scenario that is not wholly representative of those *in vivo*. It has been shown that LIF has an effect on mTOR activity although there are conflicting reports from mouse and human systems (Easley *et al.*, 2010; Cherepkova *et al.*, 2016) . Considering that SRSF1 acts in translation via this pathway and that maintenance of stemness requires mTOR signalling could explain why these cells could struggle to tolerate a lack of SRSF1. Perhaps maintenance in LIF pushes the system too far in terms of lack of translational regulation of a particular SRSF1 target mRNA meaning that only very few cells in a population could tolerate lack of cytoplasmic SRSF1. Furthermore, the clones seen to be repaired to the homozygous SRSF1-NRS-T7 genotype were not subject to downstream analysis. It is possible that these cells contain auxiliary genomic aberrations that bypass such strict requirements for SRSF1 cytoplasmic function. Indeed, in the described experiment, cells had undergone two rounds of CRISPR-targeting which inherently increases the likelihood of off-target mutagenesis so auxiliary indels are highly imaginable.

Biologically, there are numerous reasons why the generation of homozygous SRSF1-NRS-T7 knockins could be subject to intense negative selection. It is possible that SRSF1 is required for the translational regulation of pluripotency transcription factors. Other mRNA binding proteins and splicing factors such as PTBP1 have been shown to be involved in the alternative splicing to translation-control axis. For example, through the control of transcription factor YY2. Furthermore, 4E-BP proteins are required for maintenance of a pluripotent state and removal of their exerted regulation causes differentiation (Tahmasebi *et al.*, 2016). Considering that SRSF1 is known to act in translation in a mechanism involving these proteins it is possible that SRSF1 may play a role in this, albeit of specific factors. Perhaps if cells lose such regulation on generation of homozygous SRF1NRS-T7 that cannot act in translation, they cannot be maintained by LIF in a pluripotent state. The Nat1 protein is able to control mESC differentiation through translational regulation (Sugiyama *et al.*, 2016) so it is feasible that an analogous control mechanism might in play for shuttling SRSF1. In light of this, it would be interesting to investigate the differentiation potential of any repaired Clone 1 cells in future work.

Chapter 4

Investigating the mechanisms of nuclear retention of a non-shuttling SR protein

4.1 Introduction

The nuclear roles of SR proteins in constitutive and alternative splicing are well understood. In this capacity, they have been shown to interact with both core and auxiliary splicing factors in a dynamic manner, as was described for the prototype SR protein SRSF1 (Akerman *et al*. 2015). Many such interactions of SRSF1 are not reliant on RNA thus direct protein interactions of SRSF1 have a crucial role in its function. SRSF1 also interacts with components of other RNA processing complexes such as the exon junction complex (EJC) within which it physically interacts with the core helicase EIF4AIII (Saulière *et al.*, 2012).

However, a subset of SR proteins is capable of nucleocytoplasmic shuttling; also exemplified by SRSF1. In the cytoplasm, SRSF1 can act in translation to enhance the translation of a subset of mRNA targets in an mTOR-dependent mechanism. SRSF1 can interact directly with the mTOR kinase as well as polysomes and PP2A (Michlewski *et al.*, 2008). However, in comparison to the nuclear roles of SRSF1, there is still much to be understood regarding the role of SRSF1 in the cytoplasm, particularly regarding its protein interactors.

Much of the existing functional SRSF1 interactome data are reliant on overexpression systems. SR proteins, including SRSF1 are capable of autoregulation through unproductive splicing, suggesting that the tuning of their expression level is vital for cellular homeostasis (Lareau *et al.*, 2007; Sun *et al.*, 2010). When SRSF1 is overexpressed, it is possible that some of the identified interactions are artificial, therefore overexpression datasets may not accurately represent biologically relevant targets. Furthermore, there is a lack of data for SRSF1 function in varying tissue-types or undifferentiated cell systems.

4.1.2 Mechanisms of nuclear retention of the SRSF2 NRS

SRSF2 is a non-shuttling protein that contains a potent nuclear retention sequence (NRS) at the extreme C-terminus of its RS domain. Fusion of this sequence to shuttling SR proteins, such as SRSF1, causes their constitutive retention in the nucleus. This sequence has been extensively biochemically characterised by Cazalla *et al* (2002), however, the direct interacting partners of this sequence and how it exerts its retention capacity are unknown. It is possible that this sequence enables SRSF2 to specifically bind to certain interaction partners, which ultimately prevent nucleocytoplasmic shuttling.

4.2 Aims

To identify the interactome of SRSF1 in the nucleus and the cytoplasm and uncover the mechanism of nuclear retention for the NRS from SRSF2, mass spectrometry was carried out using various approaches. Firstly, three epitope tagged constructs, T7-SRSF1, T7-SRSF1-NRS and T7-SRSF2 were overexpressed in HeLa cells. In this context, SRSF1-NRS and SRSF2 are strictly nuclear, and only SRSF1 should be present in the cytoplasm. To determine the relevant binding partners of these, immunoprecipitation was performed against the T7 tag on each construct and binding partners subject to mass spectrometry. The resulting datasets were subsequently overlapped to infer both the compartment-specific SRSF1 interactome and factors which uniquely bind the NRS sequence. To further delineate specific binding partners of the latter, the NRS peptide sequence alone was also used for IP-MS analysis. Finally, to identify the SRSF1 binding partners in mouse embryonic stem cells, IP-MS was performed on cells in which SRSF1 contained an endogenous T7 tag.

In summary, the main aims of the following work are:

(1) To characterise the nuclear and cytoplasmic binding partners of SRSF1
(2) To determine the interactome of the NRS from SRSF2
(3) To characterise the binding partners of endogenous SRSF1 in mESCs

4.3 Results

4.3.1 IP-MS to investigate the binding partners of SRSF1, SRSF1-NRS and SRSF2

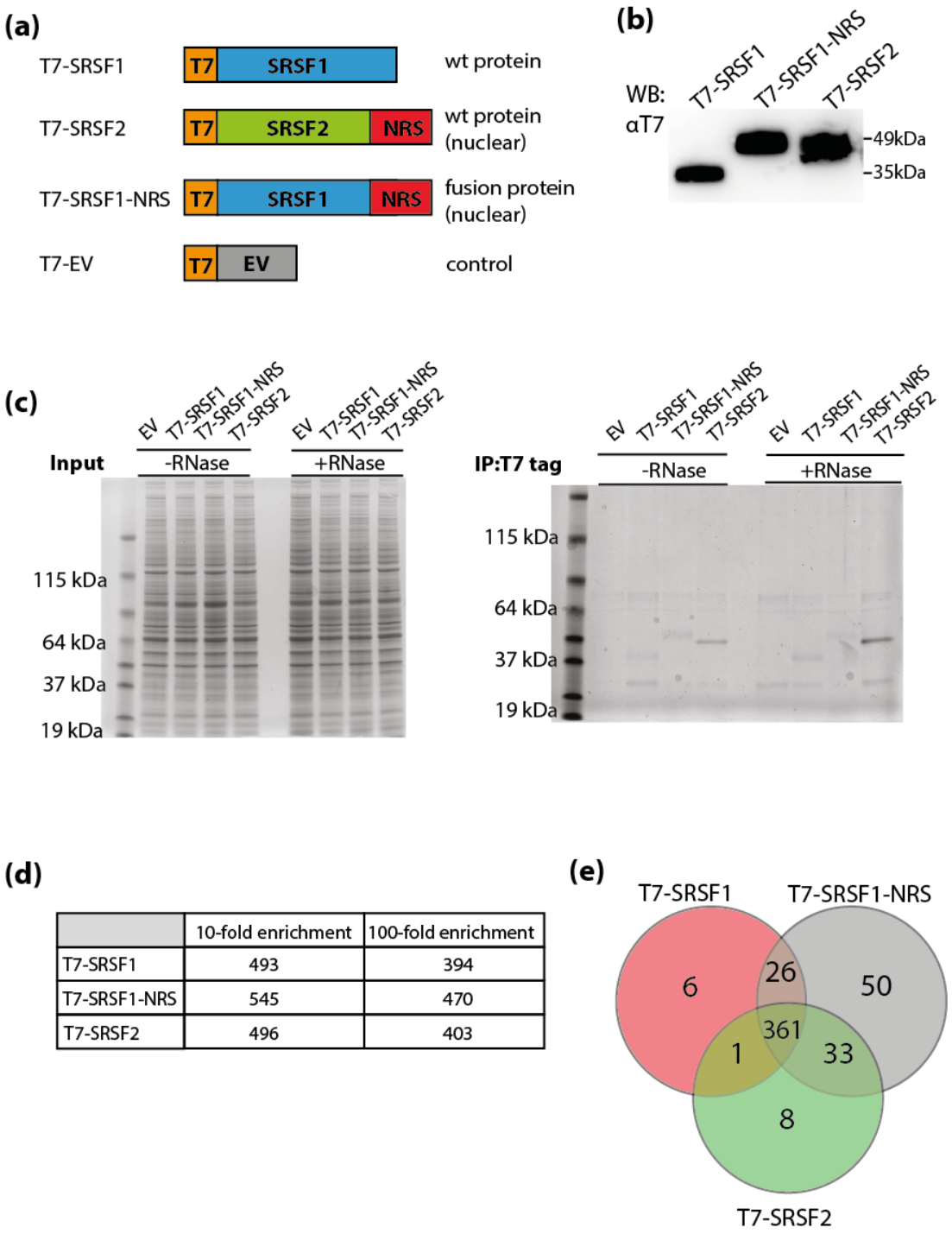
4.3.1.1 Workflow and data processing

In the first IP-MS approach, T7-tagged constructs were overexpressed in HeLa cells and immunoprecipitations carried out against the T7 tag using T7 antibody coupled beads (Figure 4.1). RNase was not used in these experiments as little difference in total protein obtained was observed by Coomassie staining on RNase treatment (Figure 4.1(c)). Therefore, it is possible that any protein interactions observed in downstream analysis are mediated by RNA. Beads were washed stringently before elution and samples subjected to mass spectrometry to determine the bound proteins for each bait. Each IP was performed in biological triplicate and run on the mass spectrometer in technical duplicate. The empty vector (EV) control was used to determine non-specific binding partners of the T7 tag, which aids downstream analysis when filtering the data. Whole cell lysates were used in this experiment

to facilitate interaction analysis in both cellular compartments. In theory, only the SRSF1 dataset should be enriched for cytoplasmic binding partners as the SRSF1-NRS and SRSF2 should not shuttle into the cytoplasm. However, accuracy of compartment annotation should be treated with caution especially since, it is possible that numerous proteins exist in both the nucleus and the cytoplasm. All mass spectrometry experiments were carried out by the in house IGMM facility. Raw data counts were analysed using MaxQuant by Jimi Wills from the facility.

The total number of interactions observed for each T7-tagged construct is extremely large (>700 each, data not shown), with a significant amount of binding in the EV control, suggesting that the T7 IP contains a certain degree of non-specific interactions. To dissect the bona fide interaction partners of the bait proteins, harsh thresholds were applied to the dataset for downstream analysis. Firstly, a student's t-test was used to determine the statistical significance of each interaction. Using this approach, the data were highly reproducible between both biological and technical replicates and a p-value threshold of <0.01 was applied to the data. Interactions that did not meet these requirements were discounted from the downstream analysis as were any duplicated gene names in the list. Secondly, the fold enrichment over the EV control was calculated and the data filtered at both 10-fold and 100-fold enrichment. Fold enrichment in this calculation, refers to ratio over the EV control of the average LFQ (label-free quantification) intensity value for each interaction as determined using the Max Quant analysis software by the mass spectrometry facility. The algorithm used to determine the LFQ value by the Max Quant software is based on the extracted ion counts of identified peptides rather than MS/MS spectral counts as in other methods. The LFQ intensity value directly reflects the peptide intensities determined. This makes it ideal for use for quantification and sample comparison in studies such as these which do not rely on metabolic labelling (Cox *et al.*, 2014).

To put biological context to interactome datasets, particularly when large, Gene Ontology (GO) analysis can be used to group gene sets into classified functional groups. A GO term annotation can contain any number of genes and a range of analogous functions. To determine the GO terms associated with the significantly enriched interacting partners for each bait, the PANTHER database (Gene Ontology Consortium) was used as it contains current GO annotations. Each GO analysis was performed using default parameters for the biological process option, with human gene set used as background, as in this case as HeLa cells were used.



	10-fold enrichment	100-fold enrichment
T7-SRSF1	493	394
T7-SRSF1-NRS	545	470
T7-SRSF2	496	403

Figure 4.1 The Interactome of shuttling and non-shuttling SR proteins using IP-MS. (a) Schematic of T7-tagged overexpression plasmid constructs used as bait for the T7 IP. (b) Overexpression of constructs depicted in (a) in HeLa cells. Note the blot for the EV construct is not shown as any protein produced is not detectable in this assay (c) RNase optimisation of IP protocol. Samples sent for MS analysis without RNAse treatment. Coomassie stained gel shown is representative of replicas used (d) Total number of hits identified as significantly enriched (>100 fold increase), over the EV control for each IP. p-value <0.01. (e) Overlap of interactions between SRSF1, SRSF2 and SRSF1-NRS datasets. (>100 fold enrichment, p value<0.01).

4.3.1.2 Overexpressed SR proteins have large but highly similar interactomes

The significantly interacting partners for each T7-IP either 10-fold or 100-fold enriched over the EV control with a p-value of less than 0.01 are shown in Figure 4.1). There are several hundred significantly interacting partners at each fold cut-off, with approximately 100 of these being lost between the 10-fold and 100-fold thresholds. The SRSF1-NRS fusion protein has the largest number of binding factors, and, naturally, displays a high degree of overlap with the T7-SRSF1 and T7-SRSF2 datasets. However, it also has several more exclusive interacting partners, 50 at 100-fold cut-off, than the other two tagged proteins (Table 4.1).

There are 361 common interactors of T7-SRSF1, T7-SRSF1-NRS and T7-SRSF2 (Figure 4.1 (e)). Exclusive interactions at the decided statistical thresholds are minimal, suggesting that the core interactomes of these three proteins are highly similar. In general, the datasets are reflective of the known functions of SR proteins in the nucleus. Indeed, many of the shared interacting proteins corroborate this as there is a large enrichment for pre-mRNA splicing and RNA processing proteins in the data. For example, other SR proteins (SRSF-3, -4, -5, -6, -7 and -9), hnRNP proteins and alternative splicing factors such as TRA2A and TRA2B. Core components of the spliceosome are also highly enriched in all three datasets such as U2AF35/65, the U4/U6 proteins Prp3 and Prp4 and the U1 snRNP factors SNRPA and SNRPC. Interestingly, numerous members of the Exosome complex are also found in all three datasets. The highest enriched in this category is EXOSC10 which is the human homologue of *S.cerevisiae* Rrp6, and a 3'-5' exonuclease with several nuclear roles such as during rRNA and snoRNA processing. Furthermore, there are several other significantly enriched factors that support a function for SRSF1 and SRSF2 in rRNA processing and ribosome biogenesis including numerous DExH/DEAD-box helicases known to be involved in such processes.

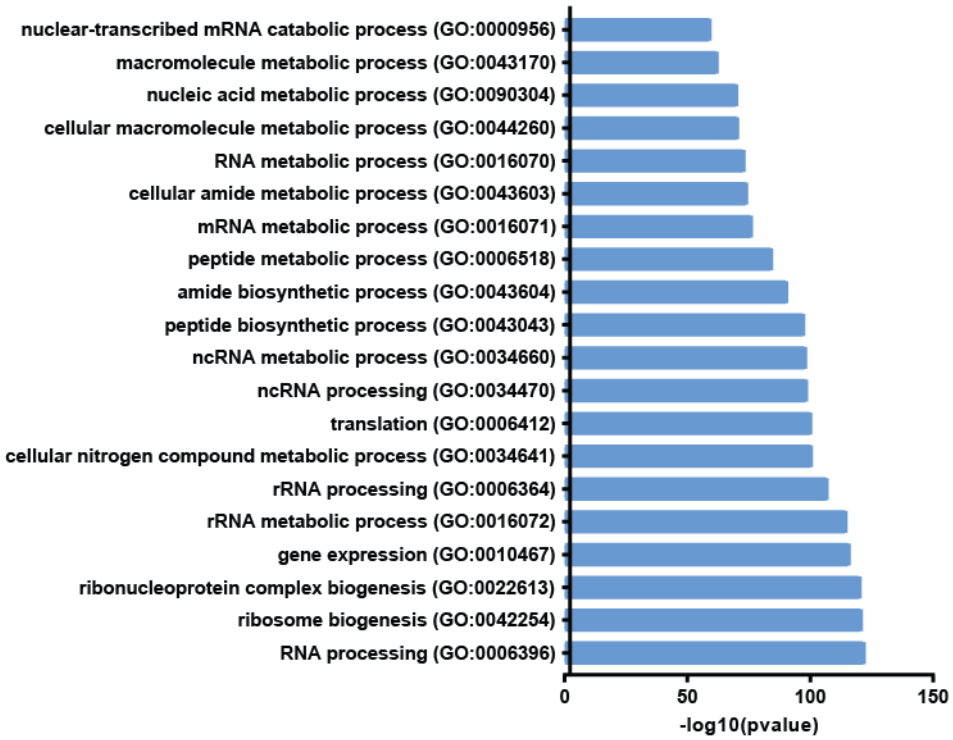
Group	No. of proteins	Protein Name
T7-SRSF1	6	APC2, ARHGEF2, COPB2, RFC5, TLN2, YWHAE
T7-SRSF1-NRS	50	ACTL6A, AP2A1, AP2B1;AP1B1, AP2M1, C8orf33, C9orf114, CCAR1, DDX28, DHX15, EIF2S2, EIF2S3; EIF2S3L, FXR2, G3BP2, GRWD1, GTF3C3, HBA1, HIST1H2BN; HIST1H2BM; HIST1H2BH; HIST2H2BF; HIST1H2BC; HIST1H2BD; HIST1H2BK; H2BFS, HIST1H4A, HIST2H3A; HIST3H3; HIST1H3A; H3F3B; H3F3A, HNRNPA0, HNRNPAB,

		HNRNPD, HP1BP3, LUC7L2; C7orf55-LUC7L2, LUC7L3, MAP4, MMTAG2, MRPL19, MRPL38, MRPL53, NOL9, PEG10, PGAM5, PHF5A, POLR2H, PPP1CC, RPL12, RPL23, RPL9, RPS10; RPS10-NUDT3, RPS11, RPS17L;RPS17, RPS25, RPS5, SRSF11, STAU1, SUPT16H, TSR1, U2AF1;U2AF1L4, YTHDC2
T7-SRSF2	8	C1QBP, DDX23, HIST2H2BE; HIST1H2BB; HIST1H2BO; HIST1H2BJ, MAK16, MRPL21, MRPS18C, SRSF2, THUMPD1
T7-SRSF1 and T7-SRSF1-NRS	26	ASF1B; ASF1A, C14orf166, C18orf21, CENPV, DDX47, EIF6, ELMSAN1, EXOSC5, FXR1, MRPL54, NAP1L1, NEMF, NOM1, POLR1D, POM121; POM121C, RBBP6, REPIN1, RFC4, RPLP0;RPLP0P6, RRP7BP;RRP7A, SRP72, SRSF1, UBA52, XRN2, ZC3H15, ZCCHC8
T7-SRSF1 and T7-SRSF2	1	MNT
T7-SRSF1-NRS and T7-SRSF2	33	AURKAIP1, CHTOP, EMG1, ERH, FOXK1, HNRNPC; HNRNPCL4; HNRNPCL1; HNRNPCL3; HNRNPCL2, HNRNPUL2; HNRNPUL2-BSCL2, LARP7, MRPL10, MRPL16, MRPL32, MRPL37, MRPL45, MRPL46, MRPL48, MRPL50, MRPL51, MRPL52, MRPS30, NHP2, PINX1, POLR3B, POP4, PRPF4B, RBM19, RPL22, RPP38 SAFB, SNRNP70, SRSF10, XRCC5, XRCC6, YTHDC1

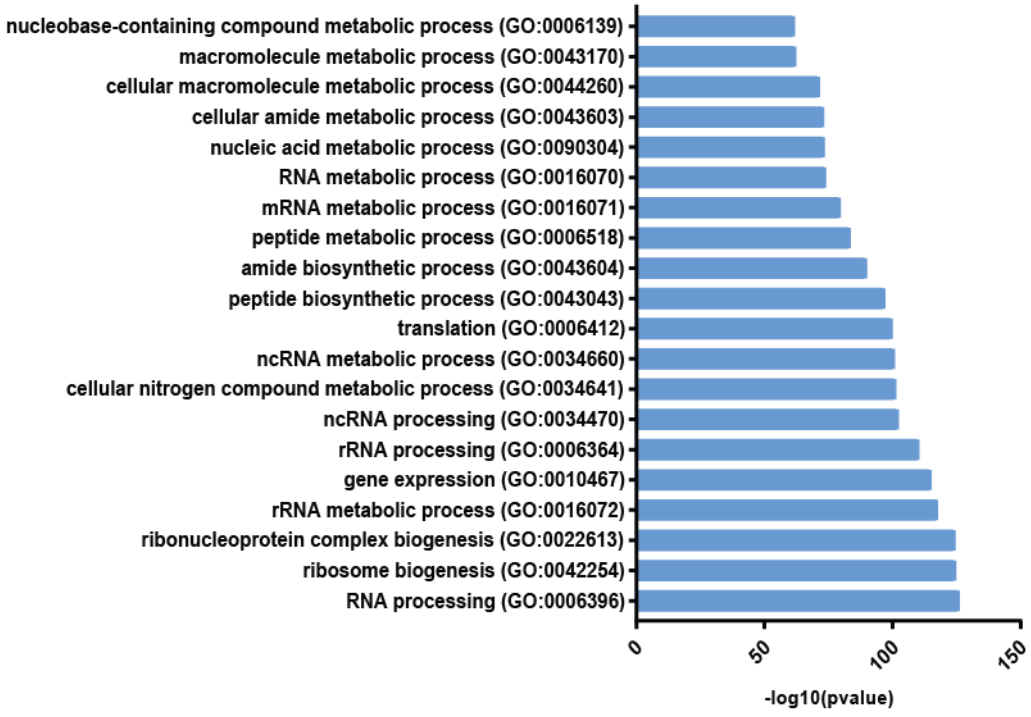
Table 4.1: Summary of exclusive interactions between SR proteins tested. Interacting proteins significantly enriched over the EV control that are unique within the sample(s) identified by the Group column (p-value <0.01, >100 fold enriched). This is considering the ratio of each dataset separately with respect to the EV control. The Group column indicates the exclusive dataset examined that are not included in the 361 common interactors, no. of proteins corresponds to Figure 4.1(e).

GO term analyses reflects a similar trend of interactome overlap. Between SRSF1, SRSF1-NRS and SRSF2 (Figure 4.2(a)) hits are significantly enriched for functional groups such as RNA processing, gene expression and RNA metabolism. This is consistent with their role as RNA processing factors and strongly implies that the data represent biologically

(a) Common interactors



(b) T7-SRSF1



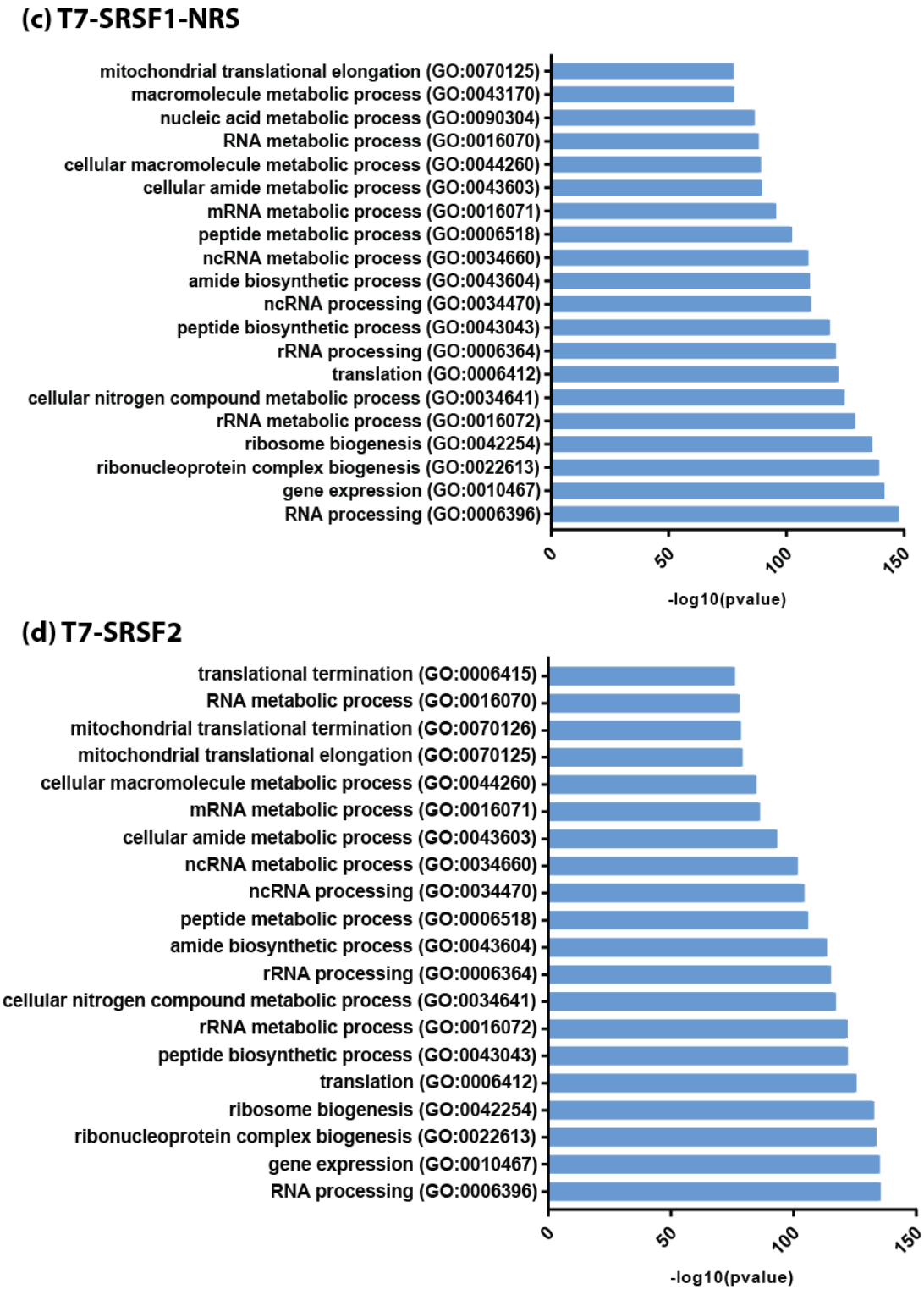


Figure 4.2 Gene Ontology (GO) term analysis for SR protein interactions. Top 20 significantly enriched GO terms for (a) the 361 common interactors of all bait proteins, (b) T7-SRSF1, (c) T7-SRSF1-NRS, (d) T7-SRSF2, all 100 fold enriched over EV control where pvalue <0.05. Significantly enriched GO term annotation is shown according to -log10(pvalue)

relevant hits. Individually, the top 20 GO term groups are largely similar when grouped by significance of enrichment. However, the GO annotation groups that are the most enriched in the data are not necessarily those with the lowest p value. Discounting p value, the highest scoring GO terms are more specific and include categories such as snoRNA processing and nuclear polyadenylation. This is consistent between SRSF1, SRSF1-NRS and SRSF2.

4.3.1.3 Cytoplasmic binding partners of SRSF1

These data can be used to identify putative interactors of SRSF1 in the cytoplasm by comparing hits for SRSF1, which shuttles to the cytoplasm and SRSF1-NRS which does not. In simplistic terms, factors that do not bind SRSF1-NRS (or SRSF2 which is also nuclear), could represent cytoplasmic interactors of SRSF1. There are 6 proteins that significantly bind SRSF1 but not SRSF1-NRS or SRSF2 Table 4.1. None of these can be found in the top 100 binding partners for SRSF1 and none has any previously annotated significant interactions with SRSF1 (STRING DB analysis, data not shown), so their functional relevance is unlikely. However, SRSF1 is predominantly a nuclear protein that shuttles to the cytoplasm transiently, therefore it may be impossible to delineate cytoplasmic binding partners using the approach described here which could further enrich for SRSF1 interactions in the nucleus.

4.3.1.4 Factors for nuclear retention

There are 33 proteins that bind to both SRSF1-NRS and SRSF2 with 100-fold enrichment over the EV control that do not bind SRSF1 with this level of significance (Figure 4.1, Table 4.1). They can be filtered out on either fold enrichment, significance or both i.e. both significance and fold enrichment must be criteria for inclusion in a section. It is possible that these are specific for the NRS sequence and thus factors involved in the nuclear retention of SRSF2.

The STRING database tool permits visualisation of functional interaction networks as inferred from numerous database annotations for a particular gene. These include protein-protein interaction data, both experimentally validated and predicted, as well as functional systems data used to infer biological pathways in which the group of proteins may be involved. The latter is advantageous as it considers indirect or previously unannotated interactions that may group similarly interacting factors. This is particularly useful here as the NRS sequence does not have an annotated interactome or functional network outside of its context within

SRSF2. Furthermore, the SRSF2 NRS peptide sequence does not share homology with any known functional domains that may hamper interactome prediction.

The functional STRING interaction network for the proteins shared between SRSF2 and SRSF1-NRS is shown in (Figure 4.3). Note that for this analysis, SRSF2 was added manually to the search list. The interactors within this group do occur more frequently than by chance over the background of the whole genome. The nodes form several clusters within the network, although only one is associated directly with SRSF2. It is possible that the other groups represent unannotated interaction networks in which SRSF2 or SRSF1 is present.

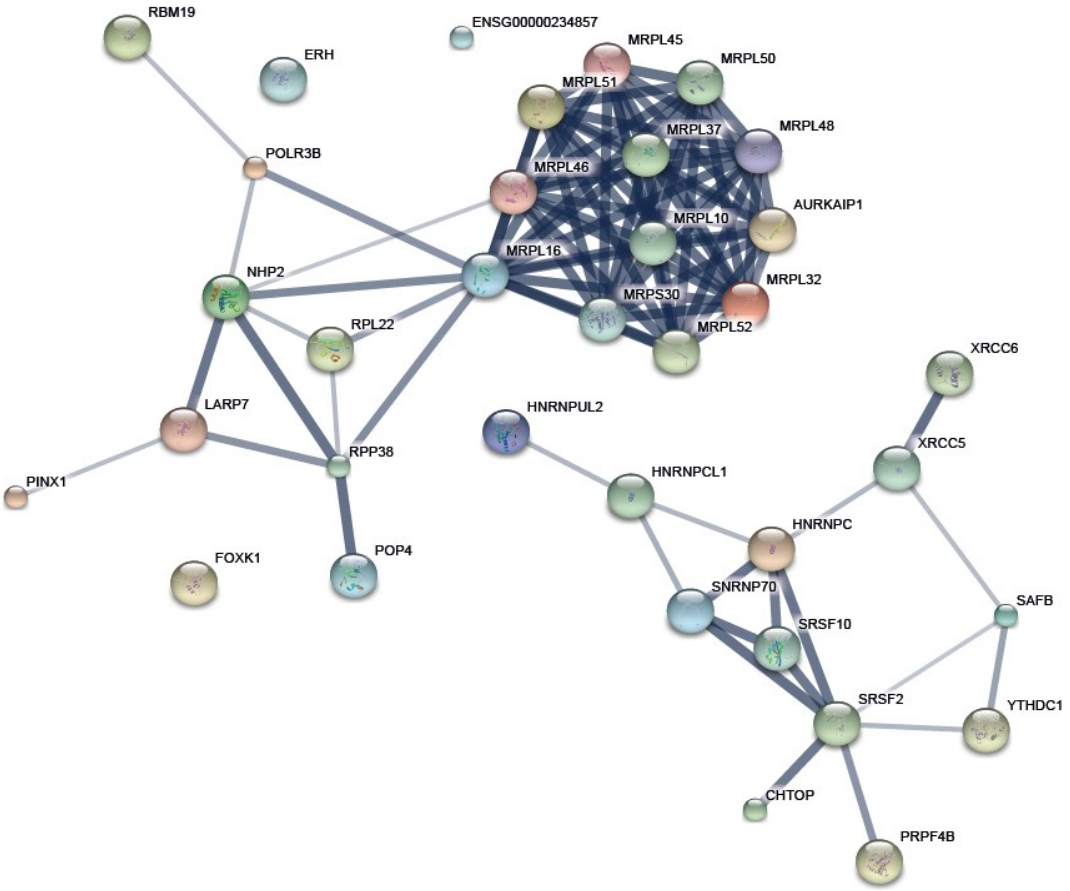


Figure 4.3 STRING network analysis of nuclear interactions. These were defined as those 33 proteins shared between T7-SRSF1-NRS and T7-SRSF2 (Table 4.1). Lines between nodes infer the confidence of the interaction based on the thickness. Image shown is output of STRING tool, found at: https://string-db.org/cgi/

The group of hits that closely associate with SRSF2 based on previous annotation were chosen for further investigation. From these hits, one protein of interest is SAFB, a

scaffolding matrix protein that binds to S/MAR DNA situated around the nuclear envelope. One hypothesis could be that SAFB interacts with the NRS of SRSF2 in close proximity to the nuclear pore to hold it in the nucleus. Despite multiple attempts, it was not possible to validate any interactions of SAFB observed in MS by IP-WB (data not shown). However, it is likely that this was due to technical problems resolving the SAFB complex from its tightly associated DNA matrix. It is possible that the stringent lysis procedures required to resolve SAFB disrupt SR protein binding, especially if it is transient.

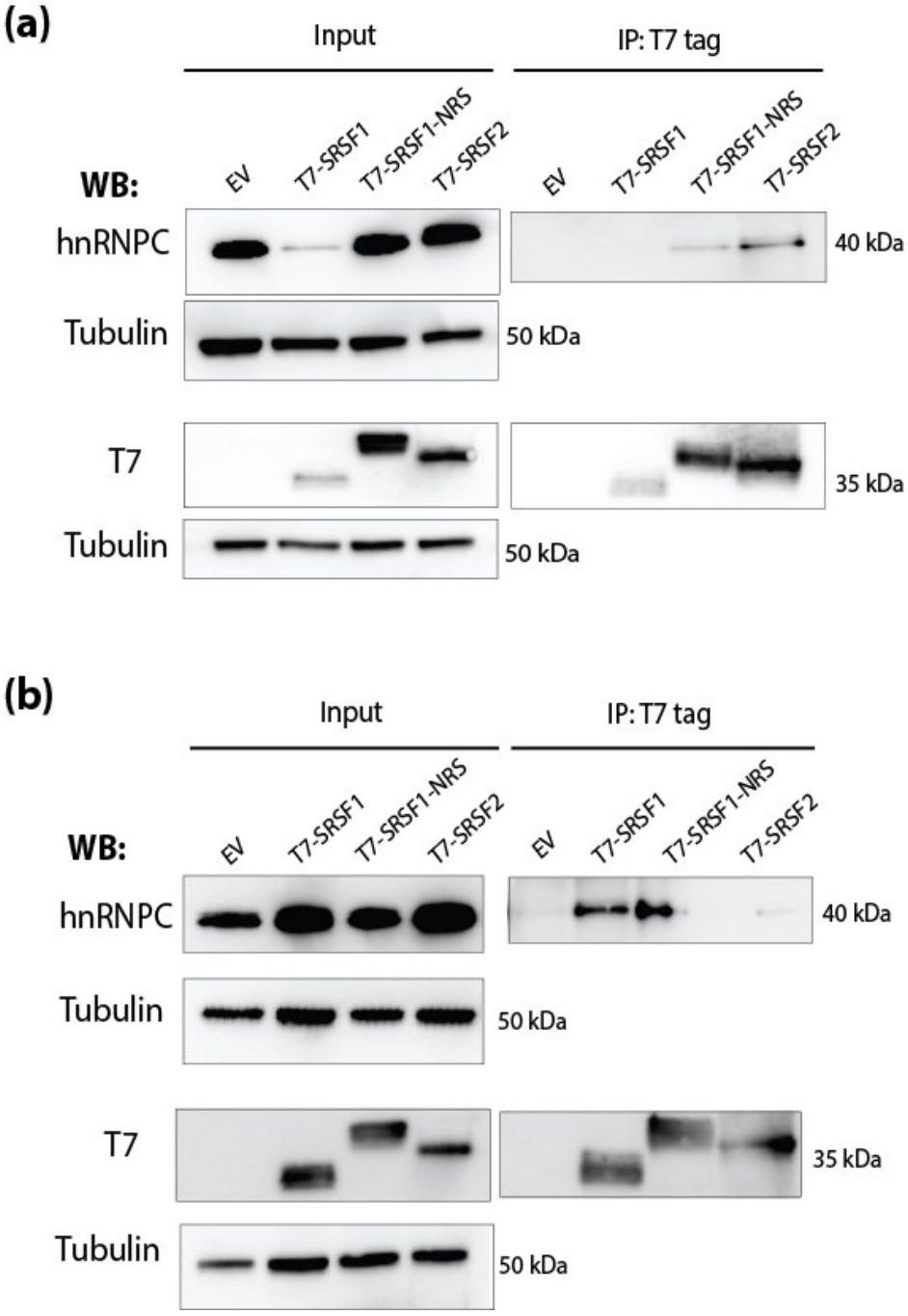


Figure 4.4 Validation of SR protein interactors. Validation of hnRNPC interactions observed by IP-MS for SRSF2 and SRSF1-NRS in HeLa cells. Two loading controls are shown as T7 western blot was performed on a second gel to avoid secondary antibody cross-reactivity. Biological replicates (a) and (b).

A high confidence interaction with HNRNPC can be seen in the STRING network analysis. It was possible to validate the interaction of HNRNPC with T7-SRSF2 and T7-SRSF1-NRS by IP-WB (Figure 4.4). However, it is unclear from these data whether T7-SRSF1 is also capable of binding as there are inconsistencies between replicates making it unlikely that any interaction with SRSF2 is specific for nuclear retention.

CHTOP is another interesting interaction partner in this group as it is known to interact with PRMT complex which has been shown to be involved in the arginine methylation of SRSF2 to control its nuclear localisation (Larsen *et al.*, 2016). This may be another interesting target to explore in the future.

4.3.2 MS using Peptides to investigate the NRS interactome

SRSF2 does not exhibit nucleocytoplasmic shuttling but resides constitutively in the nucleus, due to the unique and potent nuclear retention sequence (NRS) at the very C-terminus of the RS domain (Cazalla *et al.*, 2002). This domain is dominant and when fused to SRSF1 in the SRSF1-NRS fusion protein, it becomes trapped in the nucleus. In general, SR proteins are highly intrinsically disordered due to the RS domains. However, SRSF2 is predicted to have the highest propensity for disorder at the C-terminus, possibly due to the NRS domain (Haynes & Iakoucheva, 2006). The NRS is also highly phosphatase resistant and this property will be transferred to SRSF1 in context of the SRSF1-NRS fusion protein (Lin *et al.*, 2005). Considering that SR protein dephosphorylation is required for nuclear export after splicing, this could be partly why SRSF2 evades nuclear escape. However, the exact mechanism for nuclear retention by this sequence is unknown. Interestingly the NRS shares virtually no sequence homology with any known proteins or domains suggesting that a unique mechanism or binding of an auxiliary factor could be responsible. To investigate this possibility and determine the direct interacting partners of the SRSF2 NRS sequence, peptide sequences were designed as shown (Figure 4.5).

Each peptide was biotinylated to enable streptavidin pull-down prior to MS analysis. They also contain a small linker (SGSG) to enable flexibility of the peptide once coupled to streptavidin beads. Three additional peptides were also designed for quality control and to

facilitate comparison with the NRS and discount non-specific interactions in downstream analysis. Peptide 2 is the equivalent region from SRSF8, a vertebrate-specific SR protein and a pseudogene of SRSF2 (Soret *et al.*, 1998). This sequence shares much homology with that from SRSF2, though it is unknown whether this sequence acts as an NRS in the same manner. Peptide three is the SRSF2 NRS in which each Serine and Arginine residue has been changed to an Alanine residue. Alanine is resistant to phosphorylation so interacting partners of this peptide and the true NRS are presumably not dependent on phosphorylation. Finally, peptide four is a scrambled NRS sequence to facilitate removal of any non-specific interaction partners detected.

(a)

Peptide Name	Peptide Sequence	Description
SRSF2-NRS	SGSGPPPVSKRESKSRSRSKSPPKSPEEEGAVSS	Endogenous SRSF2 NRS
SRSF8-NRS	SGSGPPRVSKRKSKSRSRSKRPPKSPEEEGQMSS	Endogenous SRSF8 NRS
SRSF2-NRS Ala	SGSGPPPVAKAEAKAAAAAKAPPKSPEEEGAVAA	R/S to Alanine control
Scramble	SGSGSSRREEAPKSSEPGSKKPRSESPPVVSKPS	NRS scramble peptide control

(b)

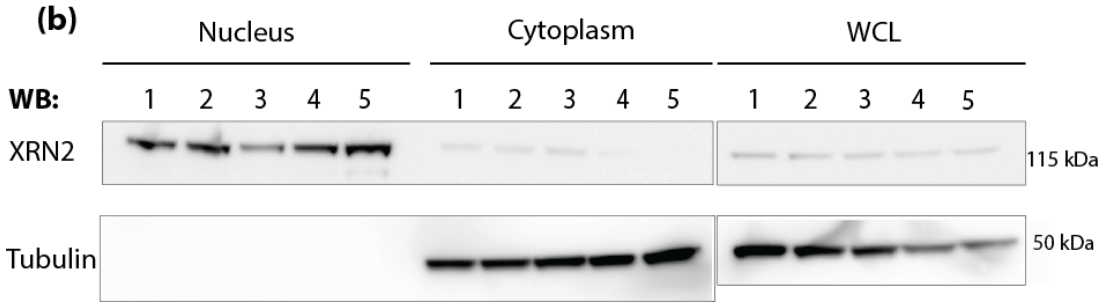


Figure 4.5 Peptide IP-MS in HeLa cells: Summary of peptide MS in HeLa cells. (a) Table to outline the peptide design implemented and the rationale for each. (b) Western blotting of fractionated cell extract to ensure clean subcellular compartments. Anti-XRN2 exclusively marks nuclear fractions; anti-tubulin is the cytoplasm -exclusive. In the whole cell extract, both proteins are detected.

In this experiment, both nuclear and whole cell lysates (WCL) were used to allow comparison between cellular compartments. Interactions of the NRS sequence with cytoplasmic proteins, inferred as those made in the whole cell extract and not the nuclear fraction, should be treated with caution considering the NRS is not present in the cytoplasm *in vivo*. However, numerous proteins persist in all cellular domains in different abundancies, thus it is conceivable, that SRSF2 could bind an abundant protein present in the whole cell lysate that is only in the nucleus in small quantities but nonetheless still binds SRSF2. In this scenario, it is possible that the approach taken may overlook low abundance protein interactions.

Peptides were coupled to streptavidin beads via their biotin linker and immobilized on a column before the cell or nuclear extracts was washed through the column. After washing, the beads were eluted and the peptide bound proteins subject to mass spectrometry analysis. Data were analysed in an analogous manner to the HeLa cell MS experiments. The IP and downstream MS was performed in triplicate for each peptide. The average LFQ value of each binding peptide was calculated for each peptide IP. A ratio of this was then taken between the sample peptide-IP and the scramble control. A student's t-test was calculated for each ratio to determine the reliability of the values. A vast number of proteins appeared in the raw data for which the LFQ values were all 0. This indicates that they were not present in any of the peptide-IP samples and thus were discounted from downstream analysis. To obtain a significant list of interaction partners, cut-offs were set at anything with a p-value under 0.05 and with a ratio of over 20 (meaning over 20-fold enriched over the scramble peptide control). Using this approach, the values shown in Table 4.2 were obtained for each peptide-IP. For the nuclear fractions, the top hit is Involucrin (IVL), a component of human skin. However, IVL is also found in the whole cell lysate, but does not bind the SRSF2 NRS peptide. This is unlikely to represent a biologically relevant significant interaction and indicates that this experiment contains a high degree of contaminants.

For the whole cell lysate data, there are no significant hits that are unique to the NRS peptide, although there are few – 8 in total - that are also significant for SRSF8. Five out of eight of these proteins are nuclear and have a role in transcription and related processes. For example, DNA ligase 3, which is involved in the nuclear DNA repair pathway that recruits XCCR6. Although this protein is not found in the peptide data, related factors such as XRCC5/6 are found to be specific for the SRSF1-NRS and SRSF2 overexpression datasets. This suggests that SRSF2 may participate in these processes through interactions of the NRS in the RS domain. None of the proteins identified in either the nuclear or WCL datasets have any annotation for function in nuclear retention, however they do show an overall enrichment for nuclear proteins. Many common contaminants and substantial binding of the scramble peptide suggest that this data is not reliable to assess a mechanism for SRSF2 nuclear retention.

Group	Gene Name	Gene symbol	Compartment
NRS only	Involucrin	IVL	Cytoplasmic
NRS and SRSF8	MAPK regulated corepressor interacting protein 2	FAM195A	Nuclear and Cytoplasmic

	Fibroblast growth factor receptor 1	FLG	Nuclear and Cytoplasmic
	Prostaglandin E synthase 2	GBF1	Nuclear
	Mitogen-activated protein kinase kinase kinase kinase 4	MAP4K4	Cytoplasmic
	E3 ubiquitin-protein ligase RBX1	RBX1	Nuclear and Cytoplasmic
	Histone deacetylase complex subunit SAP30	SAP30	Nuclear
	Paired amphipathic helix protein Sin3b	SIN3B	Nuclear
NRS, SRSF8 and NRS-Ala	Mitogen-activated protein kinase kinase kinase kinase 4	MAP4K4	Cytoplasmic
	E3 ubiquitin-protein ligase RBX1	RBX1	Nuclear and Cytoplasmic
	Paired amphipathic helix protein Sin3b	SIN3B	Nuclear

Table 4.2: Enriched interactors for IP-MS using peptides in nuclear extracts. Summary of significant hits for each peptide experiment, as shown by the Group column. Each dataset was obtained in triplicate. There are no significantly enriched hits for the NRS alone. >20-fold enrichment over Scramble shown, p value >0.05

Comparison with the HeLa IP-MS overexpression data for SRSF2 show very little overlap with hits from the peptide IP-MS, with CCDC137 being the only protein significantly present in both datasets. It is likely that this protein does represent a bone fide target of SRSF2, however it is unlikely to be specific for the NRS sequence as it also binds SRSF1. Considering that this factor does not bind to the Alanine mutant NRS, it is possible that interactions are mediated through generic binding of RS dipeptides in SR proteins. Furthermore, this interaction could depend on phosphorylation of these residues.

Group	Gene Name	Gene symbol	Compartment
NRS and SRSF8	ATP-binding cassette sub-family D member 1	ABCD1	Cytoplasm
	Coiled-coil domain-containing protein 137	CCDC137	Nucleus, Chromosome

	ER membrane protein complex subunit 1	EMC1	ER Membrane
	Histone H2A type 2-B	HIST2H2AB	Nucleus, Chromosome
	DNA ligase 3	LIG3	Mitochondria, Nucleus
	Serine/threonine-protein kinase 10	STK10	Cytoplasm
	Transcription elongation regulator 1	TCERG1	Nucleus
	WD repeat-containing protein 82	WDR82	Nucleus
NRS, SRSF8 and NRS-Ala	Histone H2A type 2-B	HIST2H2AB	Nucleus, Chromosome

Table 4.3: Enriched interactors for IP-MS using peptides in whole cell extracts. Summary of significant hits for each peptide experiment, as shown by the Group column. Each dataset was obtained in triplicate. There are no significantly enriched hits for the NRS alone. >20-fold enrichment over Scramble shown, p value >0.05.

Considering the lack of overlap or significantly bound hits it is likely that this dataset is not representative of the biologically relevant interactions of the NRS peptide. However, there were significant problems with the synthesis and subsequent solubility of the NRS peptide, perhaps as it is so highly disordered. Even after numerous attempts with different solvents, the NRS peptide failed to solubilise, despite the other three in the experiment easily dissolving in the required solvent. Therefore, when used for this experiment it was only partially in solution which could have had an adverse effect on the data. In addition, all of the sequences tested are short and disordered thus by nature have a tendency for highly promiscuous binding profiles. To properly address NRS-binding proteins, the rest of the proteins domains may be required to stabilise any NRS-specific interactions. In addition, interactions that trap SRSF2 in the nucleus may be dependent on phosphorylation, which was not addressed here. The use of a phosphomimetic NRS peptide would perhaps delineate this.

4.3.3 IP-MS in mESCs

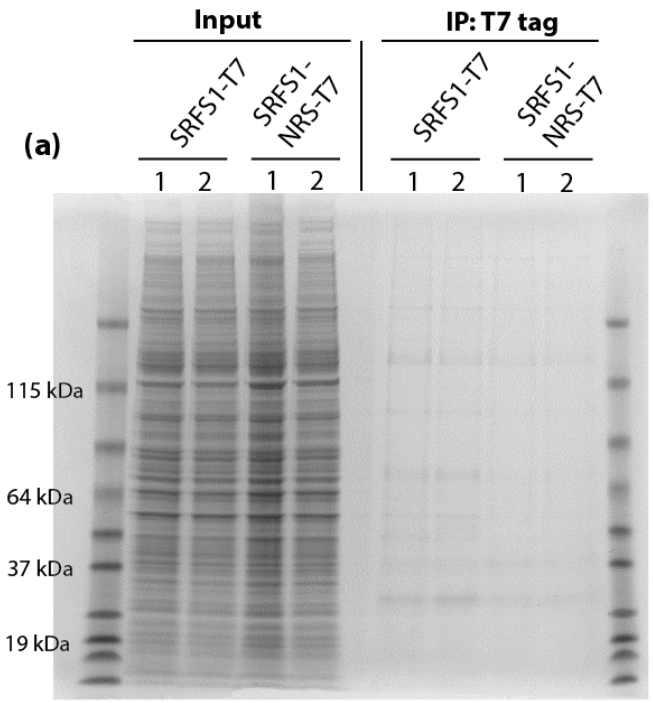
To determine the binding partners of SRSF1 in mouse embryonic stem cells (mESCs), the CRISPR generated SRSF1-T7 and SRSF1-NRS-T7 (clone 1) cell lines were used for IP-

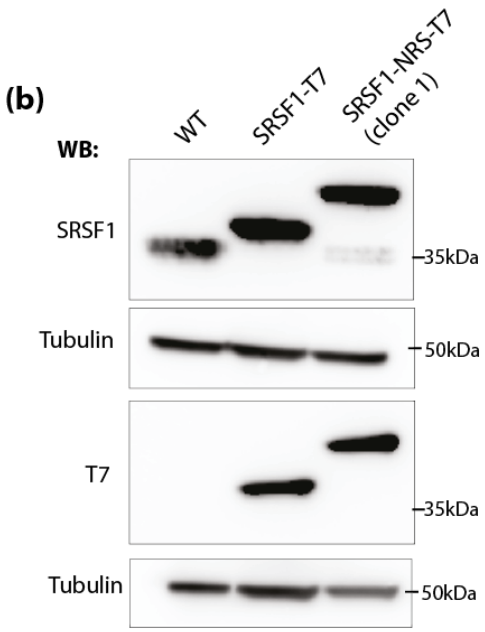
MS as described previously against the T7-tag. SRSF1-T7 cells are homozygous for a T7 tag at the endogenous SRSF1 locus. The SRSF1-NRS-T7 clone 1 cells, contain the intact NRS-T7 inset on one allele and an NRS-T7 insertion on the other allele that also has two single base pair deletions which result in the formation of a truncated protein only slightly larger than wild type SRSF1. The protein produced from this allele does not contain a T7 tag and therefore any binding partners of this protein should not be specifically pulled down by the T7 antibody. It should be noted that the T7 tag is C-terminal as opposed to the N-terminal used for the HeLa cell approach.

In this system, there is no overexpression of SRSF1 and the protein is expressed at endogenous levels in both cell lines (Figure 4.6). This is advantageous over the HeLa cell dataset which was reliant on the overexpression of each bait protein. By nature, an endogenous IP-MS should portray a more reliant picture of the binding landscape of the bait protein as overexpression may result in promiscuous non-specific binding events that do not occur or are irrelevant *in vivo*. Conversely, there is the disadvantage that less prolific or transient interactions may be missed in an endogenous IP. These data are particularly important considering that Mass Spectrometry data for SRSF1 in mouse embryonic stem cells does not exist in the literature.

The IP against the endogenous T7 tag was performed in biological quadruplicate and WT (untagged) mouse ES cells were used as the background control. Proteins that significantly bind in WT cell extract at similar levels to the tagged cells can be discounted from the downstream analysis as they most likely result from non-specific binding to the T7 antibody.

Unsurprisingly, far fewer interacting factors in total were identified in these data in comparison to the HeLa cell dataset, presumably due to endogenous expression of the bait protein. To determine interacting factors that are significantly enriched in the tagged versus the untagged cells, the average LFQ intensity value was calculated for each replicate and used to calculate a ratio over the WT cells. The student's t-test was also performed to determine the accuracy between replicates for each sample and a p-value threshold of <0.05 was set for downstream analysis.





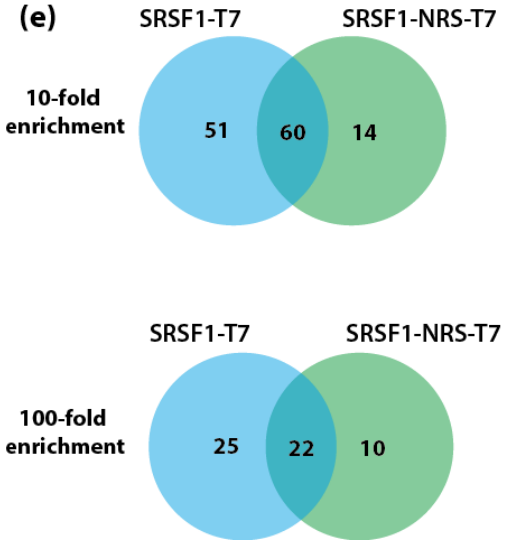
(c)

	10-fold enrichment	100-fold enrichment
SRSF1-T7	111	41
SRSF1-NRS-T7	74	32

(d) **Top 20 hits:**

SRSF1-T7	NIFK, SRSF10, CLASRP, PRPF8, REXO4, ABCF1, RRP8, ISG20L2, PWP1, BUD13, H2AFJ;HIST1H2AA;HIST1H2AH;HIST1H2AK; HIST1H2AF; HIST3H2A;HIST1H2AB, SF3A1, ABT1, SGOL2, UTP18, SAP30BP, CFAP20, CCDC137, MAK16, UTP23
SRSF1-NRS-T7	SRSF10, NIFK, SGOL2, BUD13, RRP8, MKI67 SAP30BP, ISG20L2, ABT1, ZCCHC8, PWP1, UTP18, AEBP2, DDX31, REXO4, CDK13, HP1BP3, OXA1L, MAK16, EZH1

(e)



(f) **100-fold enrichment:**

SRSF1-T7	ABCF1, ACIN1, AW146154, BCLAF1, BLM, BRD7, C12ORF66, CFAP20, CLASRP, H1F0, H2AFJ;HIST1H2AA; HIST1H2AH;HIST1H2AK;HIST1H2AF;HIST3H2A; HIST1H2AB, NOL6, POLR2H, PRPF38B, RBM15, RNPS1, SF3A1, STAU1, UTP23, YTHDC1
SRSF1-NRS-T7	AATF, AEBP2, CDK13, EZH1,MRP19,MRPS33, NOC3L OXA1L, PBRM1, ZCCHC8
Shared	ABT1, BUD13, CACTIN, CCDC137, DDX31,DHX8, DNAJC9, HPBP1, ISG20L2, MAK16,MKI67,MRPL11, MRPS30, NIFK, PWP1, REXO4, RRRP8, SAP30BP, SHOL2, SRSF1, SRSF10, UTP18

Figure 4.6 Endogenous IP-MS in mESCs: Summary of the interactome of endogenously tagged SRSF1-NRS-T7 and SRSF1-T7 mouse ESCs. (a) Coomassie stained gel as an example of samples sent for IP-MS analysis. (b) Western blotting of mouse ESCs to demonstrate fusion protein expression. (c) Number of significant hits obtained for SRSF1-T7 or SRSF1-NRS-T7 (clone1) cells, fold enrichment indicates over wild type (untagged) cells. pvalue <0.05. (d) Identity of the 20 most enriched proteins over wt cells for either SRSF1-T7 or SRSF1-NRS-T7 (e) Overlap between the SRSF1-T7 and SRSF1-NRS-T7 datasets at 10-fold (top) or 100-fold (bottom) enrichment over wt cells. Table corresponds to factors in each overlap at 100-fold enrichment.

The resultant analysis at 100-fold enrichment over WT cells, demonstrates that there were 41 and 32 significantly interacting factors for SRSF1-T7 and SRSF1-NRS-T7 (clone 1), respectively (Figure 4.6 (c)). This is strikingly fewer than the equivalent observed in the HeLa cell overexpression data and exemplifies the potential benefits of using an endogenous system. It is also possible that this reflects differences in the function of SRSF1 between species or cell type. It would be interesting to delineate this further with comparison to endogenous data from human cells.

The top 20 significantly interacting partners for SRSF1-T7 and SRSF1-NRS-T7 are shown in Figure 4.6(d). Concurrent with SRSF1 function, there are several splicing factors in this list, including CLASRP (or SRSF16), whose exact function is unclear in the literature. It was identified in a Y2H screen as a CLK4-interacting protein which also contains an RS domain and is thought to be involved in alternative splicing regulation of Clk-related kinases (Katsu *et al.*, 2002). The most enriched proteins in this dataset for each bait are the shuttling SR protein SRSF10 and NIFK, which is nuclear. NIFK or MKI67 FHA domain-interacting nucleolar phosphoprotein, interacts with the cell cycle regulator MKI67 and localises to mitotic chromosomes (Booth *et al.*, 2014; Pan *et al.*, 2015). Interestingly, MKI67 is also one of the top 20 significantly enriched binding partners for both SRSF1-T7 and SRSF1-NRS-T7, suggesting interactions with this complex. Furthermore, MKI67 has been shown to bind PP1 (Booth *et al.*, 2014), the phosphatase responsible for SR protein dephosphorylation.

Within the top 20 hits, there is a striking enrichment of rRNA processing and ribosome biogenesis factors both for SRSF1-T7 and SRSF1-NRS-T7 such as MAK16, UTP18 and UTP23. Another interesting observation is the number of transcriptional regulators in this list. For example, for SRSF1-NRS-T7 the catalytic component of the PRC complex, Ezh1, was identified.

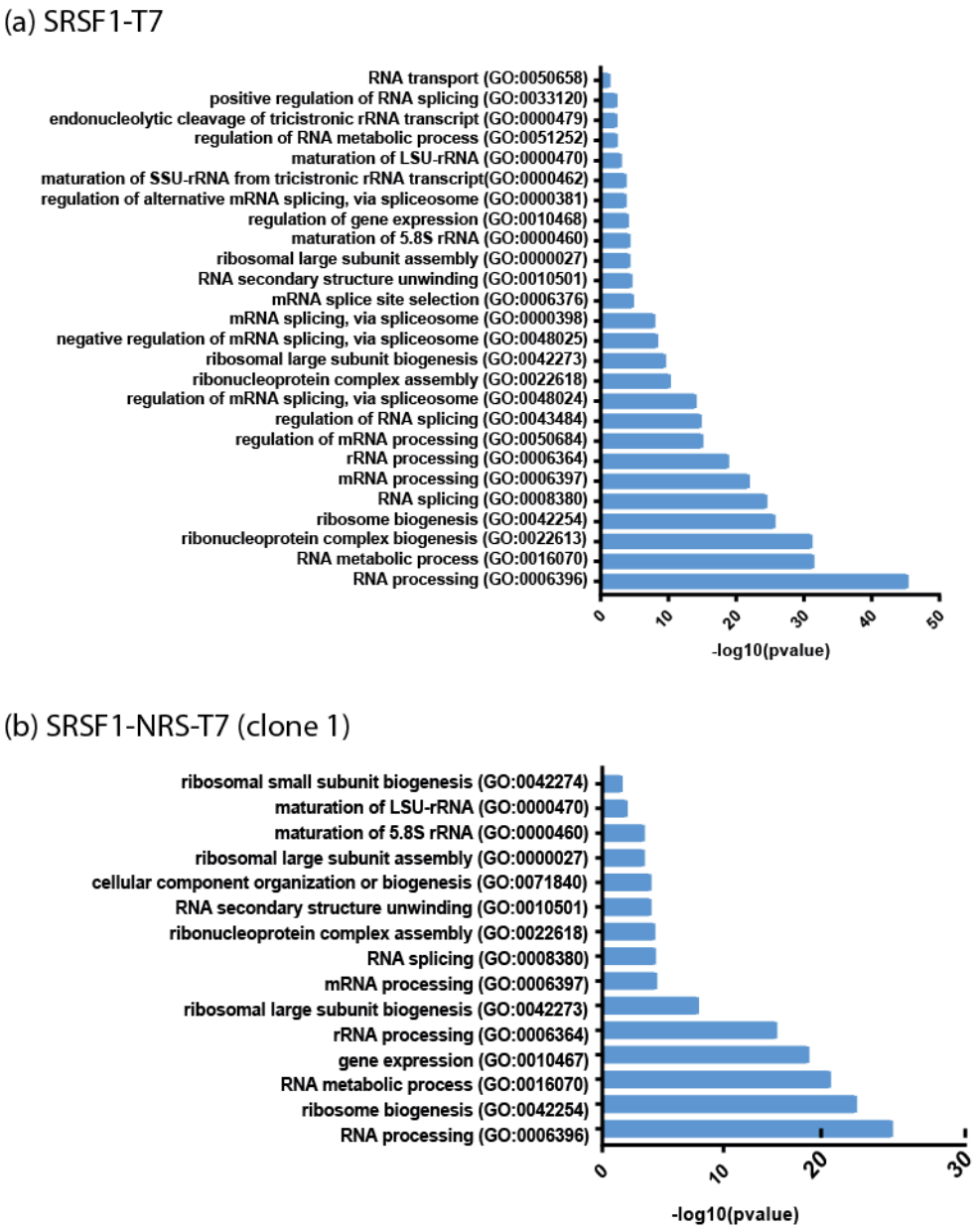


Figure 4.7 Gene Ontology (GO) analysis for mESC targets. Results represent the 10-fold enriched binding partners over wt mESCs for (a) SRSF1-T7 or (b) SRSF1-NRS-T7 (clone 1) mESCs.

Gene Ontology analysis for the SRS1-T7 and SRSF1-NRS-T7 interactomes were performed using panther as previously described, except the input list used was for 10-fold cut-off as opposed to 100-fold. Redundant GO terms were then removed using the GOtrim software (Jantzen *et al.*, 2011). The significantly overrepresented GO terms for each dataset can be seen in figure 4.7. Categories such as RNA processing, RNA splicing, splice site selection and spliceosome are present, consistent with known SRSF1 functions. It is

encouraging that this is the case for the knock-in systems as it demonstrates that the tagged and fusion proteins are both capable of normal splicing interactions. Interestingly, there are subtle differences in GO term enrichment. For example, SRSF1-T7 is enriched for RNA-transport associated factors, whereas SRSF1-NRS-T7 is not. This could illustrate cytoplasmic targets of SRSF1 or those that are lost by SRSF1 binding in the nucleus due to forced nuclear retention.

4.3.3.1 Cytoplasmic interactors of SRSF1-T7

The datasets for SRSF1-T7 and SRSF1-NRS-T7 were overlapped to infer on cytoplasmic interactors of SRSF1 (Figure4.6). Those that do not bind to the nuclear retained SRSF1-NRS-T7 could be important for SRSF1 function in the cytoplasm. STAU1 is the only protein that is significantly enriched for SRSF1-T7 but does not bind SRSF1-NRS-T7 or WT cells at all. Despite multiple attempts, this interaction could not be validated by IP:WB in any of the cell lines (Figure 4.8). It is possible that this is due to technical difficulties however, as the antibody used appears to bind numerous non-specific bands with high affinity (demonstrated by the Input samples). In this scenario, it is unclear whether lack of enrichment in the IP is for these reasons or the MS data represent a false positive interaction with SRSF1. In the future, this interaction could be further investigated following antibody optimisation to facilitate correct detection of Staufen 1 protein in the Input samples. STAU1 is a cytoplasmic dsRNA binding protein that has numerous cytoplasmic roles such as mRNA translation and quality control pathways (Vessey *et al.*, 2008). Somewhat conversely, it also has roles in mRNA stabilisation particularly during stem cell differentiation and neurogenesis (Kretz, 2013).

Overlap also demonstrated a significant enrichment in the SRSF1-T7, data for SF3A1. SRSF1-NRS-T7 also bound SF3A1 although not in all four replicates, meaning it is not significant using the chosen p-value cut-off. An interaction between SRSF1 and SF3A1 was also experimentally validated by Akerman *et al* (2012) in HeLa cells, suggesting that functional conservation of interactions between species does exist for SRSF1 at least for its roles in splicing.

The nuclear-cap binding protein 1 (NCBP1, NCBP80) was observed from the data to interact with SRSF1-T7 and not SRSF1-NRS-T7. However, this protein is absent from the dataset overlap as it is highly enriched in two of four replicas and thus failed the prescribed significance threshold. Despite this, it was possible to validate this interaction by IP:WB for not only SRSF1-T7 but also for SRSF1-NRS-T7. In the future, it would be prudent to further

validate this interaction for several reasons. Firstly, as in Figure 4.8, NCBP1 also appears to bind WT cells so it is possible that this represents a false positive interaction. Secondly, to determine the subcellular location of putative interactions, through for example fractionation.

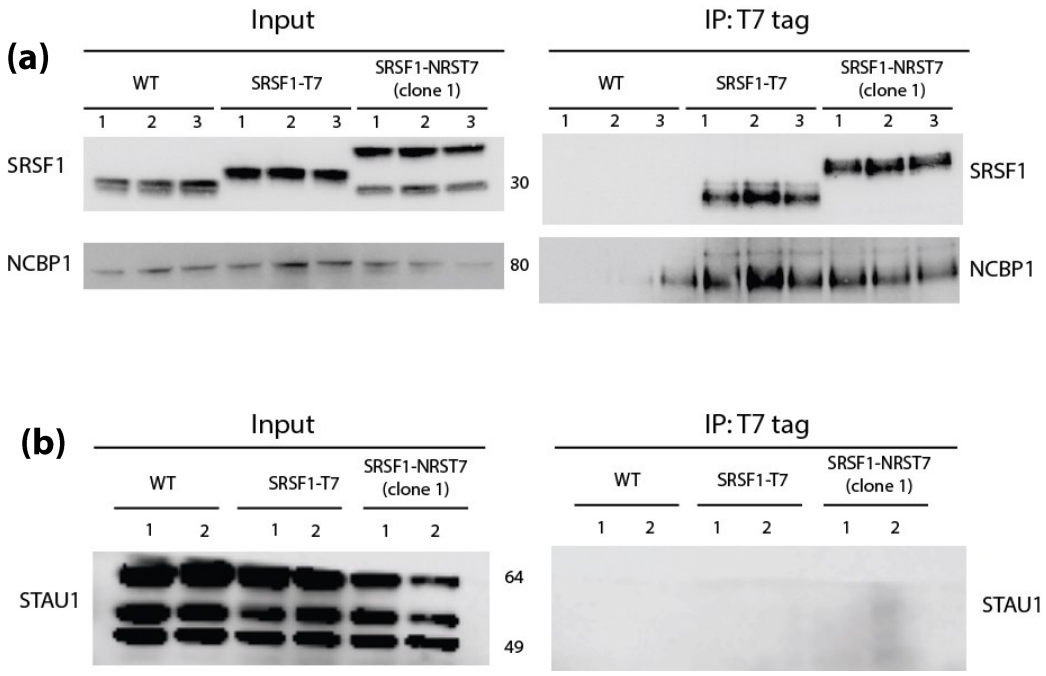


Figure 4.8: Validation of selected targets from mESCs obtained from endogenously tagged mouse ES cells. (a) IP-WB to validate interactions with NCBP1. (b) IP-WB to confirm interactions with Staufen1.

4.3.4 Overlap of data with previous dataset

A recent study used probabilistic PPI network prediction coupled with IP-MS to determine the dynamic interaction networks of SR proteins in the spliceosome (Akerman *et al.*, 2015). In this analysis, SRSF1 was used as the prototype SR protein in analogous T7-SRSF1 overexpression system to that used here. In this study SRSF1 did not require RNA to bind to spliceosomal proteins although some of the interactions found by IP-MS are nuclease sensitive. This could explain why such a little difference was observed on the gel with RNase treatment. Akerman *et al* (2015) found 204 proteins that are significantly enriched for SRSF1 in the IP-MS dataset and predicted a further 554 that were not experimentally validated by IP-MS. The data obtained here for T7-SRSF1 (100-fold enrichment) have approximately a 34% (133/395) overlap with the total (predicted and IP-MS) dataset from Akerman *et al*. Furthermore, over half (83/133) of these were also obtained in the authors IP-MS data.

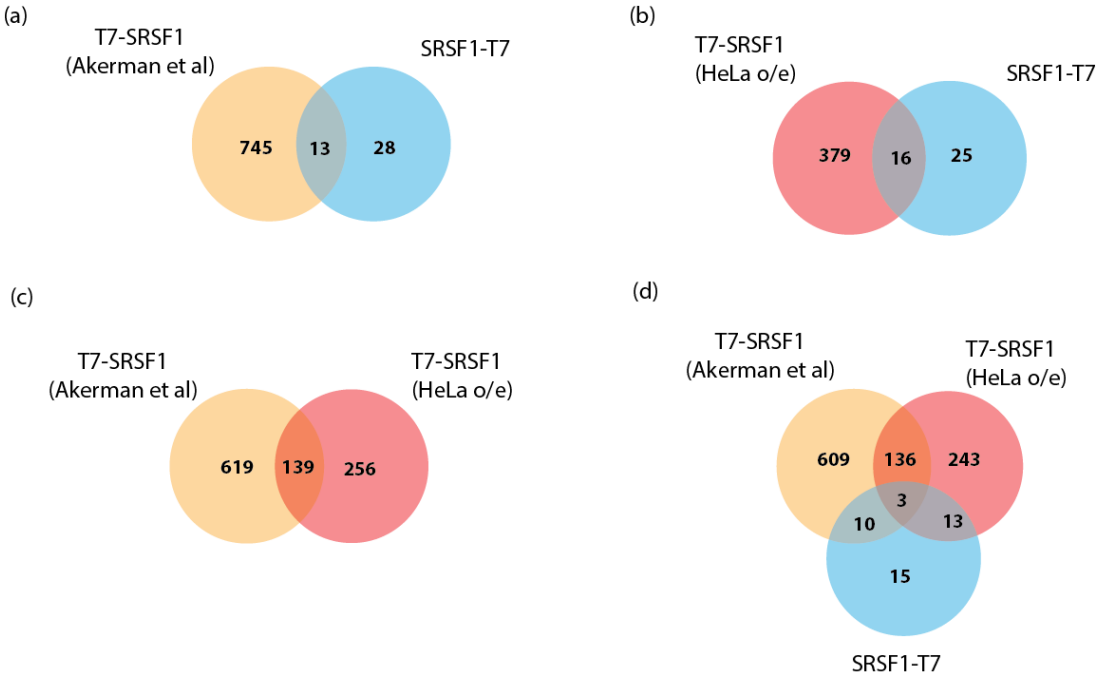


Figure 4.9: Overlap of IP-MS targets with known SRSF1 binding partners. Overlap of data from HeLa cells and mouse ES cells with previously published datasets for the interactome for T7-SRSF1 (Akerman *et al*. 2015) shown in orange. (a) Overlap with published SRSF1 data and the endogenous SRSF1-T7 mESC interactome. (b) Overlap between mouse (blue) and human (red) interacting partners of SRSF1 (c) Overlap of published data with those obtained in HeLa cells for T7-SRSF1. (d) Overlap of mouse (blue) and human (red) data with published SRSF1 interactome (orange).

Interaction group	Number of Proteins	Gene name
Common in all datasets	3	BCLAF1, PWP1, SRSF1
mESCs and HeLa o/e	13	ABCF1, CCDC137, DDX31, DKC1, DNAJC9, ISG20L2, MKI67, MRPL11, NIFK, NOL6, REXO4, RRP8, UTP18
mESCs and Akerman et al	10	ACIN1, BUD13, CACTIN, CLASRP, DHX8, PRPF38B, RBM15, SAP30BP, SF3A1, SRSF10
Exclusive to mESCs	15	ABT1, AW146154, BLM, BRD7, C12ORF66, CFAP20, H2AFJ, HP1BP3, MAK16, MRPS30, POLR2H, SGOL2, STAU1, UTP23, YTHDC1

Table 4.4: Overlap of IP-MS targets with known SRSF1 binding partners. Selected protein targets that are common between the three-dataset overlap shown in (d) above.

4.4 Discussion

4.4.1 Investigation of the SRSF1 interactome in human cells

The overwhelming observation from these data is the high degree of common binding partners between the SR proteins, despite their differential shuttling capacities and RNA binding targets. Furthermore, an extremely large number of interactions were observed in these datasets that withstand the application of harsh statistical cut-offs. This remains a large caveat in processing the data to achieve the desired informatic aims and results, most likely, from utilising an overexpression system and whole cell extracts. It is also clear from the empty vector control that the T7-antibody exhibits a high degree of promiscuous binding when exposed to the cell extract, particularly to ribosomal proteins. The latter are a common contaminant in mass spectrometry data partially due to their abundance and here comprise a large proportion of the dataset. For example, discounting ribosomal associated and processing factors, MRP and RP genes alone comprise around a third of the 361 common interacting partners between each bait. A similar pattern is reflected in the GO analysis as associated terms such as translation and ribosome biogenesis are overrepresented.

Out of context, such ribosome protein interactions may be viewed as non-specific, however they were retained in the dataset here as SR proteins are RNA processing factors. It is feasible that they could also play a role in ribosome biogenesis or the processing of ribosomal protein mRNAs. This argument is supported by the large number of proteins in the dataset that are known to function in these processes such as several members of the DEAD-box family of RNA helicases and nucleolar GTPases. It should be noted that such groups of proteins do not bind significantly in the EV control. SR-proteins do bind to polysomes in the cytoplasm (Sanford *et al.*, 2008), but it is also possible that they bind to unassembled or partially synthesised ribosomal subunits in the nucleus, perhaps to influence their export or timing of assembly. In addition to this, RP proteins can have other roles in the cell other than in a ribosomal context. For example, it has been shown that RPL5 interacts specifically with SRSF1 to control a signalling axis in cellular senescence that also involves p53 (Fregoso *et al.*, 2013).

The most significantly overrepresented group of genes in each of the datasets are those associated with RNA processing (Figure 4.2), indicating that the data are a reliable

representation of known SR protein function. SRSF1 and SRSF2 do not interact significantly with each other, despite both binding other SR proteins including their endogenous counterparts and SRSF3, 4, 5, 6,7 and 9. SRSF10 significantly binds T7-SRSF2 and T7-SRSF1-NRS and the latter also binds SRSF11. This reflects the RNA binding profiles of these proteins as determined by CLIP datasets, which show that they have distinct transcript targets for example during splicing. These data may change on the addition of RNase and in the future, it would be valuable to compare protein pull down both with and without RNase treatment as it may facilitate analysis of more subtle interactions that are directly dependent on RNA binding.

There are subtle differences between each dataset, demonstrating that although each bait binds to the same group of factors they might do so with different affinities. This could equate to functional differences or illustrate binding transience. To further delineate this in the future, a quantitative approach such as SILAC would be required as the method used here cannot be used directly to infer interaction stoichiometry.

4.4.2 Endogenous interactome in mouse ESCs

Here work aimed to investigate the interactome of endogenous SRSF1 in mouse embryonic stem cells using the tagged SRSF1-T7 and SRSF1-NRS-T7 (clone 1) cell lines. The latter contains a single copy of a non-shuttling SRSF1 protein, restricted to the nucleus by the NRS from SRSF2, but also expresses an untagged shuttling SRSF1 protein. To identify the proteins bound to both shuttling and non-shuttling SRSF1 protein, IP-MS was performed against the T7-tag.

The striking difference between the mouse and human dataset is the drastically reduced number of SRSF1 binding partners in mESCs. This is most likely due to using an endogenous system rather than overexpressing the protein; particularly as in each experiment, the extracts were prepared using the same conditions. In hindsight, it may have been prudent to optimise the IP-wash conditions to be specific to the mESC IP in order to enrich for loosely associated or transient binding partners of SRSF1. However, it is also possible that the differences in interaction numbers are representative of mouse or stem cell-specific SRSF1 binding profiles.

Comparison of the SRSF1-NRS-T7 and SRSF1-T7 datasets was used to delineate factors that bind endogenous SRSF1 in each cellular compartment. A highly specific interaction with Staufen 1 was observed for SRSF1-T7, but not SRSF1-NRS-T7, suggesting

that it could be an important cytoplasmic interactor of SRSF1 in mESCs. This interaction could not be validated by IP-WB; although this was preliminary data, which requires antibody optimisation to draw definitive conclusions. Despite this, Staufen 1 is an intriguing candidate to pursue in the future, especially as it is not found in previously published data for SRSF1 (Figure 4.9) (Akerman *et al*. 2015). Staufen 1 in s a double-stranded RNA binding protein that has roles in aspects of RNA processing, including Staufen-mediated decay and cytoplasmic RNA localisation (Gong *et al.*, 2009). In mice, it is also required during development for neurogenesis and synapse formation (Vessey *et al.*, 2008). The mRNA targets of Staufen1 include transcripts involved in cell cycle regulation (Boulay *et al.*, 2014) which has also been demonstrated for SRSF1 (Maslon *et al.*, 2014). It is possible that Staufen 1 and SRSF1 bound to a common mRNA or each other, serve to regulate the cell cycle either in a concerted or antagonistic fashion. Although not functionally validated, other cell cycle proteins were also enriched in the cytoplasmic dataset for SRSF1 such as Blm, which is a helicase involved in the DNA damage response (Yusa *et al.*, 2004). Considering that the cytoplasmic roles of SRSF1 are required for mitotic progression (Maslon *et al.*, 2014), this would be an interesting candidate for future validation studies.

Chapter 5

Investigation of SRSF1 function in the cytoplasm using mouse models

5.1 Introduction

5.1.1 Generation of animal models using CRISPR targeting

The existence of animal models in the past has been reliant on traditional gene targeting strategies which can be inefficient both financially and temporally. For such approaches in mice, embryonic stem cells are targeted with a mutation via homologous recombination and then injected into wild type blastocysts to produce chimeric animals containing a germline mutation (Wang *et al.*, 2013). CRISPR/Cas9 targeting has been revolutionary for highly efficient generation of animals carrying customised genomic modifications; dramatically reducing the number of stages required to produce transmitting lines (Graham & Root, 2015). This was exemplified in mice but has been reproduced in numerous diverse species to date from zebrafish and flies to pigs and cattle.

Mouse models can be generated in a single step through direct zygotic injection of the CRISPR/Cas9 machineries. This approach can be used for a variety of applications such as the simultaneous introduction of multiple mutations or generation of single-gene conditional mutants (Wang *et al.*, 2013; Yang *et al.*, 2013).

5.1.2 Physiological roles of SRSF1

SRSF1 is ubiquitously expressed, however, expression level can vary both across tissue types and during development (Hanamura *et al.*, 1998). SRSF1 is known to govern numerous and diverse tissue-specific splicing programmes, which, if dysregulated, have disastrous consequences. Complete ablation of SRSF1 function causes embryonic lethality prior to E9.5 in mice, presumably due to SRSF1 deficient splicing, although no absolute conclusions are ostensible (Xu *et al.*, 2005). Concurrent with this, it was shown that the nuclear functions of SRSF1 are sufficient to rescue lethality in conditional knockout mouse embryonic fibroblasts (MEFs) by transient overexpression of the SRSF1-NRS fusion protein (Lin *et al.*, 2005).

Other conditional SRSF1 knockout mouse models have been previously established, although all with respect to SRSF1 roles in splicing. For example conditional knockout of SRSF1 in the heart causes post-natal lethality due to mis-splicing of CaMKII mRNA that results in a hypercontractile heart phenotype (Xu *et al.*, 2005). Recently a novel role for SRSF1 in proliferation of smooth muscle cells was established using a smooth muscle cell-specific conditional knockout mouse model. In this system, SRSF1 is responsible for the alternative

splicing mediated generation of a truncated form of p53 that drives the expression of KLF5 to promote cell proliferation (Xie *et al.*, 2017).

5.2 Aims

In contrast to SRSF1 nuclear function, little is understood about the cytoplasmic roles of SRSF1 in physiological and developmental context. To examine this further, a mouse model of non-shuttling SRSF1 was designed and implemented using a CRISPR-based strategy. Analogous to the previously discussed mESC targeting, the nuclear retention sequence (NRS) from the non-shuttling SR protein SRSF2 was targeted to the endogenous SRSF1 locus. In addition to this sequence a T7 tag was also added at the 3'end of the insert to facilitate downstream phenotyping.

In summary, the overall aims for the following work are:

(1) To generate a mouse model for the SRSF1-NRS-T7 insertion using CRISPR targeting.
(2) To investigate the phenotypic and physiological consequences of the SRSF1-NRS-T7 allele in models generated.

5.3 Results

5.3.1 CRISPR targeting strategy to generate homozygous SRSF1-NRS-T7 mice

To target mice with the SRSF1-NRS-T7 insertion, the strategy shown in Figure 5.1 and Figure 5.2 was used. CBA x C57BL6 F1 males and females were crossed to generate CBAB6F2 zygotes for targeting. Guide RNAs and Cas9 RNA were microinjected to early mouse embryos at the single-cell stage. When the targeting strategy was implemented, data from mouse ES cells indicated that homozygosity for SRSF1 NRST7 could result in embryonic lethality. For this reason, CRISPR targeting was designed to preferentially generate heterozygous animals that could be crossed to facilitate study of potential embryonic lethal phenotypes. To increase the frequency of heterozygous animals generated, wt SRSF1 and SRSF1-NRS-T7 repair templates were delivered at 1:1 ratio to preserve the wild type SRSF1 allele. To prevent re-cleavage by the Cas9 nuclease after a successful HDR event, the PAM site of the wt SRSF1 oligo was mutated. Integration of the SRSF1-NRS-T7 allele destroys the PAM site inherently thus no mutation is required in this case.

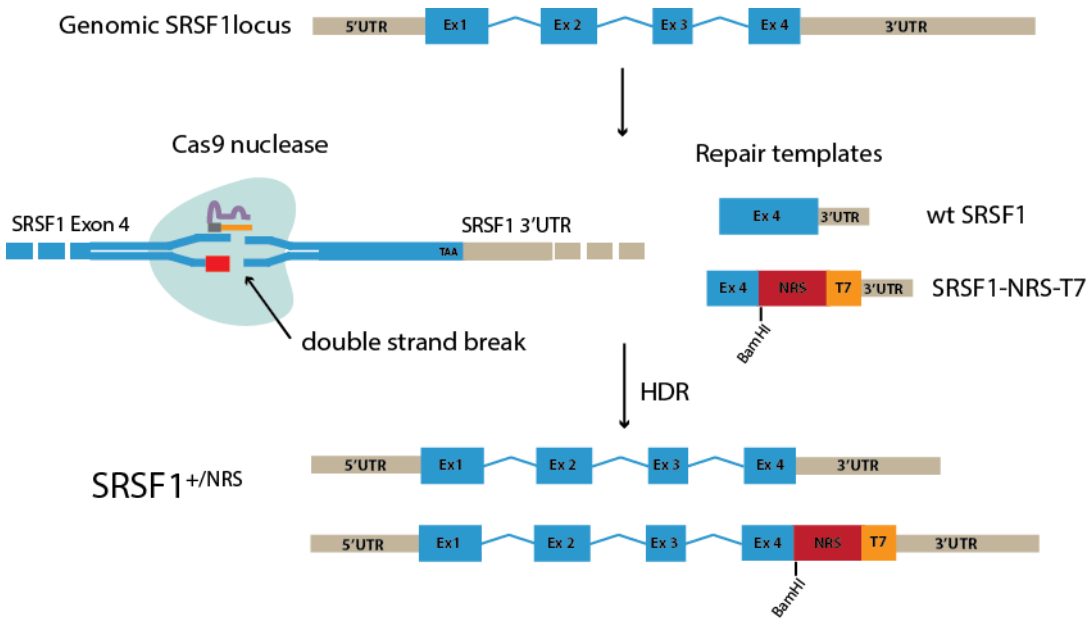


Figure 5.1: CRISPR targeting design for microinjection to mouse zygotes. The genomic SRSF1 locus was targeted in Exon 4. The gRNA was used with Cas9 nuclease RNA to generate a double stranded break. Repair templates were delivered as single stranded DNA oligonucleotides. The wt and NRS oligos were delivered in stoichiometry to generate heterozygous animals as represented by the

targeted locus. Both alleles are shown for clarity. See Materials and Methods for details of sequence design for targeting reagents.

In the approach implemented, the CRISPR reagents could potentially act either prior or subsequent to DNA replication at the single cell stage to generate uniformly mutated or mosaic animals, respectively. Therefore, founder mutant mice from initial screening should be backcrossed to wild type mice to allow any SRSF1 allelic variants to segregate.

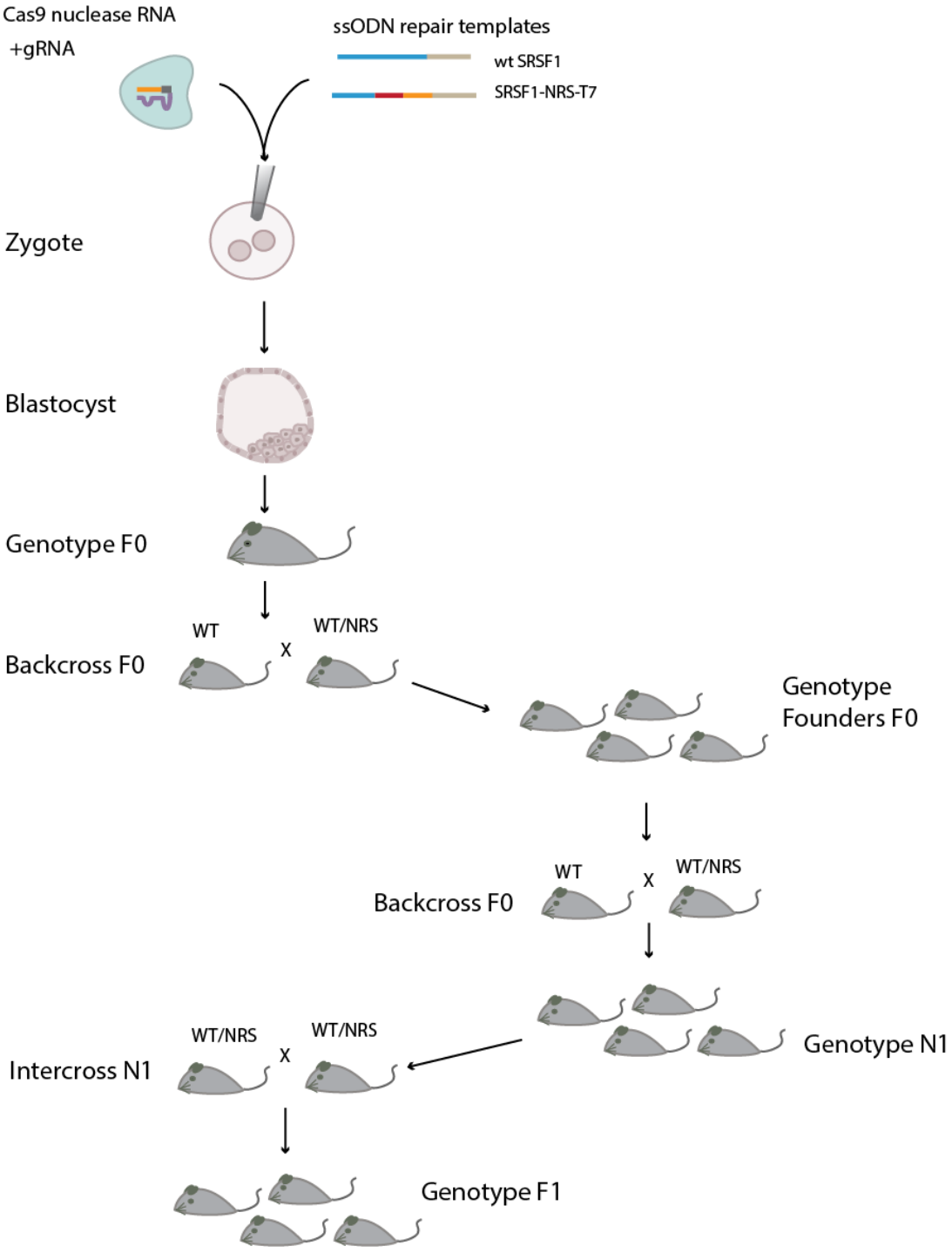


Figure 5.2: Schematic for CRISPR targeting and mouse breeding program. The Cas9 reagents were injected to single cell mouse zygotes to generated mosaic founder animals. Genotyping conformation was performed after each cross. The founders are the N1 generation. A single round of backcrossing was performed as minimum before intercrossing heterozygous mice.

5.3.2 Generation of viable heterozygous SRSF1-NRS-T7 mice

From microinjection CRISPR targeting, 57 pups were born, and genomic DNA from ear-clip tissue subject to PCR screening and subsequent sequence verification of successful mutagenesis events by Sanger sequencing (Figure 5.3). Several examples of successful targeting events can be observed in these data, including generation of heterozygotes such as animals 31 and 36 (Figure 5.3(a)(i)). Mouse 2 appears to be a compound heterozygote as there is no band at 700bp indicative of wild type SRSF1, but two larger bands between 800 and 900bp in size. In addition to this, auxiliary indels were present in some animals such as mouse 10 which contained a large deletion to result in a truncated version of SRSF1.

To ensure there was an equal allelic ratio in heterozygous mice, the PCR product was sub-cloned and 96 representative colonies sequenced for each sample. For example, mouse 2 and 31 contained the SRSF1-NRS-T7 allele in equal proportion with a second allele. Mouse 31 was a true heterozygote with approximately 50% of the PCR sequences each for wild type SRSF1 allele and SRSF1-NRS-T7. Mouse 2 was a compound heterozygote, and contained approximately 50% each of the SRSF1-NRS-T7 allele and an SRSF1-NRS-T7 allele with an additional insertion upstream of the NRS-T7 sequence. These data suggest that the founder mice are not extensively mosaic, at least in ear tissue, from which the genomic DNA was obtained for the screening process.

Two heterozygous founder mice (F0 generation) were backcrossed to C57BL/6 wild type mice. This segregates out any unlinked mutations resulting from the initial targeting that may be present. Considering that CRISPR targeting can have off-target effects that would not be identified in the screening process, this step is paramount and ideally should be repeated several times (Eisener-Dorman *et al.*, 2010).

After a single round of backcrossing, 32 pups were obtained and genotyped as the N1 generation. Of these, 10 animals, for example numbers 709 and 710, were confirmed by sequencing to be heterozygous for the correct SRSF1-NRS-T7 allele (Figure 5.3(c)). In addition, the larger allele from the compound heterozygotes segregated with wild type SRSF1, for example mice 698-700 (Figure 5.3(c)). These animals were not used in further experiments. These data demonstrate that heterozygous mice are viable and were not observed

to display an overt phenotype. Multiple N1 founder mice with the correct heterozygous SRSF1-NRS-T7 insertion were subsequently intercrossed in an attempt to generate F1 homozygous animals (Figure 5.2).

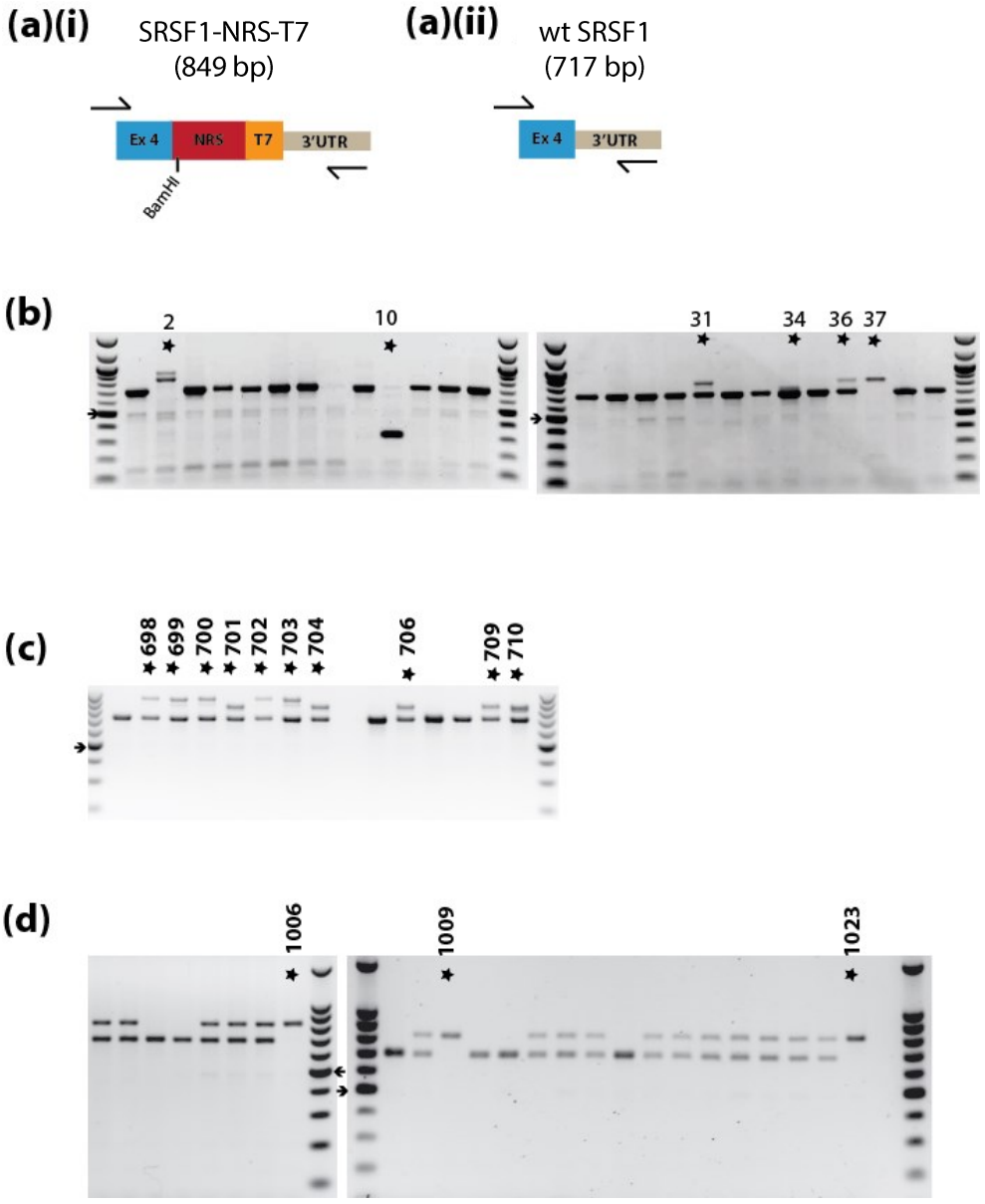


Figure 5.3: Genotyping screens of CRISPR targeted mice. (a) Schematic of PCR screening design to detect (i) SRSF1-NRS-T7 allele or (ii) wt SRSF1 allele. The amplicon size is illustrated in bp. (b) Screening of F0 generation after the initial CRISPR targeting. Representative image shown. (c) Screening of animals after a single round of backcrossing (N1 founder generation). Representative image shown. Asterisk indicates samples containing a genomic insertion that were subject to further investigation. (d) Genotyping SRSF1$^{+/NRS}$ intercross. Representative image shown from first cohort of pups. Arrowhead indicates 500bp, each ladder interval 100bp. Asterisk indicates samples that contain a homozygous NRS-T7 insertion at the genomic SRSF1 locus.

5.3.3 Generation of SRSF1-NRS-T7 homozygous mice is possible

Data from targeting attempts in mouse embryonic stem cells (Chapter 3) suggest that there is intense selection against the generation of homozygous SRSF1-NRS-T7 insertion in this cell type. Embryonic stem cells are representative of the inner cell mass of E3.5 embryos, thus it is conceivable that generation of homozygous animals is highly selected against or occurs with skewed Mendelian genetics. To determine if this is the case, pairs of N1 heterozygous SRSF1$^{+/NRS}$ mice were intercrossed.

In total, 69 pups were born from these crosses and genotyped as previously using PCR screening followed by Sanger sequencing. Surprisingly, 14 homozygous mice were born in addition to 38 heterozygotes and 17 wild type mice (Figure 5.4(a)). These numbers are not statistically different from the expected Mendelian ratio of 1:2:1. Therefore, mice homozygous for the SRSF1 NRS-T7 allele are viable and born at the expected Mendelian ratio. Contrary to expectations based on ES cell data, these data indicate homozygosity for SRSF1-NRS-T7 is compatible with postnatal development.

(a)

Genotype	Observed	Expected
+/+	17	17.25
+/NRS	38	34.5
NRS/NRS	14	17.25
Total	69	69
Chi sq.	0.971	-

(b)

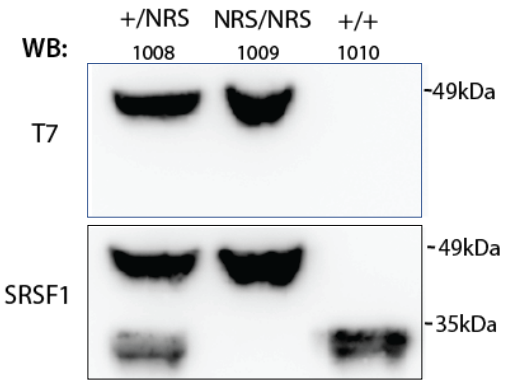


Figure 5.4: Generation of SRSF1 NRS-T7 knockin mice. (a) Homozygous SRSF-NRST7 mice are present at Mendelian ratios, with no significant difference from expected numbers as shown by Chi-squared test. (b) Western Blotting of the SRSF1-NRS-T7 fusion protein, which can be detected by both SRSF1 and T7 antibodies at the expected size.

To ensure that the SRSF1$^{NRS/NRS}$ and SRSF1$^{+/NRS}$ mice generated express the SRSF1-NRS-T7 fusion protein, liver samples from homozygous, heterozygous and wild type mice were subject to Western blotting (Figure 5.4(b)). As expected, SRSF1$^{+/+}$ mice express the canonical SRSF1 isoform as detected at approximately 33kDa using the SRSF1 antibody. A similar band was not detected for SRSF1$^{NRS/NRS}$ mice, instead a larger protein of approximately 49kDa was detected, consistent with the previously observed molecular mass (Chapter 3) for

the SRSF-NRS-T7 fusion protein. Consistent with their genotype, $SRSF1^{+/NRS}$ mice express both the wt and chimeric SRSF1 protein. Data were corroborated through use of the T7 antibody on the same liver tissue samples from each mouse to detect a band at 49kDa for $SRSF1^{NRS/NRS}$ and $SRSF1^{+/NRS}$ mice but not wild type mice. Therefore, in targeted mice, the correctly modified SRSF1 protein is expressed and homozygous mice are reliant on the SRSF1-NRS-T7 chimeric protein.

5.3.3.1 Homozygous SRSF1-NRS-T7 mice develop hydrocephaly

$SRSF1^{NRS/NRS}$ mice are born at the expected Mendelian ratio, however, it rapidly became apparent that these mice are phenotypically abnormal and develop hydrocephaly. From the first cohort of pups generated by the N1 $SRSF1^{+/NRS}$ intercrosses, there were three $SRSF1^{NRS/NRS}$ animals. One of these, mouse number 1023 developed swelling at the top of the head consistent with hydrocephaly at 2-3 weeks of age. Latterly, a second SRSF1-NRS homozygous mouse, number 1009, developed similar swelling at the top of the head indicative of hydrocephaly at 7 weeks of age. These mice were immediately euthanised in accordance with local ethical guidelines. Hydrocephaly can result in pain, distress and neurological damage, thus the remaining SRSF1-NRS-T7 homozygous mouse, number 1006, was euthanised at 7 weeks before any symptoms of hydrocephaly could develop. Similar events were not observed in any of the wild-type or heterozygous littermates in this batch of pups, suggesting that this is specific for $SRSF1^{NRS/NRS}$ animals.

A second cohort of 44 pups was obtained from additional $SRSF1^{+/NRS}$ N1 intercrosses, which were monitored daily for symptoms of hydrocephaly. At the age of 15-16 days, five of this cohort developed swelling at the top of the head analogous to that observed in the previous cohort and were euthanised (figure 5.5). Hydrocephaly is characterised as an accumulation of cerebral spinal fluid (CSF) in brain ventricles that exerts pressure on the skull causing it to bend (Vogel *et al.*, 2012). Consistent with this, pressurised fluid was seen to escape from the brains of these pups during post-mortem dissection.

Subsequent genotyping demonstrated that all five mice that displayed hydrocephalus were homozygous for the SRSF1-NRS-T7 allele. However, all the eight control littermate pups (two wild-type, six heterozygotes) examined at the same stage were devoid of all symptoms of hydrocephaly. Given that hydrocephaly can cause mice to suffer pain and neurological damage, the remaining SRSF1-NRS-T7 homozygous mice were euthanised at 2-3 weeks before any potential hydrocephalus symptoms could arise on authority of the named veterinary surgeon.

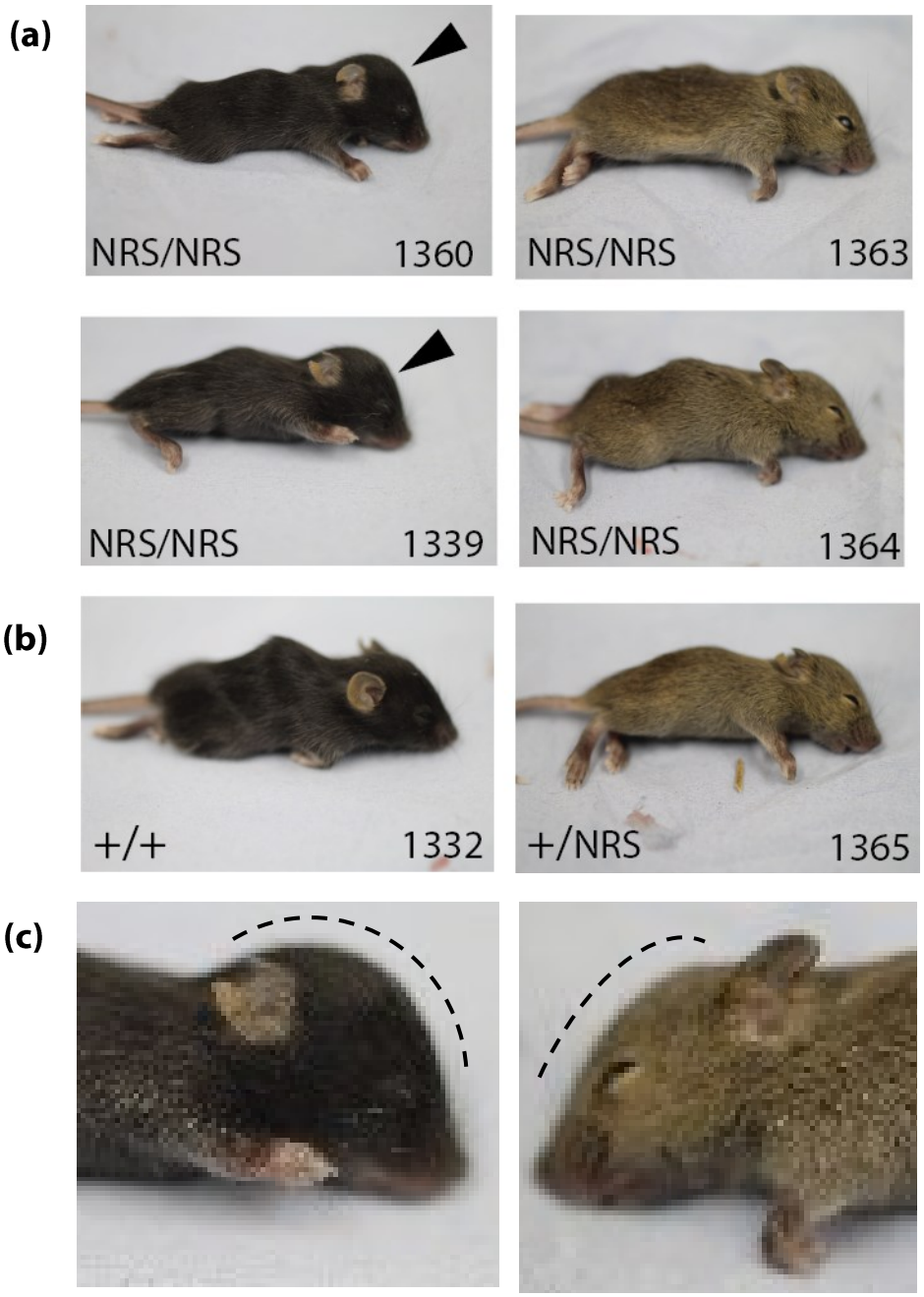


Figure 5.5: Hydrocephaly in SRSF1-NRS-T7 homozygous mice. (a) Homozygous mice, which present obvious bending of the skull indicative of hydrocephaly, are shown on the left-hand side of an unaffected counterpart, which has a normal head morphology. (b) **w**t and heterozygous mice are normal in appearance. Representative images of each genotype shown for comparison. (c) Close up of mice 1360 and 1363 shown in (a) to indicate domed skull shape in the hydrocephalic homozygous mouse.

Taken together, data from the described cohorts of intercrossed mice illustrate that there is a high risk of hydrocephaly associated with presence of a homozygous SRSF1-NRS-T7 insertion; seven of the fourteen homozygous pups developed hydrocephaly within two months. This is significantly different from the incidence of hydrocephaly in heterozygous

and wild type mice from these intercrosses (Fisher's test, $p<0.001$, Figure 5.6). Hydrocephaly can arise spontaneously in the C57BL/6 background, although at a much lower incidence than that observed in these intercrosses (1-3%, (Krebs *et al.*, 2004)).

(a)(i)

Genotype	Hydrocephaly	Unaffected	Total
+/+	0	17	17
NRS/NRS	7	7	14
Total	7	24	31
FET pvalue	0.0013	-	-

(ii)

Genotype	Hydrocephaly	Unaffected	Total
+/NRS	0	38	38
NRS/NRS	7	7	14
Total	7	45	52
FET pvalue	<0.0001	-	-

(b)

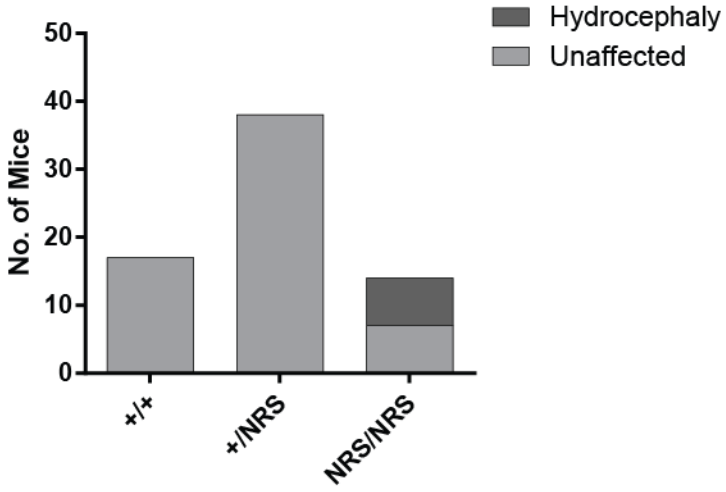


Figure 5.6: Incidence of hydrocephaly in SRSF1-NRS-T7 homozygous mice. (a) Contingency tables to demonstrate the statistically significant incidence of hydrocephaly in SRSF1$^{NRS/NRS}$ mice in comparison to (i) wt or (ii) heterozygous littermate controls. (b) Data presented as stacked bar chart for all three genotypes. Fishers exact test (FET) used for statistical analysis.

In broad terms, hydrocephaly can be categorised either as communicating or non-communicating hydrocephalus, which results from poor CSF movement or a blockage within the ventricles, respectively (Sotak & Gleeson, 2012). However, it can also be caused by an overproduction of CSF (Vogel *et al.*, 2012). At present, the underlying cause or categorisation of this phenotype is unknown. The observed phenotype is unlikely to have a strong sex bias as four of the hydrocephalic SRSF1-NRS-T7 mice were female. The asymptomatic SRSF1-NRS-T7 homozygous pups were euthanised at 2-3 weeks, and one at 7 weeks. Brains from these mice were dissected and fixed in formamide for later investigation. Although the animals did not present any deleterious symptoms at the time of death, it will be particularly interesting

to determine the presence of any pre-clinical signs of hydrocephaly, for example by taking histological sections of fixed brain tissue.

5.3.3.2 Homozygous SRSF1-NRS-T7 mice are growth restricted

In addition to the previously described hydrocephaly phenotype, the SRSF1-NRS-T7 homozygous mice were often visibly smaller than heterozygous and wild type littermates, for example mice 1362 and 1369 (Figure 5.7). To investigate if this was indeed the case empirically, mice from the SRSF1-NRS-T7 heterozygous intercrosses were weighed at 14-18 days old. Of these, all the homozygotes had a reduced body weight in comparison to littermate controls (Figure 5.7). These data suggest that there is a variable growth restriction phenotype associated with homozygosity for SRSF1-NRS-T7.

A metric such as body weight is highly affected by age of the animal, thus by nature, will be more variable between, as opposed to within litters as they were of varying age. To account for this, the weight of homozygous animals was calculated relative to the mean weight of the heterozygotes and wild type mice within each litter. The $SRSF1^{NRS/NRS}$ pups were on average 30% lighter in weight than littermate controls (Student's t-test, $p<0.0001$).

Overall, these data provide evidence to suggest that a homozygous SRSF1-NRS-T7 insertion may interfere with normal growth. In human cells, SRSF1 has roles in the translational regulation of numerous cell cycle proteins and its knockdown induces a multipolar spindle cellular phenotype that can only be rescued by shuttling SRSF1 protein. Considering that the cell cycle is tightly controlled during development and its misregulation can result in deleterious growth (Ciemerych & Sicinski, 2005), it is an intriguing possibility that the lack of cytoplasmic SRSF1 could contribute to the observed growth restriction phenotype in SRSF1 NRS-T7 homozygous mice.

Almost all of the hydrocephalus pups were also small in size with a lower body weight than control littermates. Both phenotypes are as a direct result of homozygosity for SRSF1-NRS-T7, but whether they are linked is yet to be understood. It is possible that they are linked but occur with incomplete penetrance.

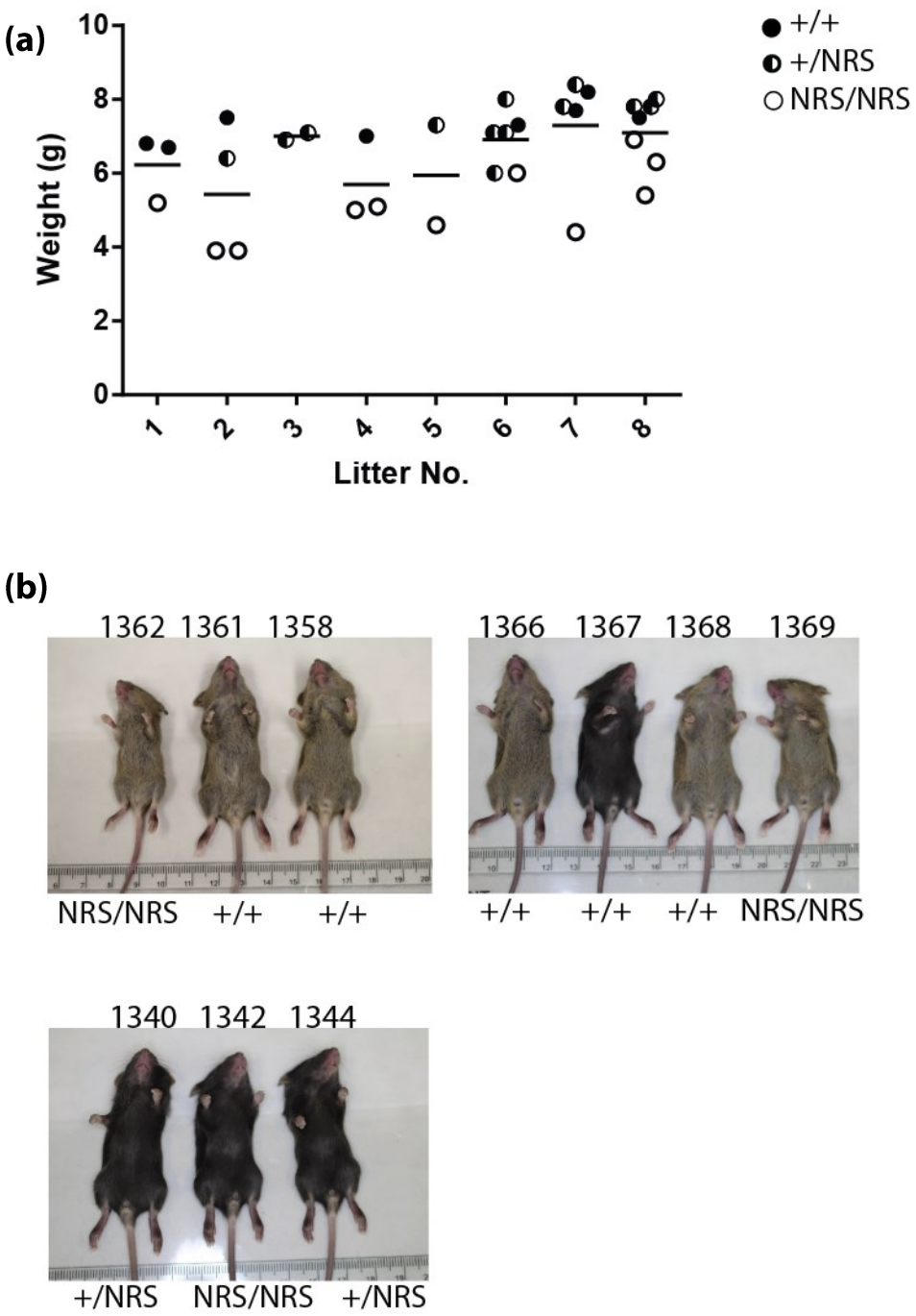
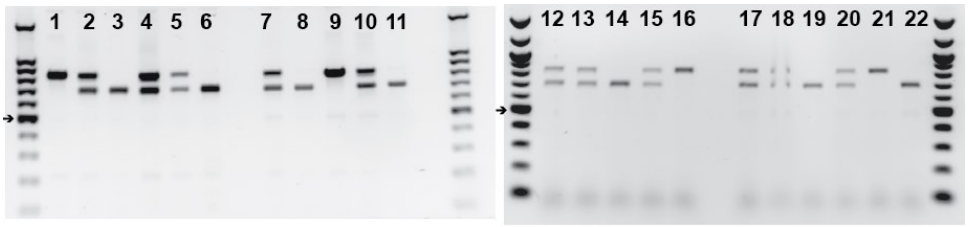


Figure 5.7: Small body size phenotype of SRSF1-NRS-T7 homozygous mice. (a) Weight of each mouse by litter number. Litter 6-8 are a day older than 1-5. (b) Two examples of visibly small homozygous mice adjacent to wild type littermates. The reduced body size phenotype appears to maintain proper proportion of the animal. (c) Example of normal sized SRSF1-NRS-T7 homozygous pup in the centre of two heterozygous littermates.

5.3.4 Embryonic Phenotypes of SRSF1-NRS-T7

To determine if the small body size and hydrocephalus phenotypes observed postnatally originate during embryonic development, heterozygous SRSF1$^{+/NRS}$ F1 mice were bred and E12.5 embryos and placentas were harvested from four female mice. As expected from the postnatal breeding data homozygote SRSF1-NRS-T7 embryos were present at Mendelian ratios (Figure 5.8 (b)). Interestingly, one of the four SRSF1-NRS-T7 homozygous embryos had exencephaly and a defect in neural tube closure around the brain.

(a)



(b)

Genotype	Observed	Expected
+/+	7	5.5
+/NRS	11	11
NRS/NRS	4	5.5
Total	22	22
Chi sq.	1.08	-
pvalue	0.58	-

Figure 5.8: Analysis of E12.5 embryos. (a) Genotype screening of E12.5 embryo tail-clip genomic DNA. The same PCR strategy was used as previously for postnatal screening. (b) Embryos are present at Mendelian ratios, with no significant difference from expected values. Arrowhead indicates 500bp, each ladder interval 100bp.

It is conceivable that this could be related to the high incidence of postnatal hydrocephaly observed in SRSF1-NRS-T7 homozygous mice. However, as only a single homozygote was affected, it is possible that this is either sporadic and unrelated, or there is a range of phenotypic penetrance. In future, it will be interesting to histologically examine the brains of the unaffected homozygous embryos to determine if there are any developmental abnormalities that could contribute to the postnatal hydrocephaly phenotype observed in SRSF1-NRS-T7 mice.

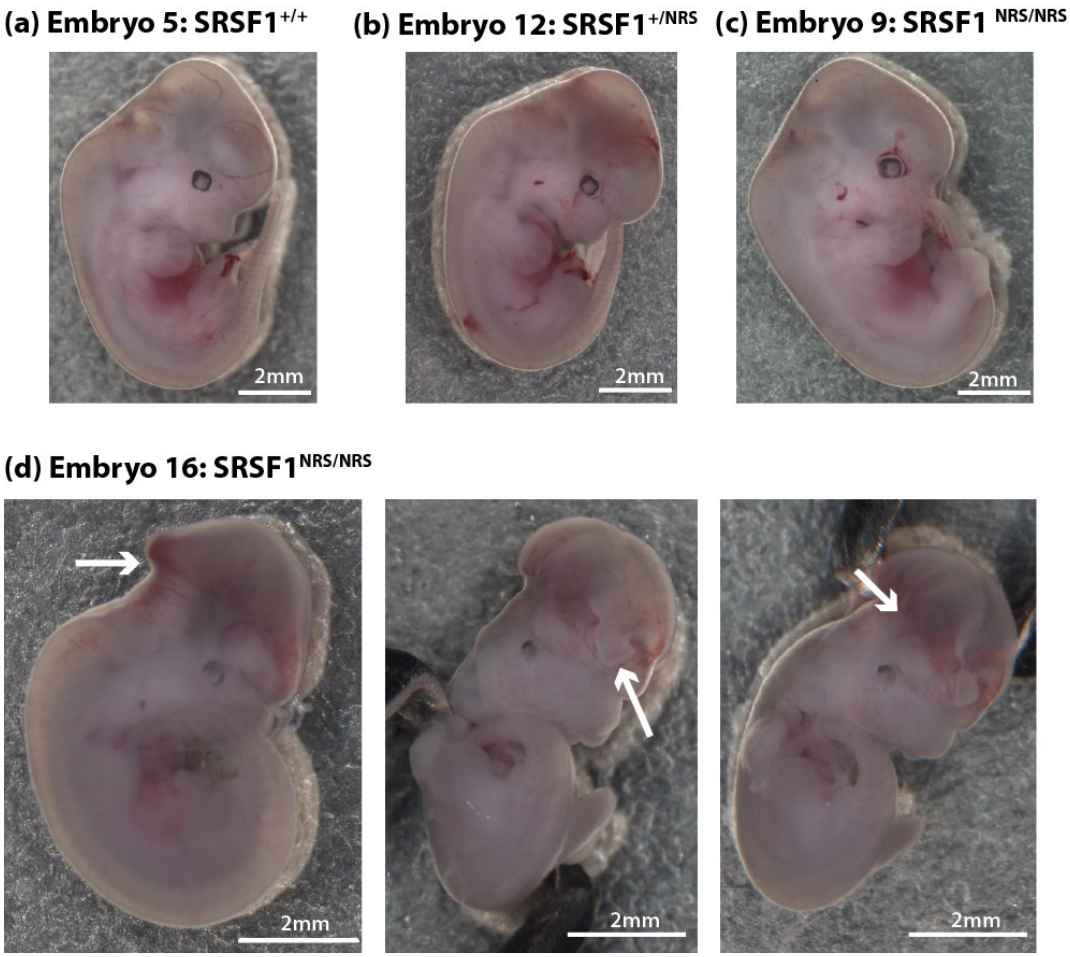


Figure 5.9: Appearance of SRSF1-NRS-T7 E12.5 embryos. Representative image (a) wt, (b) heterozygous and (c) homozygous SRSF1-NRS-T7 embryos observed to be grossly phenotypically normal. (d) SRSF1-NRS-T7 homozygous embryo with exencephaly. There are obvious gross morphological differences in the brain compared to littermates (above). The right hand images of the rotated embryo highlights the observed neural tube defect. Scale bars show represent 2 mm.

To determine if the small body size phenotype observed postnatally in SRSF1-NRS-T7 homozygous mice also occurs during embryogenesis, both embryos and placentas were weighed. 2 of 4 homozygous embryos were clearly smaller than both heterozygous and wild type littermates at E12.5. Embryo weight is strongly affected by slight differences in developmental stage and the maternal uterine environment, thus the weight of $SRSF1^{NRS/NRS}$ embryos was calculated relative to controls (heterozygous and wild type) from the same litter. Note that embryo 15 ($SRSF1^{+/NRS}$) was excluded from analyses as it was damaged during processing. The same calculation was performed for placenta weight. These data demonstrate that on average, SRSF1-NRS-T7 homozygous embryos are 22.8% smaller than the control embryos and the reciprocal placentas were 19% smaller than that of the controls (data not

shown). However, only 2 of 4 of the SRSF1-NRS-T7 homozygous embryos are affected (Figure 5.10). To determine if there is indeed a significant bimodal distribution of weight in homozygous embryos, further numbers of embryos will be required for analysis. It is conceivable that the growth restriction phenotype observed at E12.5 in the SRSF1$^{NRs/NRS}$ embryos contributes to the small body size observed postnatally.

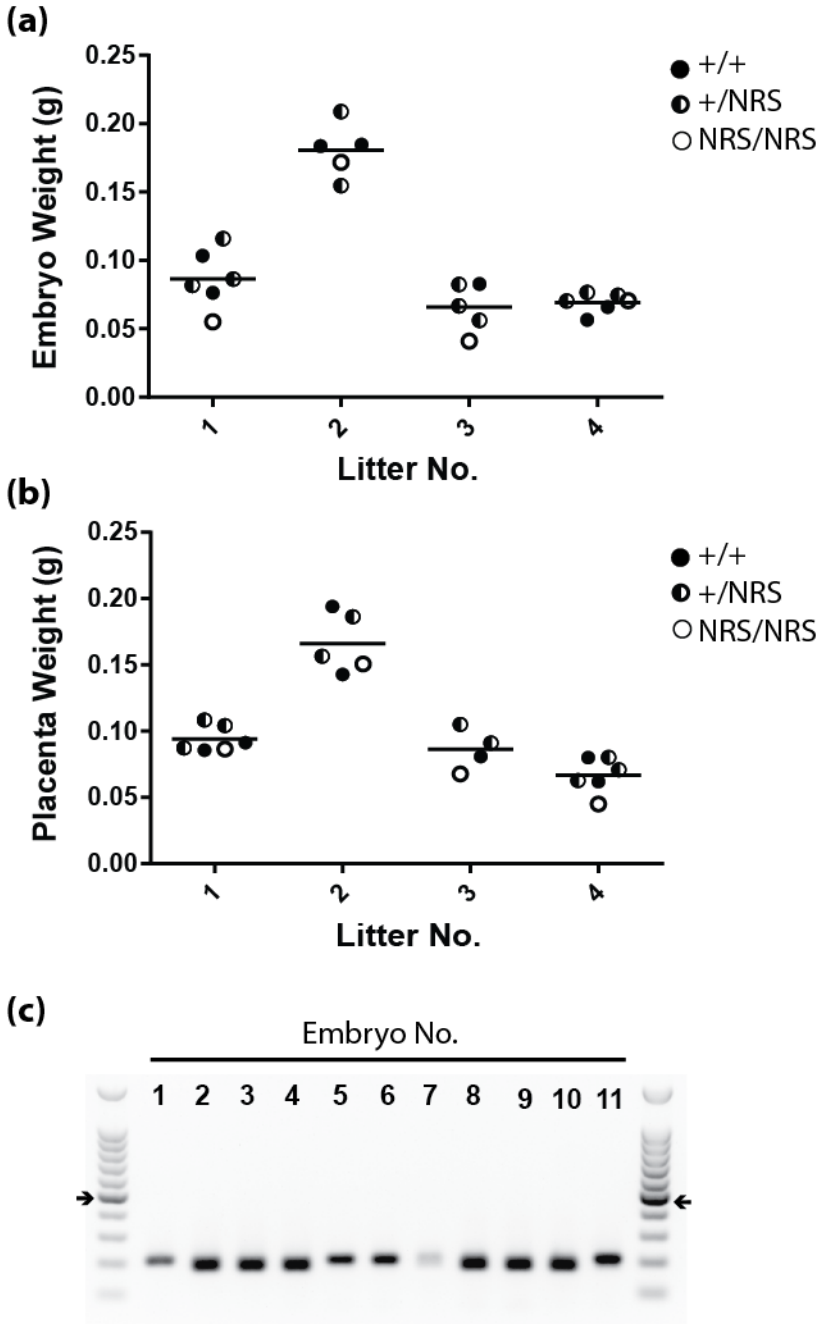


Figure 5.10: Embryo Weight at E12.5. (a) Half of the SRSF1-NRS-T7 homozygous embryos have a lower body weight than littermate controls. Note that the small heterozygote in litter two was damaged during the weighing process thus its weight appears lower than reality. (b) Placenta weight for each embryo shown in (a). Weight does not correlate with genotype. (c) the observed lower weight is not sex-linked in embryos as demonstrated by amplification of the UBE1 locus. Arrowhead indicates 500bp, each ladder interval 100bp.

The small body size observed in postnatal homozygous SRSF1-NRS-T7 mice was not sex linked, however considering that only a subset of embryos had low body weight, it is possible that it could be linked to sex at E12.5. To determine if this is the case, half of the embryos were sexed by PCR amplification of the UBE1 locus from genomic DNA obtained from embryo tail tissue. The UBE1 gene is encoded on the X/Y chromosomes and its amplification results in a larger (approximately 200bp) or smaller amplicon in females and males respectively. Of the eleven embryos tested, two were homozygous with low body weight: embryo 1 and 9 which are female and male respectively as determined by a larger or smaller PCR amplicon (Figure 5.10(b)). Therefore, concurrent with data from postnatal mice, the observed phenotypes in homozygous SRSF1-NRS-T7 embryos do not appear to correlate with gender (Figure 5.10(b)). However as only half of the embryos were tested, it will be necessary to sex the other embryos to fully conclude that this is the case.

5.4 Discussion and conclusions

The aim of this work was to investigate the physiological roles of the nucleocytoplasmic shuttling protein SRSF1. To do so, both heterozygous and homozygous SRSF1-NRS-T7 knock-in mice were successfully generated via CRISPR targeting at the endogenous SRSF1 locus. The insertion is comprised of a T7 epitope tag for screening purposes and a nuclear retention sequence (NRS) from the closely related non-shuttling SR protein SRSF2, that when fused to SRSF1 prevents nucleocytoplasmic shuttling (Cazalla *et al.*, 2002).

These data reveal that there are several, obvious and severe characteristics associated with homozygosity for SRSF1-NRS-T7, although in contrast heterozygous animals appear phenotypically normal. Assuming that *in vivo* the SRSF1-NRS-T7 fusion protein behaves as previously described (Cazalla *et al.*, 2002) and it is not capable of nucleocytoplasmic shuttling; aberrant phenotypes in SRSF1$^{NRS/NRS}$ mice are likely due to lack of shuttling SRSF1 function. This indicates the functional importance of SRSF1 presence in the cytoplasm, or at least, outside its roles in splicing. The latter should be unaffected by the C-terminal NRS fusion as the N-terminal RRM domains of SRSF1 are required for splicing function, in particular substrate specificity (Cléry *et al.*, 2013). However, to derive absolute conclusions from these data, it is essential that knock-in mice are subject to RNA-sequencing to ensure that, as predicted, splicing remains largely unaffected. Further to this, heterokaryon assays must be performed on cells from knock-in animals to determine the ablation of shuttling of

the endogenous SRSF1-NRS-T7 fusion protein. Thus, at present, alternative roles of SRSF1 in the nucleus, but not in splicing cannot be dismissed, such as mRNA stability or export.

Homozygous mice are highly susceptible to developing hydrocephaly, initially observed by exaggerated doming of the skull due to accumulation of cerebrospinal fluid (Vogel *et al.*, 2012). One of the major causes of hydrocephalus is defective cilia in the brain and it can be observed in numerous classes of ciliopathies. In healthy individuals motile cilia in the brain contract to drive movement of CSF through the ventricular system, although hydrocephaly can also arise from defective immotile primary cilia for example in Bardet-Biel syndrome (Sotak & Gleeson, 2012). The development of hydrocephalus in mouse models has previously been linked to the development of abnormal body movement, for example in the hydrocephalus with hop gait (hyh) mutant mouse (Bronson & Lane, 1990). It is possible that the homozygous mice described here to suffer hydrocephaly could also display similar phenotypes if monitored closely. It would be interesting to assay, for example the gait of homozygous mice, to determine if hydrocephaly is accompanied by ataxia or degenerative movements indicative of extensive CNS problems.

Interestingly, several diverse RNA binding proteins have been described to cause hydrocephaly on knockout. The Musashi family of RNA binding proteins are developmentally regulated and expressed specifically in neural precursor lineages. Mice lacking these in the brain develop obstructive hydrocephaly (Sakakibara *et al.*, 2002). Hydrocephaly is also a phenotype in PTB brain conditional knockout mice due to specific loss of ependymal cilia in the brain and loss of adherens junctions in specific neural lineages (Shibasaki *et al.*, 2013). The Rbfox2 RNA binding protein has also been implemented in the regulation of brain development and its perturbation results in deleterious phenotypes including hydrocephaly (Gehman *et al.*, 2012).

In addition to the hydrocephaly phenotype, homozygous mice are visibly smaller in size than both wild type and heterozygous counterparts. The latter appear unaffected indicating that expression of a single copy of shuttling SRSF1 facilitates normal growth. The cause of these observations remains an open question, although there are mouse models of small body size in existence, several of which are caused by misregulation of the cell cycle during development (Ciemerych & Sicinski, 2005). SRSF1 regulates the mRNA translation, in particular that of mRNA transcripts that encode cell cycle and mitotic proteins (Michlewski *et al.*, 2008; Maslon *et al.*, 2014). This raises the interesting possibility that the translational roles of SRSF1 in regulation of such transcripts could be linked to the observed phenotypes. Notably, homozygous deletion of the S6K1 kinase results in a small size phenotype and low

body weight (Pende *et al.*, 2004). S6K1 is an important regulator of mRNA translation. Considering that SRSF1 also functions to regulate the translation of specific mRNAs, it is an intriguing speculation that an analogous scenario is occurring in SRSF1-NRS-T7 homozygous mice. However, a significant proportion of the mouse knockout strains in existence have reduced body weight as a phenotype indicating that this is a highly variable and multigenic trait (Reed *et al.*, 2008). Considering this, it could be that the effects on body size in $SRSF1^{NRS/NRS}$ mice are due to several perturbations of SRSF1 cytoplasmic function or the translational misregulation of numerous target mRNAs.

To investigate the potential physiological roles of the SRSF1-NRS-T7 allele during prenatal development, E12.5 embryos were subject to analysis. Analogous to the postnatal dataset, homozygote embryos are present at Mendelian ratios. A subset of the homozygous embryos are small in size in comparison to control littermates, however further numbers are required to determine if this is statistically significant. Furthermore, one of the homozygous embryos had exencephaly, meaning development of the brain outside of the skull. Given the observed postnatal hydrocephaly phenotype it is of great interest to investigate potential links between the two in the future. For example, sectioning and histological analysis of the unaffected embryos may facilitate the detection of latent pre-clinical symptoms of brain developmental abnormalities. In addition, examination of postnatal brains of unaffected homozygous mice may also similarly inform on any specific consequences of homozygosity for SRSF1-NRS-T7 in development of the CNS.

In summary, the data presented here clearly demonstrate the successful generation of an animal model to determine the physiological consequences of the SRSF1-NRS-T7 allele. Preliminary analyses of this model demonstrate that there are significant consequences that result from the manipulation of the SRSF1 shuttling function *in vivo*. In the future, it will be highly interesting to mechanistically determine the causes for the observed phenotypes with regard to SRSF1 function. Considering the fundamental role that SRSF1 has in development and disease, this will be of extreme physiological and fundamental relevance.

Chapter 6

Discussion

RNA binding proteins are crucial for the regulation of all aspects of gene expression in a highly dynamic and physiologically contextual manner. The SR protein family are an essential group of such factors that are evolutionarily conserved in metazoans in terms of both structure and function (Long and Cáceres, 2009). SR proteins were originally described as splicing factors, required for both constitutive and alternative splicing. SRSF1 is the archetypal SR protein and represents a subset of the SR family that continuously shuttles between the nucleus and the cytoplasm. Such dynamics are governed by the continual phosphorylation of the eponymous Serine/Arginine (SR) repeats of the C-terminal RS domain, whereas the RNA-binding function of SRSF1 is mediated by the N-terminal RRM domains.

Further to its nuclear roles, SRSF1 is involved in numerous other cellular processes outside splicing and indeed in the cytoplasm. Importantly, SRSF1 is involved in mRNA translation and is directly associated with the translation machinery. The translational functions of SRSF1 are exerted in a previously outlined mTOR-dependent mechanism (Michlewski *et al* 2008). Latterly, studies in human cells defined the direct translational mRNA targets of SRSF1, including those encoding RNA processing and splicing factors but also, remarkably, numerous cell cycle and mitotic proteins (Maslon *et al.*, 2014).

SR proteins, including SRSF1, are known to exhibit tissue specific expression profiles and their phosphorylation state differs during development (Hanamura *et al.*, 1998; Sanford & Bruzik, 2001). Despite this, little is understood regarding the cytoplasmic roles of SRSF1 in a physiological context. To address this, both cellular and animal models were developed in which SRSF1 is exclusively nuclear.

6.1 Development of cellular models to study SRSF1 cytoplasmic function

The nuclear roles of SRSF1 in splicing are essential for cell viability and thus were maintained in model systems generated. The non-shuttling SR protein SRSF2 contains a potent nuclear retention sequence (NRS) at the very C-terminus of its RS domain (Cazalla *et al.*, 2002). When fused to SRSF1 in the context of the SRSF1-NRS fusion protein, it is sufficient to ablate shuttling while maintaining cellular viability. The primary aim of the work herein was to develop an endogenous model system that harbours the non-shuttling SRSF1 protein to study the physiological consequences of the absence of cytoplasmic SRSF1. To do so, CRISPR gene targeting was used to knockin the SRSF2 NRS to the endogenous SRSF1 locus to generate both cellular and animal model systems of SRSF1-NRS-T7.

6.1.1 Generation of non-shuttling SRSF1 is highly selected against in mESCs

Despite extensive rounds of targeting, it was not possible to generate homozygous SRSF1-NRS-T7 knockin mouse ES cells with the CRISPR strategy implemented, potentially due to extremely poor knockin efficiencies. However, this is unlikely as the endogenous locus of SRSF1 is amenable to targeting, as illustrated by the successful generation of both homozygous and heterozygous SRSF1-T7 knockin mESCs. Furthermore, only SRSF1-NRS-T7 Clone 1 cells could be generated which do contain the NRS-T7 insertion but with additional mutations that results in heterozygous SRSF1-NRS-T7 protein expression. High throughput sequencing of CRISPR targeted mESCs populations demonstrated that the generation of homozygous SRSF1-NRS-T7 clones is fundamentally possible, but targeted cells rapidly die over time due to intense negative selection pressures meaning just a single clone was recovered (Chapter 3, 3.3.6). Taken together, this strongly suggests that undifferentiated mouse ESCs struggle to tolerate a lack of shuttling SRSF1 under the growth conditions required to maintain pluripotency.

The SRSF1-NRS-T7 mESC targeting strategy used here, was rationalised by previous work in which the SRSF1-NRS fusion protein was sufficient to preserve viability in conditional SRSF1 knockout mouse embryonic fibroblasts (MEFs) (Lin *et al.*, 2005). MEFs are, in many aspects, physiologically diverse from mESCs. For instance, they are fully differentiated, which may alter cells' tolerance for ablation of SRSF1 cytoplasmic function. Interestingly, in the previously published study, MEFs complemented with a version of SRSF1 that lacked the RS domain were viable but grew slowly. Furthermore, the shuttling capacity of the ΔRS SRSF1 protein could not be determined, but it was present in both the nucleus and the cytoplasm (Lin *et al.*, 2005) The RS domain is required for SRSF1 function in translation (Michlewski *et al* 2008), thus it is an intriguing possibility that ablation of SRSF1 cytoplasmic function causes growth retardation and represents partial lack of tolerance for cytoplasmic SRSF1 in this scenario.

A recent study proposed that the tendency for SR proteins to exhibit nucleocytoplasmic shuttling is dictated by cellular differentiation state. It was demonstrated that all SR proteins are capable of shuttling to variable extents in pluripotent P19 mouse cells, but not HeLa cells. SRSF1 was considered to have high shuttling ability, whereas SRSF2 had a lower ability to shuttle (Botti *et al.*, 2017). Consequently, it is possible that in mESCs, the NRS sequence does not act with similar potency in the SRSF1-NRS fusion protein as in HeLa cells to prevent shuttling completely. Nevertheless, the NRS is expected to reduce shuttling

ability of SRSF1 in mESCs to an extent that will perturb cytoplasmic SRSF1 function sufficiently to cause aberrant consequences. Alternatively, considering that in P19 cells SRSF2 shuttling is passive and coupled to binding mRNA targets for export that are distinct from SRSF1 (Muller-McNicoll *et al.*, 2016; Botti *et al.*, 2017), the isolated NRS domain, that does not contribute to SRSF2-specific mRNA binding, is likely to confer complete nuclear retention in context of the SRSF1-NRS-T7 fusion protein analogous to differentiated cells (Cazalla *et al.*, 2002).

In either scenario, this indicates that undifferentiated cells may be more reliant on SRSF1 shuttling function which could provide explanation regarding the observed negative selection in generating homozygous SRSF1-NRS-T7 mESCs. There are several possible reasons for the latter. Undifferentiated cells have a distinct and rapid cell cycle profile from differentiated cells, characterised by a short G1 phase and altered cell-cycle checkpoints (Boward *et al.*, 2016). In human cells, SRSF1 regulates the translation of cell-cycle related mRNA transcripts and its knockdown causes mitotic defects which are independent from its roles in splicing (Maslon *et al.*, 2014). It is plausible that in mESCs, SRSF1 acts in analogous mechanisms to regulate mitosis through translation of specific targets.

A single copy of non-shuttling SRSF1 does not affect the overall cell cycle profile of targeted mESCs, as SRSF1-NRS-T7 Clone 1 cells appear normal; as expected from the presence of wt SRSF1 in these cells (Chapter 3, Figure 3.10). However, it is possible that complete lack of shuttling SRSF1 protein lowers the tolerance threshold for mitotic problems sufficiently to promote cell cycle arrest in targeted cells. Indeed, such problems cannot be rescued by nuclear retained SRSF1. Feasible mRNA targets of SRSF1 whose downregulation could cause an aberrant cell cycle phenotype, include the cyclin kinase CDK1, or the condensin complex proteins SMC2 and SMC4 (Maslon *et al.*, 2014). These proteins are highly conserved in mice and essential for proper cell cycle progression and chromosome segregation. Therefore, if the absence of SRSF1 results in reduced levels of translation of corresponding mRNA transcripts, there may be insufficient protein for consistent, accurate cell proliferation.

6.2 Interactome studies of shuttling and non-shuttling SR proteins

The potent nuclear retention sequence (NRS) naturally present in SRSF2 can ablate shuttling function when utilised in a chimeric protein (Cazalla *et al.*, 2002). However, it is a

unique peptide sequence and the mechanism of its function is currently unknown. To determine any protein interactors that could be responsible for the NRS function in nuclear retention, IP-MS studies were performed in both human and mouse ES cells.

6.2.1 Factors for nuclear retention of SRSF1-NRS-T7

Work described in Chapter 2 to determine the interactomes of SRSF1, SRSF-NRS and SRSF2 illustrate that on overexpression, these SR proteins bind and share a significantly high number of protein binding partners. Data could be corroborated by a clear overlap with a previously published dataset for SRSF1 in the same cell type (Akerman, *et al*. 2015). Consistent with the known nuclear roles of SR proteins, gene ontology analysis using the data obtained demonstrated an enrichment for RNA processing factors and those involved in splicing. The datasets obtained from IP-MS studies in HeLa cells were overlapped. In theory, factors that significantly bind to SRSF1-NRS and SRSF2 but not SRSF1 could be involved in nuclear retention.

Taking this approach, few proteins were identified, presumably because the proportion of interactions shared between all three datasets was so large. Indubitably, an interaction conferring nuclear retention could be with a shared binding factor but with differential affinity. In future, it would be interesting to use a quantitative MS approach such as SILAC to delineate if this could be the case. Alternatively, it is possible that the mechanism of nuclear retention for the NRS is not dependent on protein binding but perhaps RNA-binding specificities or events that dictate bound mRNA nuclear export. For example, mRNP remodelling at the nuclear pore through an auxiliary factor could remove SRSF2 from an mRNA to which both proteins are bound, independent of physical association with SRSF2. It is also possible that posttranslational modification of the SRSF2 protein and/or specifically the NRS could contribute to nuclear retention function. Such a model would be consistent with those previously described for SR proteins, which is dependent on RS domain dephosphorylation (Lin *et al.*, 2005). In general, phosphorylation is crucial to SR protein function, but other modifications are also present. Indeed in pluripotent cells, SRSF2 movements are also proposed to be dependent on arginine methylation (Botti *et al.*, 2017).

6.2.2 SRSF1 has species-specific protein binding partners

To determine the endogenous binding partners of SRSF1 and the SRSF1-NRS-T7, the CRISPR-generated mouse ESC cell lines SRSF1-T7 and SRSF1-NRS-T7 (Clone 1) were

used in similar interactome studies as human cells (i.e. IP-MS approach). The overwhelming observation from these data was the significant reduction in binding partners identified in either dataset in comparison to the equivalent in the overexpression system used for SRSF1 in HeLa cells. Indeed, only 16 common proteins were identified that bound T7-tagged SRSF1 in human and mouse cells (Figure 5.9). This implies either that SRSF1 overexpression could amplify promiscuous or transient binding partners, or that in mouse cells SRSF1 has a diverse interactome.

Considering the observed negative selection in generating homozygous SRSF1-NRS-T7 knockin mESCs, it was interesting to determine if the interactome of shuttling SRSF1-T7 cells could implicate binding partners responsible and thus important in SRSF1 cytoplasmic function. In total 25 cytoplasmic partners of SRSF1 were inferred as those which bound the shuttling SRSF1-T7 but not the nuclear retained version. Of these, none were strikingly apparent as those involved in translation. However, the proteins that did not appear in any previously known dataset for SRSF1 or are significantly enriched in SRSF1-NRS-T7 Clone 1 cells include factors such as BRD7, BLM, C12ORF66 and CFAP20. The Bloom helicase (BLM) is involved in sister chromatid exchange and DNA-damage response pathways and is frequently dysregulated in cancer (Luo *et al.*, 2000). Considering the roles of SRSF1 in mitosis and as an oncogene, it would be an interesting target to follow up in the future (Karni *et al.*, 2007; Maslon *et al.*, 2014). Furthermore, the BLM mRNA transcript was found to be translationally regulated by SRSF1 (Maslon *et al.*, 2014), which raises interesting possibility of a functional protein and mRNA interactome overlap.

6.2.3 SRSF1-NRS-T7 and mRNA export

At present, it is unknown from data presented here, whether homozygosity or SRSF1-NRS-T7 could influence SRSF1 function in mRNA export. It is possible that the observed negative selection in the generation of homozygous SRSF-NRS-T7 result from abnormal export of SRSF1 target transcripts. Nearly all members of the SR protein family (SRSF1-7) act as adapters for the NXF1-mediated export of specific mRNAs. Crucially, each SR protein serves to regulate distinct mRNA transcripts in this manner (Muller-McNicoll *et al.*, 2016). Interestingly, SRSF1 is involved in the export of the repeat-expansion transcript C9ORF72, which is involved in the development of neurological diseases in humans. In *Drosophila* models, SRSF1 depletion prevents neurotoxicity through ablation of C9ORF72 export to prevent toxic RNA foci accumulation in the cytoplasm (Cooper-knock *et al.*, 2014; Hautbergue *et al.*, 2017).

In the IP-MS data obtained here, ALREF, NXF1, CHTOP export adapter proteins bind to SRSF1-NRS, SRSF1 and SRSF2 suggesting that NRS does not affect protein interactions with export factors. Considering this and the fact that SRSF1 target mRNA binding should be unaffected by the NRS sequence, it is unlikely that export is dysregulated to the extent that it could demonstrate the phenotypes observed. Corroborating this data, the export factor RBM15 binds to SRSF1-NRS and SRSF1 but not SRSF2 suggesting this in an SRSF1-specific interaction but one unaffected by NRS binding. In the mESC endogenous interactome data, RBM15 and CHTOP also bind to endogenous SRSF1 and SRSF1-NRS with similar binding affinity (at least in IP-MS enrichment).

6.3 Generation of a mouse model to study SRSF1 cytoplasmic function

To investigate the relevance of SRSF1 shuttling function in a physiological and developmental model, CRISPR targeting of mouse zygotes was used to generate an endogenous model for SRSF1-NRS-T7. Contrary to expectations from mESC targeting data, homozygous SRSF1-NRS-T7 mice are born at normal Mendelian ratios. However, it was rapidly apparent that these mice exhibit several striking postnatal phenotypes that are absent in heterozygous or wild type littermates.

6.3.1 SRSF1-NRS-T7 knockin mice develop hydrocephaly

Homozygous SRSF1-NRS-T7 mice present severe postnatal hydrocephaly characterised by abnormal doming of the skull in comparison to unaffected littermates (Chapter 5 Figure 5.5), (Vogel *et al.*, 2012)). Hydrocephalus is a common phenotype of both primary and motile cilia dysregulation in the brain; it is observed in numerous ciliopathies which are often caused by multiple genes and are not transmitted in a Mendelian pattern (Valente *et al.*, 2013). The latter is consistent with the observed variable penetrance of hydrocephaly in SRSF1-NRS-T7 homozygous mice over the time monitored (Figure 5.5).

Primary cilia are organelles found at the cell surface that transmit diverse biological signals including morphogenic cues during development such as Sonic Hedgehog and Wnt signalling pathways (Valente *et al.*, 2013). Motile cilia in the brain are distinct from primary cilia and are present in several parts of the brain to exert diverse function. (Spassky & Meunier, 2017). An important example when understanding hydrocephaly are the cilia in the ependyma of the ventricles which promote the continual movement of cerebrospinal fluid (CSF). If this

is dysregulated or blocked, the CSF accumulation can be sufficient to cause hydrocephalus. On the other hand, overproduction of CSF can also result in hydrocephalus (Narita & Takeda, 2015).

Primary cilia such as those in the brain are laid down during embryogenesis from the modification of mother centrioles in a cell cycle dependent mechanism (Banizs *et al.*, 2005; Spassky & Meunier, 2017). SRSF1 has been linked to centriole biogenesis although the mechanism of this action is yet to be identified (Balestra *et al.*, 2013). Given that centrioles and cilia formation are tightly linked, it is conceivable that this is relevant for the phenotypes observed in homozygous SRSF1-NRS-T7 mice. Perhaps a lack of SRSF1-mediated translation of specific centriole components could result in dysregulated cilia biogenesis during SRSF1-NRS-T7 homozygous mice development. This could contribute to the observed neural tube defect in one of the E12.5 homozygous embryos (Figure 5.8). Interestingly in HeLa cells, SRSF1 was found to regulate the translation of both CDC42SE2 and DAAM1 mRNA (Maslon *et al.*, 2014), which are both effectors in different branches of the Wnt signalling pathway; dysregulation of which has been implicated in neural tube defects (Keller, 2002). It is tempting to speculate that this could contribute to both the embryonic and postnatal phenotypes observed in the homozygote SRSF1-NRS-T7 mice. Furthermore, both SRSF1 and SRSF9 – which incidentally is the closest SR protein paralogue of SRSF1 (Sun *et al.*, 2010)- can manipulate the Wnt/β-catenin branch of the Wnt signalling pathway. In this context, SRSF1 overexpression enhances the level of β-catenin protein through translational regulation of the reciprocal mRNA transcript in an mTOR dependent mechanism (Fu *et al.*, 2013).

Numerous other signalling pathways have also been implicated in the development of hydrocephaly such as the planar cell polarity pathway (PCP) and mTOR signalling (Spassky & Meunier, 2017). Furthermore, hydrocephaly can be of the communicating or non-communicating form both of which are underpinned by diverse and complex molecular mechanisms It will be crucial to determine into which category the observed hydrocephaly falls and if any of the likely signalling pathways is affected. This could have relevance to SRSF1 function, particularly in the translation regulation of pathway-associated factors.

Several other RNA binding proteins and splicing factors have been shown to cause hydrocephaly or aberrant neurogenesis when their function is mutated or lost such as PTBP1, RBFOX1, MBNL2 and SRRM4 (Vuong *et al.*, 2016). The dysregulation of PTBP1 function in radial glial cells can cause aberrant neural differentiation thus it would be interesting to see if SRSF1 may have similar functions in these processes (Vuong *et al.*, 2016). Interestingly, in rats, SRSF1 is predominantly cytoplasmic in dorsal root ganglion sensory neurons at steady

state but translocates to the nucleus on traumatic nerve injury. The authors propose that cytoplasmic SRSF1 in these cells is inactive, but there is no further comment on functions it may perform there (Hulse *et al.*, 2014, 2016).

6.3.2 SRSF1-NRS-T7 knockin mice are small in size

In addition to developing hydrocephalus, SRSF1-NRS-T7 mice were observed to have a considerable reduction in body size in comparison to both heterozygous and wild type littermates. This is in proportion over the entire body of the mouse and appears to present both postnatally and at the E12.5 stage of embryo development (Figure 5.7). Similar to the hydrocephalus phenotype, this is not fully penetrant within litters and does not appear to be sex-linked. It is unclear whether affected mice would have reached comparable weights as control animals if they survived longer into adulthood.

Size of an adult organism can be governed by either proliferation of individual cells, or the overall number of cells generated during development that comprise the adult organism. In turn, this is regulated by a finely tuned balance between apoptosis and cell proliferation in existence during development that can also contribute to overall body size (Yang & Xu, 2011; Lloyd, 2013). At present, the causes of the observed small phenotype in SRSF1-NRS-T7 homozygous mice are unknown; however, considering the known cytoplasmic roles of SRSF1, several speculations could be made. For example, SRSF1 is involved in both translational regulation and cell cycle control; two processes inherently linked to cell size, proliferation rate and thus overall organism size.

The IGF/PI3K/AKT/mTOR signalling axis is heavily involved in the regulation of cell growth and thus cell and overall organism size (Lloyd, 2013). Several studies have demonstrated that ablation of members of this pathway such as mTOR itself or S6K1 cause small cell and body size, which if viable result in mice that display an overall reduction in body size (Gangloff *et al.*, 2004; Pende *et al.*, 2004). Moreover, ablation of S6K1 specifically in skeletal muscle does not affect cell proliferation but is essential for the increase in cell size (Ohanna *et al.*, 2005). It will be interesting to determine in the future if SRSF1-NRS-T7 homozygous cells exhibit a slower growth rate or if the total cell number is less and if this is linked to SRSF1 function in the mTOR signalling pathway (Michlewski *et al* 2008).

Other signalling pathways that can be dysregulated in terms of size involve the kinase Akt, which acts upstream of PI3K. The latter is a crucial determinant in heart size in adult mice which inherently implicates Akt in this process (Shioi *et al.*, 2000). Akt is also known to

phosphorylate SRSF1 in the cytoplasm (Blaustein *et al.*, 2005), thus perhaps SRSF1 could acts as a similar downstream effector in regulating organ size. Maybe if there is no effector of Akt activity then expression of certain transcripts is dysregulated (Blaustein *et al.*, 2005; White *et al.*, 2010). With regards to organ size, it is interesting to note that SRSF3 conditional knockout hepatocytes grow slowly and fail to differentiate (Sen *et al.*, 2013). Similar to SRSF1, SRSF3 is capable of cytoplasmic function and can influence mRNA translation of specific targets (Kim *et al.*, 2013; Park & Jeong, 2016); tempting speculations regarding both SRSF1 and SRSF3 function in mRNA translation during proper organogenesis.

Ablation of core components of the cell cycle in mice cause embryonic lethality at various stages of both pre and postnatal development (Ciemerych & Sicinski, 2005). However, depletion of key cell cycle regulation genes, although not essential for viability, also has marked effects on body size in mouse models that could prove relevant for SRSF1 translational roles. For example, Cyclin D1 knockout mice are smaller than wild type littermates and had other skeletal abnormalities (Fantl *et al.*, 1995). In addition, mice lacking the phosphatases cdc25a and cdc25c, which mediate CDK function during the cell cycle are also small in size but do not display defects in cell cycle progression (Ferguson *et al.*, 2005). SRSF1 enhances the translation of the cyclin D kinase 1 (CDK1) and CDC27, which regulate cyclin functions to enable cell cycle progression and proper mitosis respectively (Maslon *et al.*, 2014). Taken together, it is highly feasible that cell cycle control is implicated in the size phenotype observed in SRSF1-NRS-T7 homozygous mice and it would be intriguing to interrogate cells from these mice for cell cycle abnormalities.

It is also possible that SRSF1 could enhance the translation of crucial mitogens and/or growth factors during particular developmental stages such as EGF or IGF to influence growth rate or downstream metabolism. Indeed, preliminary data implicate the cytoplasmic roles of SRSF1 in the regulation of cellular glucose uptake and other metabolic processes (M.Maslon, unpublished data). Consistent with this, SRSF1 regulates the translation of both the fatty acid Acyl-CoA synthase, AASDH and FUT8, an alpha-(1,6)-fucosyltransferase (Maslon *et al.*, 2014). Interactome data for SRSF1 (Chapter 2) also indicates that SRSF1 is capable of binding mitochondrial proteins, including numerous components of the mitochondrial ribosome in both HeLa and mouse ES cells. It is possible that SRSF1 translationally regulates such metabolite production or metabolic enzymes that if deficient could result in poor energy uptake in cells.

6.3.3 Further comments for observed phenotypes of SRSF1-NRS-T7 mice

Although the phenotypes observed in homozygous mice are severe, it appears that they are of variable penetrance and not sex-linked. For example, some mice were small with no hydrocephalus, some just hydrocephalus, some completely visually unaffected and so on. It is unknown if asymptomatic homozygous mice that were euthanised, would have developed problems with age. It is possible that these mice already had pre-clinical symptoms that were not visually apparent. Furthermore, perhaps the lack of SRSF1 shuttling function during development could contribute to the postnatal phenotypes observed. A feasible scenario could be that SRSF1 regulates the translation of an important developmental-stage specific transcription or splicing factor, which if expressed with incorrect dosage could disturb the balance of downstream gene expression profiles. With this is mind, further complexities could also be realised, considering that SRSF1 expression is likely to be highly tissue and developmentally specific.

6.4 Physiological relevance of SRSF1 shuttling function

Data presented here demonstrate that homozygosity for the SRSF1-NRS-T7 allele is viable but highly selected against in undifferentiated cells and results in serious phenotypic consequences in a mouse model. Contrary to data from previously described models using mouse embryonic fibroblasts (Lin *et al.*, 2005), this indicates that the ability of SRSF1 to shuttle is crucial for proper development and cellular homeostasis. Indeed, the nucleocytoplasmic shuttling is required for the oncogenic potential of SRSF1 (Shimoni-Sebag *et al.*, 2013) indicating that cytoplasmic function could have relevance for disease models of SRSF1 dysregulation. In future, the model systems created herein for studying SRSF1 function could serve as tools to dissect the molecular mechanisms that contribute to cytoplasmic SRSF1-mediated oncogenesis. Thus, the SRSF1-NRS-T7 model could be highly relevant to further understand molecular mechanisms in cancer.

In conclusion, this work indicates that the nucleocytoplasmic shuttling functions of SRSF1 are highly physiologically relevant and SRSF1 function outside those in splicing should be considered with equal importance. Henceforth, it will be truly exciting to use the novel SRSF1-NRS-T7 model system to further and fully dissect the cytoplasmic function of SRSF1 *in vivo*.

6.5 Future Work

The work presented in this thesis, describes the creation of a mouse model for non-shuttling SRSF1 that has striking phenotypic aberrations. Due to time constraints, causes for the latter have only received preliminarily investigations thus far; ergo the natural progression is to fully characterise the developmental and molecular mechanisms that could be responsible for the phenotypes observed in the homozygous SRSF1-NRS-T7 mice generated.

Prior to commencing phenotypic investigations, there are several pressing controls that should be undertaken. Firstly, it is crucial to perform RNA-sequencing on cells derived from SRSF1-NRS-T7 homozygous mice to ensure that SRSF1-mediated splicing remains grossly unaffected in comparison to wild type cells. Given that SRSF1 acts to regulate the expression of target transcripts during multiple and inherently coupled processing steps, it is possible that there are subtle changes in splicing. Nonetheless, SRSF1 function in splicing is essential for viability thus if there were gross splicing changes in the SRSF1-NRS-T7 mice they would not be viable. Although subtle changes in alternative splicing through insertion of the NRS-T7 to endogenous SRSF1 cannot be discounted when delineating phenotypic causes, most of the published evidence indicates that the non-shuttling SRSF1 protein should be fully functional in nuclear roles. The strongest evidence for this being that the SRSF1-NRS fusion protein could fully rescue MEFs deficient in shuttling SRSF1, suggesting that nuclear roles of SRSF1 were grossly unperturbed (Cazalla *et al.*, 2002; Lin *et al.*, 2005).

Secondly, it is necessary to ensure that in the model generated, the NRS functions as previously described and confers nuclear retention to the SRSF1-NRS fusion protein. To do this, heterokaryon assays should be further optimised and repeated to demonstrate the knockin protein is indeed incapable of shuttling into the cytoplasm.

Lastly, it cannot be yet ruled out that the NRS-T7 insertion is affecting mRNA export in the mouse model created thus it is necessary to eliminate this as a phenotypic cause. The 225 unique export targets of SRSF1 in p19 cells have been previously identified by the overlapping of RNA-seq from whole cell extracts and cytoplasmic fractions on SR protein depletion (Muller-McNicoll *et al.*, 2016). A similar approach could be used to ensure that export of these mRNAs is unaffected in homozygous mice perhaps with or without a block of export by for example depletion of NXF1 or a similar mRNA export factor. Furthermore, overlapping SRSF1 export targets and the known translational targets may indicate transcripts whose export is either linked to translation or export is affected by the presence of the NRS insertion.

6.5.1 Creating cell lines from knockin SRSF1-NRS-T7 mice

To further understand the molecular and biochemical mechanisms that underpin the cytoplasmic roles of SRSF1, and to alleviate the requirement to continuously breed mice, it will be necessary to derive stably maintained cell lines as an *in vitro* model of the system created. The logical starting point from a technical perspective, is to derive mouse embryonic fibroblasts (MEFs) from homozygous SRSF1-NRS-T7 knockin mouse embryos, which would provide a differentiated cellular system for further analysis. To continue from work in Chapter 3, it would also be interesting to determine if it is possible to derive viable and stably pluripotent homozygous SRSF1-NRS-T7 mESCs, either derived from reciprocal homozygous mice or by reprogramming SRSF1-NRS-T7 MEFs. However, this is more technically challenging and large numbers of mice are required, which limits the potential attempt number. If achieved, it would be interesting to determine if homozygous SRSF1-NRS-T7 cells are capable of differentiation into for example neurons, which could be used to investigate the potential cause of hydrocephaly.

To continue from work in Chapter 4, and pursue both the protein factors responsible for nuclear retention and those that SRSF1 binds in the cytoplasm. It would be interesting to perform quantitative mass spectrometry such as SILAC for endogenous SRSF1 in both wt and SRSF-NRS-T7 cell lines. Indeed, if generation of mESCs from the mouse model is possible, it would be prudent to use these and compare resultant data to those obtained from the SRSF1-T7 mESCs.

6.5.2 Translational roles of SRSF1 in the SRSF1-NRS-T7 mouse model

All previous datasets for SRSF1 function in translation and its targets therein were obtained in differentiated human cell lines. However, mRNA translation and its regulation is not wholly ubiquitous between species or cellular differentiation state. Therefore, it is feasible that SRSF1 regulates diverse, specific mRNA transcripts in adult mice or during mouse development in comparison to those found for human cells. This is of particular interest given the aberrant phenotypes of SRSF1-NRS-T7 knockin mice.

During mRNA translation, the ribosome protects a small fragment of associated mRNA termed a footprint which is typically 25 to 30 nucleotides in length. Recently, numerous methods, namely ribosome profiling and variations thereof, have been described to accurately measure translation state at nucleotide resolution under given conditions (Ingolia *et al.*, 2009, 2011; Ingolia, 2016). Ribosome profiling takes advantage of the ribosome

footprint by freezing ribosomes on the bound mRNA using cyclohexmide then subjecting bound transcripts to deep sequencing analysis. In the future, such methods could be used to determine the translational targets of SRSF1 in the mouse model created.

This could be approached in several ways. To investigate SRSF1 in translation initiation and the 40S subunit scanning process, translation complex profiling (TCP-seq) could be used, which is a modified version of the ribosome profiling protocol that enriches for pre-initiation complexes on the mRNA (Archer *et al.*, 2016; Shirokikh *et al.*, 2017). Alternatively, use of Harringtonine to freeze initiating ribosomes would derive similar information (Ingolia *et al.*, 2012). Furthermore, ribosome footprinting and sequencing of monosome and polysomal fractions separately would facilitate quantitative understanding of the translational abundance of SRSF1 translationally regulated transcripts (Heyer & Moore, 2016). It would also be interesting to perform ribosome profiling coupled to cell synchronization to reveal quantitative translatome dynamics during the cell cycle and in mitosis, which could be highly relevant in context of SRSF1 (Tanenbaum *et al.*, 2015).

Chapter 7

Bibliography

Aitken, C.E., Beznosková, P., Vlčkova, V., Chiu, W.L., Zhou, F., Valášěk, L.S., Hinnebusch, A.G., & Lorsch, J.R. (2016) Eukaryotic translation initiation factor 3 plays distinct roles at the mrna entry and exit channels of the ribosomal preinitiation complex. *Elife*, **5**, 1–37.

Aitken, C.E. & Lorsch, J.R. (2012) A mechanistic overview of translation initiation in eukaryotes. *Nat. Struct. Mol. Biol.*, **19**, 568–576.

Akaike, Y., Kurokawa, K., Kajita, K., Kuwano, Y., Masuda, K., Nishida, K., Wan Kang, S., Tanahashi, T., & Rokutan, K. (2011) Skipping of an alternative intron in the srsf1 3' untranslated region increases transcript stability. *J. Med. Investig.*, **58**, 180–187.

Akerman, M., Fregoso, O.I., Das, S., Ruse, C., Jensen, M. a., Pappin, D.J., Zhang, M.Q., & Krainer, A.R. (2015) Differential connectivity of splicing activators and repressors to the human spliceosome. *Genome Biol.*, **16**, 119.

Anczuków, O., Akerman, M., Cléry, A., Wu, J., Shen, C., Shirole, N.H., Raimer, A., Sun, S., Jensen, M.A., Hua, Y., Allain, F.H.-T., & Krainer, A.R. (2015) SRSF1-Regulated Alternative Splicing in Breast Cancer. *Mol. Cell*, **60**, 105–117.

Anczuków, O., Rosenberg, A.Z., Akerman, M., Das, S., Zhan, L., Karni, R., Muthuswamy, S.K., & Krainer, A.R. (2012) The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat. Struct. Mol. Biol.*, **19**, 220–228.

Andreev, D.E., O'Connor, P.B.F., Loughran, G., Dmitriev, S.E., Baranov, P. V., & Shatsky, I.N. (2017) Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Res.*, **45**, 513–526.

Änkö, M.-L., Morales, L., Henry, I., Beyer, A., & Neugebauer, K.M. (2010) Global analysis reveals SRp20- and SRp75-specific mRNPs in cycling and neural cells. *Nat. Struct. Mol. Biol.*, **17**, 962–970.

Änkö, M.-L., Müller-McNicoll, M., Brandl, H., Curk, T., Gorup, C., Henry, I., Ule, J., & Neugebauer, K.M. (2012) The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.*, **13**, R17.

Änkö, M.L. (2014) Regulation of gene expression programmes by serine-arginine rich splicing factors. *Semin. Cell Dev. Biol.*, **32**, 11–21.

Archer, S.K., Shirokikh, N.E., Beilharz, T.H., & Preiss, T. (2016) Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature*, **535**, 570–574.

Arribere, J.A., Cenik, E.S., Jain, N., Hess, G.T., Lee, C.H., Bassik, M.C., & Fire, A.Z. (2016) Translation readthrough mitigation. *Nature*, 1–17.

Ast, G. (2004) How did alternative splicing evolve? *Nat. Rev. Genet.*, **5**, 773–782.

Atkinson, G.C., Baldauf, S.L., & Hauryliuk, V. (2008) Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components. *BMC Evol. Biol.*, **8**, 290.

Aubol, B.E., Chakrabarti, S., Ngo, J., Shaffer, J., Nolen, B., Fu, X.-D., Ghosh, G., & Adams, J. a (2003) Processive phosphorylation of alternative splicing factor/splicing factor 2. *Proc.*

Natl. Acad. Sci. U. S. A., **100**, 12601–12606.

Aubol, B.E., Hailey, K.L., Fattet, L., Jennings, P.A., & Adams, J.A. (2017) Redirecting SR Protein Nuclear Trafficking Through An Allosteric Platform. *J. Mol. Biol.*,.

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

Aubol, B.E., Wu, G., Keshwani, M.M., Movassat, M., Fattet, L., Hertel, K.J., Fu, X.-D., & Adams, J.A. (2016) Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. *Mol. Cell*, 1–11.

Aviner, R., Hofmann, S., Elman, T., Shenoy, A., Geiger, T., Elkon, R., Ehrlich, M., & Elroy-Stein, O. (2017) Proteomic analysis of polyribosomes identifies splicing factors as potential regulators of translation during mitosis. *Nucleic Acids Res.*, **45**, 5945–5957.

Baierlein, C., Hackmann, A., Gross, T., Henker, L., Hinz, F., & Krebber, H. (2013) Monosome formation during translation initiation requires the serine/arginine-rich protein Npl3. *Mol. Cell. Biol.*, **33**, 4811–4823.

Balestra, F., Strnad, P., Flückiger, I., & Gönczy, P. (2013) Discovering regulators of centriole biogenesis through siRNA-based functional genomics in human cells. *Dev. Cell*, **25**, 555–571.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C., & Landthaler, M. (2012) The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol. Cell*, **46**, 674–690.

Banizs, B., Pike, M.M., Millican, C.L., Ferguson, W.B., Komlosi, P., Sheetz, J., Bell, P.D., Schwiebert, E.M., & Yoder, B.K. (2005) Dysfunctional cilia lead to altered ependyma and choroid plexus function, and result in the formation of hydrocephalus. *Development*, **132**, 5329–5339.

Bentley, D.L. (2014) Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.*, **15**, 163–175.

Berget, S.M., Moore, C., & Sharp, P.A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.*, **74**, 3171–3175.

Bermingham, J.R., Arden, K.C., Naumova, a K., Sapienza, C., Viars, C.S., Fu, X.D., Khotz, J., Manley, J.L., & Rosenfeld, M.G. (1995) Chromosomal localization of mouse and human genes encoding the splicing factors ASF/SF2 (SFRS1) and SC-35 (SFRS2). *Genomics*,.

Bertram, K., Agafonov, D.E., Liu, W.-T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., & Lührmann, R. (2017) Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature*, **542**, 318–323.

Björk, P., Jin, S.B., Zhao, J., Singh, O.P., Persson, J.O., Hellman, U., & Wieslander, L. (2009) Specific combinations of SR proteins associate with single pre-messenger RNAs in vivo and contribute different functions. *J. Cell Biol.*, **184**, 555–568.

Blaustein, M., Pelisch, F., Tanos, T., Muñoz, M.J., Wengier, D., Quadrana, L., Sanford, J.R.,

Muschietti, J.P., Kornblihtt, A.R., Cáceres, J.F., Coso, O. a, & Srebrow, A. (2005) Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat. Struct. Mol. Biol.*, **12**, 1037–1044.

Booth, D.G., Takagi, M., Sanchez-Pulido, L., Petfalski, E., Vargiu, G., Samejima, K., Imamoto, N., Ponting, C.P., Tollervey, D., Earnshaw, W.C., & Vagnarelli, P. (2014) Ki-67 is a PP1-interacting protein that organises the mitotic chromosome periphery. *Elife*, **3**, e01641.

Bothmer, A., Phadke, T., Barrera, L.A., Margulies, C.M., Lee, C.S., Buquicchio, F., Moss, S., Abdulkerim, H.S., Selleck, W., Jayaram, H., Myer, V.E., & Cotta-Ramusino, C. (2017) Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat. Commun.*, **8**, 13905.

Botti, V., McNicoll, F., Steiner, M.C., Richter, F.M., Solovyeva, A., Wegener, M., Schwich, O.D., Poser, I., Zarnack, K., Wittig, I., Neugebauer, K.M., & Müller-McNicoll, M. (2017) Cellular differentiation state modulates the mRNA export activity of SR proteins. *J. Cell Biol.*, jcb.201610051.

Boucher, L., Ouzounis, C. a, Enright, a J., & Blencowe, B.J. (2001) A genome-wide survey of RS domain proteins. *RNA*, **7**, 1693–1701.

Boulay, K., Ghram, M., Viranaicken, W., Trépanier, V., Mollet, S., Fréchina, C., & Desgroseillers, L. (2014) Cell cycle-dependent regulation of the RNA-binding protein Staufen1. *Nucleic Acids Res.*, **42**, 7867–7883.

Bourgeois, C.F., Mortreux, F., & Auboeuf, D. (2016) The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat. Rev. Mol. Cell Biol.*, **17**, 426–438.

Boward, B., Wu, T., & Dalton, S. (2016) Concise Reviews : Control of Cell Fate Through Cell Cycle and Pluripotency Networks. *Stem Cells*,.

Bressan, R.B., Dewari, P.S., Kalantzaki, M., Gangoso, E., Matjusaitis, M., Garcia-Diaz, C., Blin, C., Grant, V., Bulstrode, H., Gogolok, S., Skarnes, W.C., & Pollard, S.M. (2017) Efficient CRISPR/Cas9-assisted gene targeting enables rapid and precise genetic manipulation of mammalian neural stem cells. *Development*, **44**, dev.140855.

Bronson, R.T. & Lane, P.W. (1990) Hydrocephalus with hop gait (hyh): a new mutation on chromosome 7 in the mouse. *Dev. Brain Res.*, **54**, 131–136.

Burdon, T., Smith, A., & Savatier, P. (2002) Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol.*, **12**, 432–438.

Cáceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L., & Krainer, A.R. (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J. Cell Biol.*, **138**, 225–238.

Cáceres, J.F., Screaton, G.R., & Krainer, A.R. (1998) A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev.*, **12**, 55–66.

Calabretta, S. & Richard, S. (2015) Emerging Roles of Disordered Sequences in RNA-Binding Proteins. *Trends Biochem. Sci.*, **40**, 662–672.

Cammas, A., Lacroix-Triki, M., Pierredon, S., Le Bras, M., Iacovoni, J.S., Teulade-Fichou, M.-P., Favre, G., Roché, H., Filleron, T., Millevoi, S., & Vagner, S. (2016) hnRNP A1-mediated translational regulation of the G quadruplex-containing RON receptor

tyrosine kinase mRNA linked to tumor progression. *Oncotarget*, **7**, 16793–16805.

Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.M., Foehr, S., Curk, T., Krijgsveld, J., & Hentze, M.W. (2016) Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell*, **63**, 696–710.

Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A.R., & Cáceres, J.F. (2002) Nuclear export and retention signals in the RS domain of SR proteins. *Mol. Cell. Biol.*, **22**, 6871–6882.

Ceccaldi, R., Rondinelli, B., & D'Andrea, A.D. (2016) Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol.*, **26**, 52–64.

Chen, L., Weinmeister, R., Kralovicova, J., Eperon, L.P., Vorechovsky, I., Hudson, A.J., & Eperon, I.C. (2016) Stoichiometries of U2AF35, U2AF65 and U2 snRNP reveal new early spliceosome assembly pathways. *Nucleic Acids Res.*, **45**, gkw860.

Chen, X., Rinsma, M., Janssen, J.M., Liu, J., Maggio, I., & Gonçalves, M.A.F.V. (2016) Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res.*, **44**, 6482–6492.

Cheng, Y., Luo, C., Wu, W., Xie, Z., Fu, X., & Feng, Y. (2016) Liver-specific deletion of SRSF2 caused acute liver failure and early death in mice. *Mol. Cell. Biol.*, **36**, MCB.01071-15.

Cherepkova, M.Y., Sineva, G.S., & Pospelov, V.A. (2016) Leukemia inhibitory factor (LIF) withdrawal activates mTOR signaling pathway in mouse embryonic stem cells through the MEK/ERK/TSC2 pathway. *Cell Death Dis.*, **7**, 1–10.

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

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., & Kim, J. (2014) Analysis of off-target effects of CRISPR Cas-derived RNA-guided endonucleases and nickases sup2. *Genome Res.*, **24**, 132–141.

Choi, Y.D. & Dreyfuss, G. (1984) Isolation of the heterogeneous nuclear RNA-ribonucleoprotein complex (hnRNP): a unique supramolecular assembly. *Proc. Natl. Acad. Sci. U. S. A.*, **81**, 7471–7475.

Choi, Y. Do, Grabowski, P.J., Sharp, P.A., & Dreyfuss, G. (1986) Ribonucleoprotems : Role in RNA Splicing. *Science*, **231**, 1534–1539.

Chow, L.T., Gelinas, R.E., Broker, T.R., & Roberts, R.J. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, **12**, 1–8.

Chu, J., Cargnello, M., Topisirovic, I., & Pelletier, J. (2016) Translation Initiation Factors: Reprogramming Protein Synthesis in Cancer. *Trends Cell Biol.*, **26**, 918–933.

Ciemerych, M. a & Sicinski, P. (2005) Cell cycle in mouse development. *Oncogene*, **24**, 2877–2898.

Cléry, A., Sinha, R., Anczuków, O., Corrionero, A., Moursy, A., Daubner, G.M., Valcárcel, J., Krainer, A.R., & Allain, F.H.-T. (2013) Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc.*

Natl. Acad. Sci. U. S. A., **110**, E2802–E2811.

Cohen-Eliav, M., Golan-Gerstl, R., Siegfried, Z., Andersen, C.L., Thorsen, K., Ørntoft, T.F., Mu, D., & Karni, R. (2013) The splicing factor SRSF6 is amplified and is an oncoprotein in lung and colon cancers. *J. Pathol.*, **229**, 630–639.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L. a, & Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.

Conti, L., Pollard, S.M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.L., Cattaneo, E., & Smith, A. (2005) Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.*, **3**, 1594–1606.

Cooper-knock, J., Walsh, M.J., Higginbottom, A., Highley, J.R., Dickman, M.J., Edbauer, D., Ince, P.G., Wharton, S.B., Wilson, S.A., Kirby, J., Hautbergue, G.M., & Shaw, P.J. (2014) Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain*, **137**, 2040–2051.

Cordin, O. & Beggs, J.D. (2013) RNA helicases in splicing. *RNA Biol.*, **10**, 83–95.

Cowling, V.H. (2010) Regulation of mRNA Cap Methylation. *Biochem. J.*, **1**, 576–579.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., & Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics MCP*, **13**, 2513–2526.

Das, S. & Krainer, A.R. (2014) Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Mol. Cancer Res.*,.

Dever, T.E. & Green, R. (2015) Phases of Translation in Eukaryotes 1–16.

Ding, J.-H., Xu, X., Yang, D., Chu, P.-H., Dalton, N.D., Ye, Z., Yeakley, J.M., Cheng, H., Xiao, R.-P., Ross, J., Chen, J., & Fu, X.-D. (2004) Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J.*, **23**, 885–896.

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

Dominguez, D., Tsai, Y., Weatheritt, R., Wang, Y., Blencowe, B.J., Wang, Z., Hill, C., Comprehensive, L., Biology, C., & Centre, D. (2016) An Extensive Program of Periodic Alternative Splicing Linked to Cell Cycle Progression. *Elife*,.

Doudna, J.A. & Charpentier, E. (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.).*, **346**, 1258096–1258096.

Easley, C.A., Ben-yehudah, A., Redinger, C.J., Oliver, S.L., Varum, S.T., Eisinger, V.M., Carlisle, D.L., Donovan, P.J., & Schatten, G.P. (2010) mTOR-Mediated Activation of p70 S6K Induces Differentiation of Pluripotent Human Embryonic Stem Cells. *Cell. Reprogram.*, **12**.

Eisener-Dorman, A., Lawrence, D., & Bolivar, V. (2010) Cautionary insite on knockout mouse studies: the gene or not the gene. *Brain, Behav. Immun.*, **23**, 318–324.

Ellis, J.D., Llères, D., Denegri, M., Lamond, A.I., & Cáceres, J.F. (2008) Spatial mapping of

splicing factor complexes involved in exon and intron definition. *J. Cell Biol.*, **181**, 921–934.

Estrella, L.A., Wilkinson, M.F., & González, C.I. (2009) The Shuttling Protein Npl3 Promotes Translation Termination Accuracy in Saccharomyces cerevisiae. *J. Mol. Biol.*, **394**, 410–422.

Ezponda, T., Pajares, M.J., Agorreta, J., Echeveste, J.I., López-Picazo, J.M., Torre, W., Pio, R., & Montuenga, L.M. (2010) The oncoprotein SF2/ASF promotes non-small cell lung cancer survival by enhancing survivin expression. *Clin. Cancer Res.*, **16**, 4113–4125.

Fantl, V., Stamp, G., Andrews, A., Rosewell, I., & Dickson, C. (1995) Mice lacking cyclin D1 are small and show defects in eye and mammary gland development. *Genes Dev.*, **9**, 2364–2372.

Ferguson, A.M., White, L.S., Donovan, P.J., Piwnica-worms, H., Al, F.E.T., & Iol, M.O.L.C.E.L.L.B. (2005) Normal Cell Cycle and Checkpoint Responses in Mice and Cells Lacking Cdc25B and Cdc25C Protein Phosphatases. *Mol. Cell. Biol.*, **25**, 2853–2860.

Fica, S.M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., & Piccirilli, J.A. (2013) RNA catalyses nuclear pre-mRNA splicing. *Nature*, **503**, 229–234.

Fonseca, B.D., Zakaria, C., Jia, J.J., Graber, T.E., Svitkin, Y., Tahmasebi, S., Healy, D., Hoang, H.D., Jensen, J.M., Diao, I.T., Lussier, A., Dajadian, C., Padmanabhan, N., Wang, W., Matta-Camacho, E., Hearnden, J., Smith, E.M., Tsukumo, Y., Yanagiya, A., Morita, M., Petroulakis, E., González, J.L., Hernández, G., Alain, T., & Damgaard, C.K. (2015) La-related protein 1 (LARP1) represses terminal oligopyrimidine (TOP) mRNA translation downstream of mTOR complex 1 (mTORC1). *J. Biol. Chem.*, **290**, 15996–16020.

Fraser, H.B. & Xie, X. (2009) Common polymorphic transcript variation in human disease Common polymorphic transcript variation in human disease. *Genome Res*, **19**, 567–575.

Fregoso, O.I., Das, S., Akerman, M., & Krainer, A.R. (2013) Splicing-Factor Oncoprotein SRSF1 Stabilizes p53 via RPL5 and Induces Cellular Senescence. *Mol. Cell*, **50**, 56–66.

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

Fu, Y., Huang, B., Shi, Z., Han, J., Wang, Y., Huangfu, J., & Wu, W. (2013) SRSF1 and SRSF9 RNA binding proteins promote Wnt signalling-mediated tumorigenesis by enhancing β-catenin biosynthesis. *EMBO Mol. Med.*, **5**, 737–750.

Gama-Carvalho, M. & Carmo-Fonseca, M. (2006) Heterokaryons: An assay for nucleocytoplasmic shuttling. *Cell Biol. Four-Volume Set*, **2**, 277–283.

Gangloff, Y., Mueller, M., Dann, S.G., Svoboda, P., Sticker, M., Spetz, J., Um, H., Brown, E.J., Cereghini, S., Thomas, G., Kozma, S.C., & Um, S.H. (2004) Disruption of the Mouse mTOR Gene Leads to Early Postimplantation Lethality and Prohibits Embryonic Stem Cell Development. *Mol. Cell. Biol.*, **24**, 9508–9516.

Ge, H., Zuo, P., & Manley, J.L. (1991) Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. *Cell*, **66**, 373–382.

Gebauer, F. & Hentze, M.W. (2004) Molecular mechanisms of translational control. *Nat.*

Rev. Mol. Cell Biol., **5**, 827–835.

Gehman, L.T., Meera, P., Stoilov, P., Shiue, L., O'Brien, J.E., Meisler, M.H., Ares, M., Otis, T.S., & Black, D.L. (2012) The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. *Genes Dev.*, **26**, 445–460.

Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P.M., Green, M.R., Riva, S., & Biamonti, G. (2005) Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Mol. Cell*, **20**, 881–890.

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

Gilbert, W. (1978) Why genes in pieces? *Nature*,.

Goldberg, S., Weber, J., & Darnell, J.E. (1977) The definition of a large viral transcription unit late in Ad2 infection of HeLa cells: Mapping by effects of ultraviolet irradiation. *Cell*, **10**, 617–621.

Gong, C., Kim, Y.K., Woeller, C.F., Tang, Y., & Maquat, L.E. (2009) SMD and NMD are competitive pathways that contribute to myogenesis: Effects on PAX3 and myogenin mRNAs. *Genes Dev.*, **23**, 54–66.

Goren, A., Ram, O., Amit, M., Keren, H., Lev-Maor, G., Vig, I., Pupko, T., & Ast, G. (2006) Comparative Analysis Identifies Exonic Splicing Regulatory Sequences-The Complex Definition of Enhancers and Silencers. *Mol. Cell*, **22**, 769–781.

Grabowski, P.J., Seiler, S.R., & Sharp, P.A. (1985) A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell*, **42**, 345–353.

Graham, D.B. & Root, D.E. (2015) Resources for the design of CRISPR gene editing experiments. *Genome Biol.*, **16**, 260.

Gueroussov, S., Gonatopoulos-Pournatzis, T., Irimia, M., Bushra, R., Lin, Z.-Y., Gingras, A.-C., & Blencowe, B.J. (2015) An alternative splicing event amplifies evolutionary differences between vertebrates. *Science (80-.).*, **349**, 868–873.

Guertin, D.A., Stevens, D.M., Thoreen, C.C., Burds, A.A., Kalaany, N.Y., Moffat, J., Brown, M., Fitzgerald, K.J., & Sabatini, D.M. (2006) Ablation in Mice of the mTORC Components raptor, rictor, or mLST8 Reveals that mTORC2 Is Required for Signaling to Akt-FOXO and PKCα, but Not S6K1. *Dev. Cell*, **11**, 859–871.

Hamelberg, D., Shen, T., & McCammon, J.A. (2007) A proposed signaling motif for nuclear import in mRNA processing via the formation of arginine claw. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 14947–14951.

Hanamura, a, Cáceres, J.F., Mayeda, a, Franza, B.R., & Krainer, a R. (1998) Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA*, **4**, 430–444.

Hang, J., Wan, R., Yan, C., & Shi, Y. (2015) Structural basis of pre-mRNA splicing. *Sci. 327, 580*, **349**, 1191–1198.

Hargous, Y., Hautbergue, G.M., Tintaru, A.M., Skrisovska, L., Golovanov, A.P., Stevenin, J., Lian, L.-Y., Wilson, S.A., & Allain, F.H.-T. (2006) Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J.*, **25**, 5126–5137.

Hautbergue, G.M., Castelli, L.M., Ferraiuolo, L., Sanchez-Martinez, A., Cooper-Knock, J., Higginbottom, A., Lin, Y.-H., Bauer, C.S., Dodd, J.E., Myszczynska, M.A., Alam, S.M., Garneret, P., Chandran, J.S., Karyka, E., Stopford, M.J., Smith, E.F., Kirby, J., Meyer, K., Kaspar, B.K., Isaacs, A.M., El-Khamisy, S.F., De Vos, K.J., Ning, K., Azzouz, M., Whitworth, A.J., & Shaw, P.J. (2017) SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. *Nat. Commun.*, **8**, 16063.

Hay, N. & Sonenberg, N. (2004) Upstream and Downstream of mTOR. *Genes Dev.*, **18**, 1926–1945.

Haynes, C. & Iakoucheva, L.M. (2006) Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res.*, **34**, 305–312.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., & Stelzl, U. (2012) Dynamic Protein-Protein Interaction Wiring of the Human Spliceosome. *Mol. Cell*, **45**, 567–580.

Hendel, A., Fine, E.J., Bao, G., & Porteus, M.H. (2015) Quantifying on- and off-target genome editing. *Trends Biotechnol.*, **33**, 132–140.

Heyer, E.E. & Moore, M.J. (2016) Article Redefining the Translational Status of 80S Monosomes Authors. *Cell*, **164**, 757–769.

Hinnebusch, A.G. (2017) Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. *Trends Biochem. Sci.*, **xx**, 1–23.

Howard, J.M. & Sanford, J.R. (2014) The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdiscip. Rev. RNA*,.

Hsu, P.D., Scott, D. a, Weinstein, J. a, Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L. a, Bao, G., & Zhang, F. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.*, **31**, 827–832.

Huang, Y., Gattoni, R., Stévenin, J., & Steitz, J. a. (2003) SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol. Cell*, **11**, 837–843.

Huang, Y. & Steitz, J.A. (2001) Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol. Cell*, **7**, 899–905.

Hug, N., Longman, D., & Cáceres, J.F. (2015) Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.*, **44**, 1483–1495.

Hulse, R.P., Beazley-Long, N., Hua, J., Kennedy, H., Prager, J., Bevan, H., Qiu, Y., Fernandes, E.S., Gammons, M. V., Ballmer-Hofer, K., Gittenberger de Groot, A.C., Churchill, A.J., Harper, S.J., Brain, S.D., Bates, D.O., & Donaldson, L.F. (2014) Regulation of alternative VEGF-A mRNA splicing is a therapeutic target for analgesia. *Neurobiol. Dis.*, **71**, 245–259.

Hulse, R.P., Drake, R.A.R., Bates, D.O., & Donaldson, L.F. (2016) The control of alternative splicing by SRSF1 in myelinated afferents contributes to the development of neuropathic pain. *Neurobiol. Dis.*, **96**, 186–200.

Ingolia, N.T. (2014) Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.*, **15**, 205–213.

Ingolia, N.T. (2016) Ribosome Footprint Profiling of Translation throughout the Genome. *Cell*, **165**, 22–33.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., & Weissman, J.S. (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.*, **7**, 1534–1550.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., & Weissman, J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.

Ingolia, N.T., Lareau, L.F., & Weissman, J.S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.

Irimia, M. & Roy, S.W. (2014) Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb. Perspect. Biol.*, **6**.

Jantzen, S.G., Sutherland, B.J., Minkley, D.R., & Koop, B.F. (2011) GO Trimming: Systematically reducing redundancy in large Gene Ontology datasets. *BMC Res. Notes*, **4**, 267.

Jensen, M. a, Wilkinson, J.E., & Krainer, A.R. (2014) Splicing factor SRSF6 promotes hyperplasia of sensitized skin. *Nat. Struct. Mol. Biol.*, **21**, 189–197.

Jiang, L., Huang, J., Higgs, B.W., Hu, Z., Xiao, Z., Yao, X., Conley, S., Zhong, H., Liu, Z., Brohawn, P., Shen, D., Wu, S., Ge, X., Jiang, Y., Zhao, Y., Lou, Y., Morehouse, C., Zhu, W., Sebastian, Y., Czapiga, M., Oganesyan, V., Fu, H., Niu, Y., Zhang, W., Streicher, K., Tice, D., Zhao, H., Zhu, M., Xu, L., Herbst, R., Su, X., Gu, Y., Li, S., Huang, L., Gu, J., Han, B., Jallal, B., Shen, H., & Yao, Y. (2016) Genomic Landscape Survey Identifies SRSF1 as a Key Oncodriver in Small Cell Lung Cancer. *PLOS Genet.*, **12**, e1005895.

Jiang, Q., Crews, L.A., Holm, F., & Jamieson, C.H.M. (2017) RNA editing-dependent epitranscriptome diversity in cancer stem cells. *Nat. Rev. Cancer*, **17**, 381–392.

Jumaa, H. & Nielsen, P.J. (1997) The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.*, **16**, 5077–5085.

Jumaa, H., Wei, G., & Nielsen, P.J. (1999) Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20. *Curr. Biol.*, **9**, 899–902.

Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., & Krainer, A.R. (2007) The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.*, **14**, 185–193.

Karni, R., Hippo, Y., Lowe, S.W., & Krainer, A.R. (2008) The splicing-factor oncoprotein SF2/ASF activates mTORC1. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 15323–15327.

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

Katsu, R., Onogi, H., Wada, K., Kawaguchi, Y., & Hagiwara, M. (2002) Novel SR-rich-related protein Clasp specifically interacts with inactivated Clk4 and induces the exon EB inclusion of Clk. *J. Biol. Chem.*, **277**, 44220–44228.

Keller, R. (2002) Shaping the vertebrate body plan by polarized embryonic cell movements.

Science (80-.)., **298**, 1950–1954.

Keshwani, M.M., Aubol, B.E., Fattet, L., Ma, C., Qiu, J., Jennings, P.A., Fu, X., & Adams, J.A. (2016) Protein Conformation and Splicing Function **466**, 311–322.

Kim, H.-J., Lee, H.-R., Seo, J.-Y., Ryu, H.G., Lee, K.-H., Kim, D.-Y., & Kim, K.-T. (2017) Heterogeneous nuclear ribonucleoprotein A1 regulates rhythmic synthesis of mouse Nfil3 protein via IRES-mediated translation. *Sci. Rep.*, **7**, 42882.

Kim, J., Park, R.Y., Chen, J.-K., Jeong, S., & Ohn, T. (2013) Splicing factor SRSF3 represses the translation of programmed cell death 4 mRNA by associating with the 5'-UTR region. *Cell Death Differ.*, **21**, 1–10.

Klessig, D.F. (1977) Two adenovirus mRNAs have a common 5' terminal leader sequence encoded at least 10 kb upstream from their main coding regions. *Cell*, **12**, 9–21.

Ko, T.K., Kelly, E., & Pines, J. (2001) CrkRS : a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *J. Cell Biol.*, **114**.

Komor, A.C., Badran, A.H., & Liu, D.R. (2016) CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell*, **168**, 1–17.

Kornblihtt, A.R., Schor, I.E., Alló, M., Dujardin, G., Petrillo, E., & Muñoz, M.J. (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.*, **14**, 153–165.

Krainer, A.R., Conway, G.C., & Kozak, D. (1990) Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev.*, **4**, 1158–1171.

Krainer, a R. & Maniatis, T. (1985) Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell*, **42**, 725–736.

Krainer, a R., Mayeda, a, Kozak, D., & Binns, G. (1991) Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and Drosophila splicing regulators. *Cell*, **66**, 383–394.

Krebs, D.L., Metcalf, D., Merson, T.D., Voss, A.K., Thomas, T., Zhang, J.-G., Rakar, S., O'bryan, M.K., Willson, T. a, Viney, E.M., Mielke, L. a, Nicola, N. a, Hilton, D.J., & Alexander, W.S. (2004) Development of hydrocephalus in mice lacking SOCS7. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 15446–15451.

Kress, T.L., Krogan, N.J., & Guthrie, C. (2008) A Single SR-like Protein, Npl3, Promotes Pre-mRNA Splicing in Budding Yeast. *Mol. Cell*, **32**, 727–734.

Kretz, M. (2013) TINCR, staufen1, and cellular differentiation. *RNA Biol.*, **10**, 1597–1601.

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

Lai, M.C. & Tarn, W.Y. (2004) Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. *J. Biol. Chem.*, **279**, 31745–31749.

Lareau, L.F. & Brenner, S.E. (2015) Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Mol. Biol. Evol.*, **32**, 1072–1079.

Lareau, L.F., Hite, D.H., Hogan, G.J., & Brown, P.O. (2014) Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife*, **2014**, 1–16.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., & Brenner, S.E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, **446**, 926–929.

Larsen, S.C., Sylvestersen, K.B., Mund, A., Lyon, D., Mullari, M., Madsen, M. V, Daniel, J.A., Jensen, L.J., & Nielsen, M.L. (2016) Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells. *Sci. Signal.*, **9**, 1–15.

Le Hir, H., Gatfield, D., Izaurralde, E., & Moore, M.J. (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.*, **20**, 4987–4997.

Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., & Qian, S.-B. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci.*, **109**, E2424–E2432.

Lee, S. & Stevens, S.W. (2016) Spliceosomal intronogenesis. *Proc. Natl. Acad. Sci.*, **113**, 6514–6519.

Lee, Y. & Rio, D.C. (2014) Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.*, **84**, 291–323.

Lemaire, R., Prasad, J., Kashima, T., Gustafson, J., Manley, J.L., & Lafyatis, R. (2002) Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: A novel function for SR proteins. *Genes Dev.*, **16**, 594–607.

Lewis, K.M. & Ke, A. (2017) Building the Class 2 CRISPR-Cas Arsenal. *Mol. Cell*, **65**, 377–379.

Li, M., Pokharel, S., Wang, J.-T., Xu, X., & Liu, Y. (2015) RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. *Nat. Commun.*, **6**, 6720.

Li, X. & Manley, J.L. (2005) Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell*, **122**, 365–378.

Li, X., Wang, J., & Manley, J.L. (2005) Loss of splicing factor ASF/SF2 induces G2 cell cycle arrest and apoptosis, but inhibits internucleosomal DNA fragmentation. *Genes Dev.*, **19**, 2705–2714.

Lieber, M.R. (2011) The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Mol. Microbiol.*, **79**, 181–211.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., & Fu, X.-D. (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.*, **15**, 819–826.

Lin, S., Staahl, B., Alla, R.K., & Doudna, J.A. (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR / Cas9 delivery. *Elife*,.

Lin, S., Xiao, R., Sun, P., Xu, X., & Fu, X.D. (2005) Dephosphorylation-dependent sorting of SR splicing factors during mRNP maturation. *Mol. Cell*, **20**, 413–425.

Lloyd, A.C. (2013) The regulation of cell size. *Cell*, **154**, 1194–1205.

Long, J.C. & Cáceres, J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.*, **417**, 15–27.

Longman, D., Johnstone, I.L., & Caceres, J.F. (2000) Functional characterization of SR and SR-related genes in Caenorhabditis elegans. *EMBO J.*, **19**, 1625–1637.

Loomis, R.J., Naoe, Y., Parker, J.B., Savic, V., Bozovsky, M.R., Macfarlan, T., Manley, J.L., & Chakravarti, D. (2009) Chromatin Binding of SRp20 and ASF/SF2 and Dissociation from Mitotic Chromosomes Is Modulated by Histone H3 Serine 10 Phosphorylation. *Mol. Cell*, **33**, 450–461.

Luo, G., Santoro, I.M., Mcdaniel, L.D., Nishijima, I., Mills, M., Vogel, H., Schultz, R.A., & Bradley, A. (2000) Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nat. Genet.*, **26**.

Ma, C.T., Ghosh, G., Fu, X.D., & Adams, J. a. (2010) Mechanism of Dephosphorylation of the SR Protein ASF/SF2 by Protein Phosphatase 1. *J. Mol. Biol.*, **403**, 386–404.

Manley, J.L. & Krainer, A.R. (2010) A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev.*, **24**, 1073–1074.

Manning, K.S. & Cooper, T.A. (2016) The roles of RNA processing in translating genotype to phenotype. *Nat. Rev. Mol. Cell Biol.*, **18**, 102–114.

Maris, C., Dominguez, C., & Allain, F.H.T. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, **272**, 2118–2131.

Martello, G. & Smith, A. (2014) The nature of embryonic stem cells. *Annu. Rev. Cell Dev. Biol.*, **30**, 647–675.

Martínez-Lumbreras, S., Taverniti, V., Zorrilla, S., Séraphin, B., & Pérez-Cañadillas, J.M. (2016) Gbp2 interacts with THO/TREX through a novel type of RRM domain. *Nucleic Acids Res.*, **44**, 437–448.

Maslon, M.M., Heras, S.R., Bellora, N., Eyras, E., & Cáceres, J.F. (2014) The translational landscape of the splicing factor SRSF1 and its role in mitosis. *Elife*, **2014**.

Matera, A.G. & Wang, Z. (2014) A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.*, **15**, 108–121.

Matlin, A.J., Clark, F., & Smith, C.W.J. (2005) Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.*, **6**, 386–398.

Michlewski, G., Sanford, J.R., & Cáceres, J.F. (2008) The Splicing Factor SF2/ASF Regulates Translation Initiation by Enhancing Phosphorylation of 4E-BP1. *Mol. Cell*, **30**, 179–189.

Misteli, T., Cáceres, J.F., & Spector, D.L. (1997) The dynamics of a pre-mRNA splicing factor in living cells. *Nature*, **387**, 523–527.

Mitsunobu, H., Teramoto, J., Nishida, K., & Kondo, A. (2017) Beyond Native Cas9: Manipulating Genomic Information and Function. *Trends Biotechnol.*, 1–14.

Moore, M.J., Wang, Q., Kennedy, C.J., & Silver, P. a. (2010) An alternative splicing network

links cell-cycle control to apoptosis. *Cell*, **142**, 625–636.

Möröy, T. & Heyd, F. (2007) The impact of alternative splicing in vivo: mouse models show the way. *RNA*, **13**, 1155–1171.

Moulton, V.R., Gillooly, A.R., Perl, M. a., Markopoulou, A., & Tsokos, G.C. (2015) Serine Arginine-Rich Splicing Factor 1 (SRSF1) Contributes to the Transcriptional Activation of CD3ζ in Human T Cells. *PLoS One*, **10**, e0131073.

Moulton, V.R., Gillooly, A.R., & Tsokos, G.C. (2014) Ubiquitination regulates expression of the serine/arginineric splicing factor 1 (SRSF1) in normal and systemic lupus erythematosus (SLE) t cells. *J. Biol. Chem.*, **289**, 4126–4134.

Muddukrishna, B., Jackson, C.A., & Yu, M.C. (2017) Protein Arginine Methylation of Npl3 Promotes Splicing of the SUS1 Intron harboring Non-consensus 5' Splice Site and Branch Site. *Biochim. Biophys. Acta - Gene Regul. Mech.*,.

Muller-McNicoll, M., Botti, V., de Jesus Domingues, A.M., Brandl, H., Schwich, O.D., Steiner, M.C., Curk, T., Poser, I., Zarnack, K., & Neugebauer, K.M. (2016) SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev.*, **30**, 553–566.

Murakami, M. & Ichisaka, T. (2004) mTOR is essential for growth and proliferation in early mouse embryos and embryonic stem cells. *Mol. Cell. Biol.*, **24**, 6710–6718.

Narita, K. & Takeda, S. (2015) Cilia in the choroid plexus: their roles in hydrocephalus and beyond. *Front. Cell. Neurosci.*, **9**, 1–5.

Neumann, B., Walter, T., Hériché, J.-K., Bulkescher, J., Erfle, H., Conrad, C., Rogers, P., Poser, I., Held, M., Liebel, U., Cetin, C., Sieckmann, F., Pau, G., Kabbe, R., Wünsche, A., Satagopam, V., Schmitz, M.H.A., Chapuis, C., Gerlich, D.W., Schneider, R., Eils, R., Huber, W., Peters, J.-M., Hyman, A.A., Durbin, R., Pepperkok, R., & Ellenberg, J. (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, **464**, 721–727.

Ngo, J.C.K., Chakrabarti, S., Ding, J.H., Velazquez-Dones, A., Nolen, B., Aubol, B.E., Adams, J. a., Fu, X.D., & Ghosh, G. (2005) Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2. *Mol. Cell*, **20**, 77–89.

Ngo, J.C.K., Giang, K., Chakrabarti, S., Ma, C.T., Huynh, N., Hagopian, J.C., Dorrestein, P.C., Fu, X.D., Adams, J.A., & Ghosh, G. (2008) A Sliding Docking Interaction Is Essential for Sequential and Processive Phosphorylation of an SR Protein by SRPK1. *Mol. Cell*, **29**, 563–576.

Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E., & Ares, M. (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.*, **21**, 708–718.

Ohanna, M., Sobering, A.K., Lapointe, T., Lorenzo, L., Praud, C., Sonenberg, N., Kelly, P.A., Sotiropoulos, A., & Pende, M. (2005) Atrophy of S6K1 − / − skeletal muscle cells reveals distinct mTOR effectors for cell cycle and size control. *Nat. Cell Biol.*, **7**.

Orford, K.W. & Scadden, D.T. (2008) Deconstructing stem cell self-renewal: genetic insights

into cell-cycle regulation. *Nat. Rev. Genet.*, **9**, 115–128.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., & Blencowe, B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–1415.

Pan, W., Tsai, H., Wang, S., Hsiao, M., Wu, P., & Tsai, M. (2015) The RNA recognition motif of NIFK is required for rRNA maturation during cell cycle progression. *RNA Biol.*, **12**, 255–267.

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

Papasaikas, P. & Valcárcel, J. (2016) The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem. Sci.*, **41**, 33–45.

Park, S.K. & Jeong, S. (2016) SRSF3 represses the expression of PDCD4 protein by coordinated regulation of alternative splicing, export and translation. *Biochem. Biophys. Res. Commun.*, **470**, 431–438.

Patel, A.A. & Steitz, J.A. (2003) Splicing double: insights from the second spliceosome. *Nat. Rev. Mol. Cell Biol.*, **4**, 960–970.

Pauklin, S. & Vallier, L. (2013) The cell cycle state of stem cells determines cell fate propensity. *Cell*, **155**, 135–147.

Pelisch, F., Gerez, J., Druker, J., Schor, I.E., Muñoz, M.J., Risso, G., Petrillo, E., Westman, B.J., Lamond, A.I., Arzt, E., & Srebrow, A. (2010) The serine/arginine-rich protein SF2/ASF regulates protein sumoylation. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 16119–16124.

Pende, M., Um, S.H., Mieulet, V., Goss, V.L., Mestan, J., Mueller, M., Fumagalli, S., Kozma, S.C., Thomas, G., & Sticker, M. (2004) S6K1-/-/S6K2-/- Mice Exibit Preinatal Lethality and Rapamycin-Sensitive 5 ′ -Terminal Oligopyrimidine mRNA Translation and Reveal a Mitogen-Activated Protein Kinase-Dependent S6 Kinase Pathway. *Mol. Cell. Biol.*, **24**, 3112–3124.

Piñol-Roma, S. & Dreyfuss, G. (1992) Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature*, **355**, 730–732.

Plaschka, C., Lin, P.-C., & Nagai, K. (2017) Structure of a pre-catalytic spliceosome. *Nature*, 1–27.

Quesnel-Vallières, M., Dargaei, Z., Irimia, M., Gonatopoulos-Pournatzis, T., Ip, J.Y., Wu, M., Sterne-Weiler, T., Nakagawa, S., Woodin, M.A., Blencowe, B.J., & Cordes, S.P. (2016) Misregulation of an Activity-Dependent Splicing Network as a Common Mechanism Underlying Autism Spectrum Disorders. *Mol. Cell*, **64**, 1023–1034.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D. a., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. (2013) Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380–1389.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D. a, & Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu,

M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L.H., Dale, R.K., Smith, S.A., Yarosh, C.A., Kelly, S.M., Nabet, B., Mecenas, D., Li, W., Laishram, R.S., Qiao, M., Lipshitz, H.D., Piano, F., Corbett, A.H., Carstens, R.P., Frey, B.J., Anderson, R.A., Lynch, K.W., Penalva, L.O.F., Lei, E.P., Fraser, A.G., Blencowe, B.J., Morris, Q.D., & Hughes, T.R. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.

Reed, D.R., Lawler, M.P., & Tordoff, M.G. (2008) Reduced body weight is a common effect of gene knockout in mice. *BMC Genet.*, **9**, 4.

Richter, J.D. & Coller, J. (2015) Pausing on Polyribosomes: Make Way for Elongation in Translational Control. *Cell*, **163**, 292–300.

Risso, G., Pelisch, F., Quaglino, A., Pozzi, B., & Srebrow, A. (2012) Critical Review Regulating the Regulators : Serine / Arginine-rich Proteins Under Scrutiny. *IUBMB Life*, **64**, 809–816.

Rogozin, I.B., Carmel, L., Csuros, M., & Koonin, E. V (2012) Origin and evolution of spliceosomal introns. *Biol. Direct*, **7**, 11.

Roscigno, R.F. & Garcia-Blanco, M.A. (1995) SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA*, **1**, 692–706.

Roth, M.B., Murphy, C., & Gall, J.G. (1990) A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *J. Cell Biol.*, **111**, 2217–2223.

Roundtree, I.A., Evans, M.E., Pan, T., & Chuan He (2017) Dynamic RNA Modifications in Gene Expression Regulation. *Cell*, **169**, 1187.

Roy, R., Durie, D., Hui, L., Qian, L.B., Skehel, J.M., Mauri, F., Veronica, C.L., Mattia, B., Guo, L., Holcik, M., Seckl, M.J., & Pardo, O.E. (2014) hnRNPA1 couples nuclear export and translation of specific mRNAs downstream of FGF-2 / S6K2 signalling. *Nucleic Acids Res.*, **42**, 1–15.

Saitoh, N., Spahr, C.S., Patterson, S.D., Bubulya, P., Neuwald, A.F., & Spector*, and D.L. (2004) Proteomic Analysis of Interchromatin Granule Clusters. *Mol. Biol. Cell*, **16**, 1–13.

Sakakibara, S., Nakamura, Y., Yoshida, T., Shibata, S., Koike, M., Takano, H., Ueda, S., Uchiyama, Y., Noda, T., & Okano, H. (2002) RNA-binding protein Musashi family: roles for CNS stem cells and a subpopulation of ependymal cells revealed by targeted disruption and antisense ablation. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 15194–15199.

Sampath, P., Pritchard, D.K., Pabon, L., Reinecke, H., Schwartz, S.M., Morris, D.R., & Murry, C.E. (2008) A Hierarchical Network Controls Protein Translation during Murine Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell*, **2**, 448–460.

Sander, J.D. & Joung, J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.*, **32**, 347–355.

Sanford, J.R. & Bruzik, J.P. (2001) Regulation of SR protein localization during development. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 10184–10189.

Sanford, J.R., Coutinho, P., Hackett, J. a., Wang, X., Ranahan, W., & Cáceres, J.F. (2008)

Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS One*, **3**.

Sanford, J.R., Ellis, J.D., Cazalla, D., & Cáceres, J.F. (2005) Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 15042–15047.

Sanford, J.R., Gray, N.K., Beckmann, K., & Cáceres, J.F. (2004) A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.*, **18**, 755–768.

Sanford, J.R., Wang, X., Mort, M., Vanduyn, N., Cooper, D.N., Mooney, S.D., Edenberg, H.J., & Liu, Y. (2009) Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.*, **19**, 381–394.

Sapra, A.K., Änkö, M.L., Grishina, I., Lorenz, M., Pabis, M., Poser, I., Rollins, J., Weiland, E.M., & Neugebauer, K.M. (2009) SR Protein Family Members Display Diverse Activities in the Formation of Nascent and Mature mRNPs In Vivo. *Mol. Cell*, **34**, 179–190.

Saulière, J., Murigneux, V., Wang, Z., Marquenet, E., Barbosa, I., Le Tonquèze, O., Audic, Y., Paillard, L., Crollius, H.R., & Le Hir, H. (2012) CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat. Struct. Mol. Biol.*, **19**, 1124–1131.

Saxton, R.A. & Sabatini, D. (2012) mTOR Signaling in Growth Control and Disease. *Cell*, **149**, 274–293.

Schmidt, C., Beilsten-Edmands, V., & Robinson, C. V. (2016) Insights into Eukaryotic Translation Initiation from Mass Spectrometry of Macromolecular Protein Assemblies. *J. Mol. Biol.*, **428**, 344–356.

Schutz, P., Bumann, M., Oberholzer, A.E., Bieniossek, C., Trachsel, H., Altmann, M., & Baumann, U. (2008) Crystal structure of the yeast eIF4A-eIF4G complex : An RNA-helicase controlled by protein – protein interactions. *Proc. Natl. Acad. Sci.*, **105**.

Schwartz, S. & Ast, G. (2010) Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. *EMBO J.*, **29**, 1629–1636.

Schwarz, D.S., Ding, H., Kennington, L., Moore, J.T., Schelter, J., Burchard, J., Linsley, P.S., Aronin, N., Xu, Z., & Zamore, P.D. (2006) Designing siRNA that distinguish between genes that differ by a single nucleotide. *PLOS Genet.*, **2**.

Sen, S., Jumaa, H., & Webster, N.J.G. (2013) Splicing factor SRSF3 is crucial for hepatocyte differentiation and metabolic function. *Nat. Commun.*, **4**, 1336.

Serrano, P., Aubol, B.E., Keshwani, M.M., Forli, S., Ma, C.-T., Dutta, S.K., Geralt, M., Wüthrich, K., & Adams, J.A. (2016) Directional phosphorylation and nuclear transport of the splicing factor SRSF1 is regulated by an RNA recognition motif. *J. Mol. Biol.*,.

Shalem, O., Sanjana, N.E., & Zhang, F. (2015) High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.*, **16**, 299–311.

Sharp, P.A. (1985) On the Origin of RNA Splicing and Introns. *Cell*, **42**, 397–400.

Sharp, P.A., Konarska, M.M., Grabowski, P.J., Lamond, A.I., Marcinlak, R., & Seiler, S.R. (1987) Splicing of nuclear messenger RNA precursors. *Cold Spring Harb. Symp. Quant.*

Biol., **52**, 277–285.

Shen, H. & Green, M.R. (2004) A pathway of sequential Arginine-Serine-Rich Domain-Splicing signal interactions during mammalian spliceosome assembly. *Mol. Cell*, **16**, 363–373.

Shen, H. & Green, M.R. (2007) RS domain–splicing signal interactions in splicing of U12-type and U2-type introns. *Nat. Struct. Mol. Biol.*, **14**, 597–603.

Shibasaki, T., Tokunaga, A., Sakamoto, R., Sagara, H., Noguchi, S., Sasaoka, T., & Yoshida, N. (2013) PTB deficiency causes the loss of adherens junctions in the dorsal telencephalon and leads to lethal hydrocephalus. *Cereb. Cortex*, **23**, 1824–1835.

Shimoni-Sebag, A., Lebenthal-Loinger, I., Zender, L., & Karni, R. (2013) RRM1 domain of the splicing oncoprotein SRSF1 is required for MEK1-MAPK-ERK activation and cellular transformation. *Carcinogenesis*, **34**, 2498–2504.

Shin, C., Kleiman, F.E., & Manley, J.L. (2005) Multiple properties of the splicing repressor SRp38 distinguish it from typical SR proteins. *Mol. Cell. Biol.*, **25**, 8334–8343.

Shioi, T., Kang, P.M., Douglas, P.S., Hampe, J., Yballe, C.M., Lawitts, J., Cantley, L.C., & Izumo, S. (2000) The conserved phosphoinositide 3-kinase pathway determines heart size in mice. *EMBO J.*, **19**, 2537–2548.

Shirokikh, N.E., Archer, S.K., Beilharz, T.H., Powell, D., & Preiss, T. (2017) Translation complex profile sequencing to study the in vivo dynamics of mRNA–ribosome interactions during translation initiation, elongation and termination. *Nat. Protoc.*, **12**, 697–731.

Shuda, M., Velásquez, C., Cheng, E., Cordek, D.G., Kwun, H.J., Chang, Y., & Moore, P.S. (2015) CDK1 substitutes for mTOR kinase to activate mitotic cap-dependent protein translation. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 5875–5882.

Sibley, C.R., Blazquez, L., & Ule, J. (2016) Lessons from non-canonical splicing. *Nat. Rev. Genet.*, **17**, 407–421.

Simsek, D., Tiu, G.C., Flynn, R.A., Xu, A.F., Chang, H.Y., Barna, M., Simsek, D., Tiu, G.C., Flynn, R.A., Byeon, G.W., Leppek, K., Xu, A.F., & Chang, H.Y. (2017) The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity Article The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell*, **169**, 1051–1057.e18.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S. a., Weng, Z., & Moore, M.J. (2012) The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell*, **151**, 750–764.

Singh, G., Pratt, G., Yeo, G.W., & Moore, M.J. (2013) The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu. Rev. Biochem.*, **84**, 150317182619002.

Sinha, R., Allemand, E., Zhang, Z., Karni, R., Myers, M.P., & Krainer, A.R. (2010) Arginine methylation controls the subcellular localization and functions of the oncoprotein splicing factor SF2/ASF. *Mol. Cell. Biol.*, **30**, 2762–2774.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., & Zhang, F. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science (80-.).*, **351**, 84–88.

Sonenberg, N. & Hinnebusch, A.G. (2009) Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell*, **136**, 731–745.

Soret, J., Gattoni, R., Guyon, C., Sureau, a, Popielarz, M., Le Rouzic, E., Dumon, S., Apiou, F., Dutrillaux, B., Voss, H., Ansorge, W., Stévenin, J., & Perbal, B. (1998) Characterization of SRp46, a novel human SR splicing factor encoded by a PR264/SC35 retropseudogene. *Mol. Cell. Biol.*, **18**, 4924–4934.

Sotak, B.N. & Gleeson, J.G. (2012) Can't get there from here: cilia and hydrocephalus. *Nat. Med.*, **18**, 1742–1743.

Spassky, N. & Meunier, A. (2017) The development and functions of multiciliated epithelia. *Nat. Rev. Mol. Cell Biol.*, **18**, 423–436.

Sterne-Weiler, T., Martinez-Nunez, R.T., Howard, J.M., Cvitovik, I., Katzman, S., Tariq, M. a., Pourmand, N., & Sanford, J.R. (2013) Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res.*, **23**, 1615–1623.

Stickeler, E., Kittrell, F., Medina, D., & Berget, S.M. (1999) Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. *Oncogene*, **18**, 3574–3582.

Stumpf, C., Moreno, M., Olshen, A., Taylor, B., & Ruggero, D. (2013) The translational landscape of the mammalian cell cycle. *Mol. Cell*, **52**, 574–582.

Sugiyama, H., Takahashi, K., Yamamoto, T., Iwasaki, M., Narita, M., Nakamura, M., Rand, T.A., Nakagawa, M., Watanabe, A., & Yamanaka, S. (2016) *Nat1* promotes translation of specific proteins that induce differentiation of mouse embryonic stem cells. *Proc. Natl. Acad. Sci.*, **114**, 201617234.

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

Susor, A., Jansova, D., Cerna, R., Danylevska, A., Anger, M., Toralova, T., Malik, R., Supolikova, J., Cook, M.S., Oh, J.S., & Kubelka, M. (2015) Temporal and spatial regulation of translation in the mammalian oocyte via the mTOR-eIF4F pathway. *Nat. Commun.*, **6**, 1–12.

Tahmasebi, S., Jafarnejad, S.M., Tam, I.S., Gonatopoulos-Pournatzis, T., Matta-Camacho, E., Tsukumo, Y., Yanagiya, A., Li, W., Atlasi, Y., Caron, M., Braunschweig, U., Pearl, D., Khoutorsky, A., Gkogkas, C.G., Nadon, R., Bourque, G., Yang, X.-J., Tian, B., Stunnenberg, H.G., Yamanaka, Y., Blencowe, B.J., Giguère, V., & Sonenberg, N. (2016) Control of embryonic stem cell self-renewal and differentiation via coordinated alternative splicing and translation of YY2. *Proc. Natl. Acad. Sci.*, **113**, 12360–12367.

Tanenbaum, M.E., Stern-Ginossar, N., Weissman, J.S., & Vale, R.D. (2015) Regulation of mRNA translation during mitosis. *Elife*, **4**, 1–19.

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., & Sabatini, D.M. (2012) A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature*, **485**, 109–113.

Tintaru, A.M., Hautbergue, G.M., Hounslow, A.M., Hung, M.-L., Lian, L.-Y., Craven, C.J., & Wilson, S.A. (2007) Structural and functional analysis of RNA and TAP binding to

SF2/ASF. *EMBO Rep.*, **8**, 756–762.

Tress, M.L., Abascal, F., & Valencia, A. (2017) Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends Biochem. Sci.*, **42**, 98–110.

Tsouroula, K., Furst, A., Rogier, M., Heyer, V., Maglott-Roth, A., Ferrand, A., Reina-San-Martin, B., & Soutoglou, E. (2016) Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Mol. Cell*, **63**, 293–305.

Turunen, J.J., Niemelä, E.H., Verma, B., & Frilander, M.J. (2013) The significant other: Splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA*, **4**, 61–76.

Twyffels, L., Gueydan, C., & Kruys, V. (2011) Shuttling SR proteins: More than splicing factors. *FEBS J.*, **278**, 3246–3255.

Valacca, C., Bonomi, S., Buratti, E., Pedrotti, S., Baralle, F.E., Sette, C., Ghigna, C., & Biamonti, G. (2010) Sam68 regulates EMT through alternative splicing-activated nonsense-mediated mRNA decay of the SF2/ASF proto-oncogene. *J. Cell Biol.*, **191**, 87–99.

Valente, E.M., Rosti, R.O., Gibbs, E., & Gleeson, J.G. (2013) Primary cilia in neurodevelopmental disorders. *Nat. Rev. Neurol.*, **10**, 27–36.

Vessey, J.P., Macchi, P., Stein, J.M., Mikl, M., Hawker, K.N., Vogelsang, P., Wieczorek, K., Vendra, G., Riefler, J., Tübing, F., Aparicio, S. a J., Abel, T., & Kiebler, M. a (2008) A loss of function allele for murine Staufen1 leads to impairment of dendritic Staufen1-RNP delivery and dendritic spine morphogenesis. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 16374–16379.

Vogel, P., Read, R.W., Hansen, G.M., Payne, B.J., Small, D., Sands, a T., & Zambrowicz, B.P. (2012) Congenital hydrocephalus in genetically engineered mice. *Vet. Pathol.*, **49**, 166–181.

Vuong, C.K., Black, D.L., & Zheng, S. (2016) The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.*, **17**, 265–281.

Wahl, M.C. & Lührmann, R. (2015) SnapShot : Spliceosome Dynamics II SnapShot : Spliceosome Dynamics II. *Cell*, **162**, 456–456.e1.

Wahl, M.C., Will, C.L., & Lührmann, R. (2009) The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, **136**, 701–718.

Wallace, E.W.J. & Beggs, J.D. (2017) Extremely fast and incredibly close: cotranscriptional splicing in budding yeast. *Rna*, **23**, 601–610.

Wang, F. & Qi, L.S. (2016) Applications of CRISPR Genome Engineering in Cell Biology. *Trends Cell Biol.*, **26**, 875–888.

Wang, H.-Y., Xu, X., Ding, J.-H., Bermingham, J.R., & Fu, X.-D. (2001) SC35 Plays a Role in T Cell Development and Alternative Splicing of CD45. *Mol. Cell*, **7**, 331–342.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., & Jaenisch, R. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell*, **153**, 910–918.

Wang, J., Takagaki, Y., & Manley, J.L. (1996) Targeted disruption of an essential vertebrate gene : ASF / SF2 is required for cell viability. *Genes Dev.*, 2588–2599.

Wang, J., Xiao, S.H., & Manley, J.L. (1998) Genetic analysis of the SR protein ASF/SF2: Interchangeability of RS domains and negative control of splicing. *Genes Dev.*, **12**, 2222–2233.

Wang, Z., Murigneux, V., & Le Hir, H. (2014) Transcriptome-wide modulation of splicing by the exon junction complex. *Genome Biol.*, **15**, 551.

Wells, S.E., Hillner, P.E., Vale, R.D., & Sachs, A.B. (1998) Circularization of mRNA by Eukaryotic Translation Initiation Factors. *Mol. Cell*, **2**, 135–140.

White, E.S., Sagana, R.L., Booth, A.J., Yan, M., Cornett, A.M., Bloomheart, C.A., Tsui, J.L., Wilke, C.A., Moore, B.B., Ritzenthaler, J.D., Roman, J., & Muro, A.F. (2010) Control of fibroblast fibronectin expression and alternative splicing via the PI3K/Akt/mTOR pathway. *Exp. Cell Res.*, **316**, 2644–2653.

Will, C.L. & Lührmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.*, **3**, 1–2.

Wong, Q.W.L., Vaz, C., Lee, Q.Y., Zhao, T.Y., Luo, R., Archer, S.K., Preiss, T., Tanavde, V., & Vardy, L.A. (2016) Embryonic stem cells exhibit mRNA isoform specific translational regulation. *PLoS One*, **11**, 1–19.

Wurth, L. & Gebauer, F. (2014) RNA-binding proteins, multifaceted translational regulators in cancer. *BBA - Gene Regul. Mech.*, 1–6.

Xiang, S., Gapsys, V., Kim, H.Y., Bessonov, S., Hsiao, H.H., Möhlmann, S., Klaukien, V., Ficner, R., Becker, S., Urlaub, H., Lührmann, R., De Groot, B., & Zweckstetter, M. (2013) Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure*, **21**, 2162–2174.

Xiao, R., Sun, Y., Ding, J.-H., Lin, S., Rose, D.W., Rosenfeld, M.G., Fu, X.-D., & Li, X. (2007) Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol. Cell. Biol.*, **27**, 5393–5402.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y.S., Hao, Y.J., Sun, B.F., Sun, H.Y., Li, A., Ping, X.L., Lai, W.Y., Wang, X., Ma, H.L., Huang, C.M., Yang, Y., Huang, N., Jiang, G. Bin, Wang, H.L., Zhou, Q., Wang, X.J., Zhao, Y.L., & Yang, Y.G. (2016) Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell*, **61**, 507–519.

Xie, N., Chen, M., Dai, R., Zhang, Y., Zhao, H., Song, Z., Zhang, L., Li, Z., Feng, Y., Gao, H., Wang, L., Zhang, T., Xiao, R.-P., Wu, J., & Cao, C.-M. (2017) SRSF1 promotes vascular smooth muscle cell proliferation through a Δ133p53/EGR1/KLF5 pathway. *Nat. Commun.*, **8**, 16016.

Xu, X., Yang, D., Ding, J.H., Wang, W., Chu, P.H., Dalton, N.D., Wang, H.Y., Bermingham, J.R., Ye, Z., Liu, F., Rosenfeld, M.G., Manley, J.L., Ross, J., Chen, J., Xiao, R.P., Cheng, H., & Fu, X.D. (2005) ASF/SF2-regulated CaMKIIδ alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. *Cell*, **120**, 59–72.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., & Jaenisch, R. (2013) XOne-step generation of mice carrying reporter and conditional alleles by CRISPR/cas-mediated genome engineering. *Cell*, **154**, 1370–1379.

Yang, X. & Xu, T. (2011) Molecular mechanism of size control in development and human diseases. *Nat. Publ. Gr.*, **21**, 715–729.

Yoffe, Y., David, M., Kalaora, R., Povodovski, L., Friedlander, G., Feldmesser, E., Ainbinder, E., Saada, A., Bialik, S., & Kimchi, A. (2016) Cap-independent translation by DAP5 controls cell fate decisions in human embryonic stem cells. *Genes Dev.*, **30**, 1–16.

Yusa, K., Horie, K., Kondoh, G., & Kouno, M. (2004) Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom ' s syndrome gene. *Nature*, **429**, 896–899.

Zaher, H.S. & Green, R. (2009) Fidelity at the Molecular Level: Lessons from Protein Synthesis. *Cell*, **136**, 746–762.

Zahler, A.M., Neugebauer, K.M., Lane, W.S., & Roth, M.B. (1993) Distinct functions of SR proteins in alternative pre-messenger RNA splicing. *Science (80-.).*, **260**, 219–222.

Zhang, X., Yan, C., Hang, J., Finci, L.I., Lei, J., & Shi, Y. (2017) An Atomic Structure of the Human Spliceosome. *Cell*, **169**, 918–919.e14.

Zhang, Z. & Krainer, A.R. (2004) Involvement of SR proteins in mRNA surveillance. *Mol. Cell*, **16**, 597–607.

Zhao, B.S., Roundtree, I.A., & He, C. (2016) Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell Biol.*, **18**, 31–42.

Zhao, X., Yang, Y.Y.-G., Sun, B.-F., Shi, Y., Yang, X., Xiao, W., Hao, Y.-J., Ping, X.-L., Chen, Y.-S., Wang, W.-J., Jin, K.-X., Wang, X.X.-J., Huang, C.-M., Fu, Y., Ge, X.-M., Song, S.-H., Jeong, H.S., Yanagisawa, H., Niu, Y., Jia, G.-F., Wu, W., Tong, W.-M., Okamoto, A., He, C., Rendtlew Danielsen, J.M., Wang, X.X.-J., & Yang, Y.Y.-G. (2014) FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.*, **24**, 1403–1419.

Zhou, Z., Licklider, L.J., Gygi, S.P., & Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419**, 182–185.

Zhu, J. & Krainer, A.R. (2000) Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes Dev.*, **14**, 3166–3178.

Chapter 8

Appendix

Supplementary Table 1: IP-MS interactome of overexpressed SR proteins in HeLa cells (described Section 4.3.1). Fold enrichment over EV (empty vector control cells) and p-values for all hits obtained in the IP-MS experiments in HeLa cells that overexpressed either T7-SRSF1, T7-SRSF1-NRS or T7-SRSF2. Only processed data values are shown, raw intensity values and LFQ values omitted for ease. For processing methods used, see Section 4.3.1.1 for details. Where N/A is shown for p-value and 1.00 is shown for fold enrichment, this represents a scenario where there was no binding in either condition tested. All values shown to two decimal places. Hits are in alphabetical order according to Gene Name as imputed from the UniProt database. Protein ID refers to the given UniProt ID for each protein identified. Where multiple proteins are listed, this indicates that the interacting peptide identified could map to multiple proteins or isoforms of a protein. Database search to map the raw data hits was carried out by Jimi Wills of the IGMM Mass Spectrometry facility.

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
A2M;PZP	P01023; P20742	Alpha-2-macroglobulin;Pregnancy zone protein	4.09	1.64E-04	4.51	1.85E-06	1.70	3.05E-01
AATF	Q9NY61	Protein AATF	4009022.50	1.10E-05	9301803.49	1.85E-08	2676352.76	1.21E-03
ABCF1	F5GYK6	ATP-binding cassette sub-family F member 1	2323383.50	1.48E-07	5193165.94	1.27E-07	1371286.87	5.86E-09
ABCF2	Q9UG63	ATP-binding cassette sub-family F member 2	17.70	7.27E-09	32.13	1.34E-13	12.51	3.90E-09
ACTA1;ACTC1;ACTG2;ACTA2	P68133;P68032;P63267;P62736;	Actin, alpha skeletal muscle;Actin, alpha cardiac muscle 1;Actin, gamma-enteric smooth muscle;Actin, aortic smooth muscle	13.78	2.43E-03	40.74	7.40E-07	10.00	1.97E-02
ACTB	P60709	Actin, cytoplasmic 1;Actin, cytoplasmic 1, N-terminally processed	3.79	1.78E-08	5.60	1.99E-10	2.51	6.99E-08
ACTG1	P63261	Actin, cytoplasmic 2;Actin, cytoplasmic 2, N-terminally processed	2.47	5.50E-03	2.38	4.46E-02	0.85	7.52E-01
ACTL6A	O96019	Actin-like protein 6A	405367.81	5.34E-02	1478035.87	1.09E-03	53989.61	3.41E-01
ADAM15	Q13444	Disintegrin and metalloproteinase domain-containing protein 15	2.79	2.25E-02	1.13	8.62E-01	0.91	8.52E-01
ADAM9	Q13443	Disintegrin and metalloproteinase domain-containing protein 9	2.18	1.28E-03	1.89	7.59E-03	1.49	6.15E-02
ADSL	P30566	Adenylosuccinate lyase	3.18	9.11E-04	2.02	3.10E-01	1.44	4.17E-01

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
ADSSL1	Q8N142	Adenylosuccinate synthetase isozyme 1	3.29	2.72E-04	4.50	5.68E-06	1.25	7.01E-01
AGRN	O00468	Agrin;	2.81	4.89E-05	3.90	8.34E-08	1.79	2.61E-03
AHSG	P02765	Alpha-2-HS-glycoprotein	25.53	4.83E-03	15.54	2.85E-02	1.81	5.89E-01
AIFM1	O95831	Apoptosis-inducing factor 1, mitochondrial	15.17	1.72E-01	13.59	3.77E-01	25.98	1.38E-03
ALB	A0A0C4DGB6	Serum albumin	1.41	5.31E-01	3.59	2.78E-05	1.41	4.03E-01
ALDH9A1	P49189	4-trimethylaminobutyraldehyde dehydrogenase	1.83	5.05E-02	1.48	5.19E-01	1.20	5.34E-01
ALDOA	J3KPS3	Fructose-bisphosphate aldolase	2.37	2.67E-03	0.43	2.16E-01	1.09	7.61E-01
ALYREF	Q86V81	THO complex subunit 4	12544.43	5.88E-10	40493.24	7.45E-11	13897.67	4.77E-11
ALYREF	E9PB61	THO complex subunit 4	496176.94	1.58E-02	998851.55	5.12E-02	686099.70	3.24E-03
ANXA2;ANXA2P2	P07355;A6NMY6	Annexin;Annexin A2;Putative annexin A2-like protein	3.30	1.31E-05	7.90	1.43E-07	3.60	3.49E-04
AP2A1	O95782	AP-2 complex subunit alpha-1	37.63	7.16E-05	142.00	2.93E-07	29.98	3.64E-06
AP2B1;AP1B1	P63010;Q10567	AP-2 complex subunit beta;AP-1 complex subunit beta-1	33.41	6.31E-07	128.66	4.60E-06	35.29	1.42E-08
AP2M1	C9JJD3	AP-2 complex subunit mu	78.99	2.99E-07	309.69	2.81E-08	71.35	8.26E-04
APC2	O95996	Adenomatous polyposis coli protein 2	574628.32	8.03E-04	208017.13	1.52E-01	94869.99	3.41E-01
APOE	E9PEV4	Apolipoprotein E	2.10	5.59E-02	2.24	4.04E-07	1.47	5.09E-04
ARG1	P05089	Arginase-1	0.00	1.29E-01	0.00	1.29E-01	0.00	1.29E-01
ARHGEF2	V9GYF5	Rho guanine nucleotide exchange factor 2	764736.01	8.84E-04	829739.96	5.49E-02	398478.68	5.04E-02
ASF1B;ASF1A	Q9NVP2;Q9Y294	Histone chaperone ASF1B;Histone chaperone ASF1A	2787530.06	6.96E-09	4959581.36	1.27E-03	543424.76	1.46E-01
ATAD3A;ATAD3B;ATAD3C	Q9NVI7;Q5T9A4;Q5T2N8	ATPase family AAA domain-containing protein 3A;ATPase family AAA domain-containing protein	5.38	1.49E-05	5.42	6.02E-02	2.55	2.29E-02

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
		3B;ATPase family AAA domain-containing protein 3C						
ATIC	C9JLK0	Bifunctional purine biosynthesis protein PURH	3.90	1.54E-05	4.29	5.21E-08	2.21	2.90E-04
ATP5A1	P25705	ATP synthase subunit alpha, mitochondrial	3.89	2.32E-06	5.16	3.78E-09	2.38	7.26E-06
ATP5C1	P36542	ATP synthase subunit gamma, mitochondrial	5.44	2.45E-03	9.93	2.01E-02	4.64	1.65E-02
ATXN2L	H3BUF6	Ataxin-2-like protein	14.73	3.56E-07	29.35	4.32E-09	11.66	1.28E-06
AURKAIP1	Q9NWT8	Aurora kinase A-interacting protein	1857409.20	1.08E-02	7190657.05	1.73E-11	2234536.71	4.30E-07
BAG2	O95816	BAG family molecular chaperone regulator 2	5.25	3.00E-04	7.67	7.05E-07	2.62	3.07E-02
BBX	C9JA69	HMG box transcription factor BBX	2787690.02	2.06E-09	6060211.23	2.97E-08	1931178.84	4.56E-10
BCAS2	O75934	Pre-mRNA-splicing factor SPF27	4977533.12	1.12E-10	15430085.76	1.08E-12	4429470.94	4.50E-07
BCL9	O00512	B-cell CLL/lymphoma 9 protein	1.55	4.17E-02	0.00	8.99E-05	0.00	8.99E-05
BCLAF1	E9PK09	Bcl-2-associated transcription factor 1	6250120.07	4.04E-09	15245083.53	1.47E-09	9697325.13	2.99E-10
BIN2	S4R418	Bridging integrator 2	3.42	7.52E-02	1.64	7.07E-01	1.51	5.05E-01
BMS1	Q14692	Ribosome biogenesis protein BMS1 homolog	5405392.52	8.36E-09	13042824.04	2.57E-10	4115624.74	4.42E-11
BRIX1	Q8TDN6	Ribosome biogenesis protein BRX1 homolog	1008.12	4.85E-12	2690.85	1.65E-12	614.43	1.36E-06
BYSL	Q13895	Bystin	8062032.89	1.13E-09	22145522.49	1.43E-11	8478850.42	3.81E-06
C11orf98	E9PRG8	Uncharacterized protein C11orf98	51820025.70	3.03E-12	107370143.92	5.29E-12	39148194.30	9.67E-16
C14orf166	Q9Y224	UPF0568 protein C14orf166	111.30	2.02E-05	215.44	1.66E-07	66.10	9.62E-04
C15orf62	A8K5M9	Uncharacterized protein C15orf62, mitochondrial	1.21	7.95E-01	0.00	3.18E-07	0.83	6.79E-01
C18orf21	L7N2F3	UPF0711 protein C18orf21	769829.30	1.18E-08	2008290.47	1.52E-03	368018.15	1.14E-02

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
C19orf53	Q9UNZ5	Leydig cell tumor 10 kDa protein homolog	1819.32	2.45E-09	4445.23	1.54E-10	1570.17	6.52E-07
C19orf84	K7EN44	Uncharacterized protein C19orf84	766.90	2.02E-09	2070.28	3.98E-06	978.23	4.26E-08
C1QBP	Q07021	Complement component 1 Q subcomponent-binding protein, mitochondrial	6.82	4.18E-01	0.00	3.41E-01	105.65	1.49E-07
C3orf17	Q6NW34	Uncharacterized protein C3orf17	1947761.15	1.31E-08	5493433.67	2.26E-06	1473678.69	1.61E-07
C4B;C4A	F5GXS0	Complement C4-B; Complement C4-A alpha chain	2.82	5.29E-02	0.00	6.05E-02	0.00	6.05E-02
C7orf50	C9JQV0	Uncharacterized protein C7orf50	22984532.18	3.82E-14	46554268.58	5.56E-13	12768637.69	7.37E-06
C8orf33	Q9H7E9	UPF0488 protein C8orf33	94.49	7.90E-13	238.46	4.36E-11	57.99	3.56E-07
C9orf114	Q5T280	Putative methyltransferase C9orf114	490251.88	7.12E-02	3097304.95	4.02E-07	598574.39	1.02E-02
CAD	F8VPD4	CAD protein	0.97	9.62E-01	0.47	3.35E-01	1.25	7.15E-01
CAPRIN1	Q14444	Caprin-1	21.32	2.55E-11	73.41	1.02E-14	26.78	6.95E-13
CARM1	K7EK20	Histone-arginine methyltransferase CARM1	2.55	5.96E-02	1.56	3.30E-01	1.21	6.61E-01
CBLN1	H3BQD0	Cerebellin-1	0.00	5.08E-02	0.00	5.08E-02	0.00	5.08E-02
CCAR1	F5H3E1	Cell division cycle and apoptosis regulator protein 1	213119.16	1.46E-01	2094800.67	6.46E-04	26573.28	3.41E-01
CCAR2	Q8N163	Cell cycle and apoptosis regulator protein 2	62982.19	1.56E-01	518894.95	1.48E-02	53871.30	1.76E-01
CCDC137	I3L0U5	Coiled-coil domain-containing protein 137	9928582.10	7.46E-12	18771452.64	4.85E-15	6954365.61	1.74E-13
CCDC59	Q9P031	Thyroid transcription factor 1-associated protein 26	4373082.31	2.04E-09	9063665.86	2.64E-11	3458757.91	3.55E-10
CCDC86	Q9H6F5	Coiled-coil domain-containing protein 86	42023407.28	9.65E-10	127096867.69	5.81E-10	45747604.28	1.68E-11
CCT2	F5GWF6	T-complex protein 1 subunit beta	2.68	2.99E-05	2.21	1.83E-02	1.76	2.56E-03
CCT3	P49368	T-complex protein 1 subunit gamma	3.52	9.37E-07	4.11	1.33E-08	2.09	3.67E-05

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
CCT4	P50991	T-complex protein 1 subunit delta	3.11	3.10E-05	3.48	1.34E-08	1.80	6.01E-05
CCT5	P48643	T-complex protein 1 subunit epsilon	3.20	6.32E-06	4.15	2.68E-05	1.98	4.78E-05
CCT6A	P40227	T-complex protein 1 subunit zeta	3.02	1.44E-06	3.59	2.29E-05	1.55	9.88E-03
CCT6B	Q92526	T-complex protein 1 subunit zeta-2	7.63	4.10E-03	0.00	1.46E-01	0.00	1.46E-01
CCT7	Q99832	T-complex protein 1 subunit eta	3.34	1.29E-05	4.06	1.10E-09	2.06	8.11E-04
CCT8	P50990	T-complex protein 1 subunit theta	3.00	3.15E-05	3.57	4.72E-08	1.71	5.33E-04
CD320	Q9NPF0	CD320 antigen	2.90	3.97E-02	4.99	2.65E-03	2.83	1.89E-03
CD3EAP	O15446	DNA-directed RNA polymerase I subunit RPA34	26094533.50	1.20E-11	48373198.10	2.02E-13	13526128.20	4.70E-13
CD97	P48960	CD97 antigen	2.76	2.19E-03	1.34	5.88E-01	1.38	2.06E-01
CDC5L	Q99459	Cell division cycle 5-like protein	244.93	1.90E-11	724.00	5.06E-14	221.95	1.42E-11
CDK12	Q9NYV4	Cyclin-dependent kinase 12	4458910.12	3.69E-11	9031184.97	5.23E-11	4209738.41	8.74E-08
CDK13	Q14004	Cyclin-dependent kinase 13	2805742.19	3.29E-10	4096339.22	8.45E-09	1329084.57	8.44E-04
CDKN2A	Q8N726	Tumor suppressor ARF	4500750.04	1.87E-11	11779845.11	6.32E-11	4036540.32	2.44E-05
CEBPZ	Q03701	CCAAT/enhancer-binding protein zeta	517.77	1.12E-08	1156.80	1.32E-12	350.77	1.64E-08
CENPV	Q7Z7K6	Centromere protein V	1879094.33	1.21E-08	4559722.07	8.39E-08	976098.18	1.04E-02
CFL1;CFL2	E9PP50	Cofilin-1;Cofilin-2	3.86	6.71E-06	5.71	5.69E-08	2.92	2.55E-02
CHERP	J3QK89	Calcium homeostasis endoplasmic reticulum protein	1453938.35	8.09E-04	4948587.66	2.65E-09	1901602.56	1.26E-09
CHTOP	Q5T7Y7	Chromatin target of PRMT1 protein	649462.75	1.11E-02	2737644.57	1.63E-03	911882.11	6.32E-03
CIRBP	Q14011	Cold-inducible RNA-binding protein	1130454.80	1.09E-03	7792080.88	1.24E-08	2385753.52	1.13E-09
CLTC	Q00610	Clathrin heavy chain	0.00	4.76E-06	0.00	4.76E-06	0.00	4.76E-06
CMAS	Q8NFW8	N-acylneuraminate cytidylyltransferase	1100328.06	8.88E-10	2798602.97	2.30E-10	802220.79	8.82E-04
CMSS1	C9J384	Protein CMSS1	6874336.87	1.18E-06	21065713.53	1.12E-10	10140065.27	1.20E-12
CNBP	P62633	Cellular nucleic acid-binding protein	3.26	1.96E-05	4.00	2.34E-07	1.89	2.71E-04

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
COPA	P53621	Coatomer subunit alpha	9.62	2.78E-06	16.88	2.39E-06	7.45	9.76E-07
COPB2	P35606	Coatomer subunit beta	1130.79	6.08E-04	1486.33	5.00E-02	373.18	1.48E-01
CP	E9PFZ2	Ceruloplasmin	3.12	3.16E-06	3.85	8.97E-07	1.84	1.22E-03
CPS1	P31327	Carbamoyl-phosphate synthase [ammonia], mitochondrial	3.11	2.68E-03	3.35	1.58E-03	0.79	7.21E-01
CREBBP	Q92793	CREB-binding protein	2.23	9.16E-04	0.07	2.72E-04	0.00	6.54E-05
CRELD2	Q6UXH1	Cysteine-rich with EGF-like domain protein 2	1.12	8.33E-01	0.00	1.56E-03	0.00	1.56E-03
CRIM1	Q9NZV1	Cysteine-rich motor neuron 1 protein	2.48	1.06E-04	3.08	5.97E-07	1.55	2.76E-02
CSDE1	E9PLT0	Cold shock domain-containing protein E1	13.60	8.91E-10	23.63	8.91E-10	10.50	2.19E-02
CSE1L	P55060	Exportin-2	1.32	5.38E-01	0.48	3.24E-01	2.58	1.04E-01
CSRP1	P21291	Cysteine and glycine-rich protein 1	3.96	5.34E-04	3.99	5.25E-02	3.11	5.06E-04
CTCF	P49711	Transcriptional repressor CTCF	6439805.01	2.65E-07	14723834.74	6.82E-10	6235191.67	2.08E-12
CTGF	P29279	Connective tissue growth factor	2.71	1.68E-02	5.61	1.63E-05	1.41	3.65E-01
CTSZ	Q9UBR2	Cathepsin Z	4.36	1.39E-01	9.51	4.39E-06	4.98	5.73E-02
CYR61	O00622	Protein CYR61	3.46	2.34E-07	4.86	3.97E-09	2.26	2.86E-04
DAP3	P51398	28S ribosomal protein S29, mitochondrial	108.40	8.07E-04	674.11	2.14E-07	337.22	2.76E-09
DARS	H7BZ35	Aspartate--tRNA ligase, cytoplasmic	4.65	4.39E-06	4.22	4.28E-03	2.29	8.37E-07
DAZAP1	K7EQ55	DAZ-associated protein 1	3.53	4.81E-05	2.22	2.19E-05	0.84	3.34E-01
DCD	P81605	Dermcidin	0.16	2.96E-01	0.05	2.37E-01	0.01	2.18E-01
DDX1	Q92499	ATP-dependent RNA helicase DDX1	15.21	8.97E-06	33.12	1.22E-11	9.65	1.82E-03
DDX10	E9PIF2	Probable ATP-dependent RNA helicase DDX10	1617127.56	2.56E-07	3874775.18	1.27E-05	1479226.25	1.93E-07
DDX17	H3BLZ8	Probable ATP-dependent RNA helicase DDX17	7.42	1.80E-05	26.55	1.42E-12	8.24	7.41E-07

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
DDX18	Q9NVP1	ATP-dependent RNA helicase DDX18	850.19	5.16E-09	2662.19	5.24E-10	638.90	4.26E-06
DDX20	Q9UHI6	Probable ATP-dependent RNA helicase DDX20	314.92	1.59E-08	591.76	4.04E-07	253.80	7.80E-04
DDX21	Q9NR30	Nucleolar RNA helicase 2	2352.94	1.50E-08	5793.07	1.88E-18	2232.66	6.10E-10
DDX23	F8VVA2	Probable ATP-dependent RNA helicase DDX23	500342.98	1.46E-01	296792.83	3.41E-01	773868.14	7.91E-04
DDX24	G3V529	ATP-dependent RNA helicase DDX24	31668313.97	7.80E-12	74947003.35	1.20E-14	24066698.99	1.09E-10
DDX27	B7Z6D5	Probable ATP-dependent RNA helicase DDX27	921.82	7.87E-10	2472.55	8.70E-13	711.46	5.06E-08
DDX28	Q9NUL7	Probable ATP-dependent RNA helicase DDX28	65.61	2.81E-08	183.80	2.04E-14	80.78	1.86E-10
DDX31	F8WAJ0	Probable ATP-dependent RNA helicase DDX31	499.52	2.18E-13	1074.69	2.37E-12	318.86	2.48E-14
DDX3X;DDX3Y	O00571;O15523	ATP-dependent RNA helicase DDX3X;ATP-dependent RNA helicase DDX3Y	3.49	1.59E-04	9.83	4.58E-09	3.43	1.66E-03
DDX47	Q9H0S4	Probable ATP-dependent RNA helicase DDX47	130.82	9.97E-07	387.20	1.28E-09	92.17	1.54E-10
DDX5	P17844	Probable ATP-dependent RNA helicase DDX5	9.96	2.00E-07	33.15	4.87E-13	9.43	2.25E-08
DDX50	Q9BQ39	ATP-dependent RNA helicase DDX50	9992221.73	3.93E-09	25952115.57	3.28E-10	11455426.44	9.58E-10
DDX52	Q9Y2R4	Probable ATP-dependent RNA helicase DDX52	13436805.27	2.91E-09	28875617.25	3.69E-13	9194186.68	2.31E-08
DDX54	Q8TDD1	ATP-dependent RNA helicase DDX54	399.31	7.07E-08	1079.95	4.15E-10	347.60	6.02E-08
DDX55	F5H5U2	ATP-dependent RNA helicase DDX55	4103122.32	4.53E-09	11375027.80	3.33E-09	3047703.77	3.93E-06
DDX56	Q9NY93	Probable ATP-dependent RNA helicase DDX56	15370010.43	1.85E-10	31975348.55	6.86E-13	8586305.57	1.76E-07
DHX15	O43143	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	44.81	6.35E-12	130.56	2.51E-14	49.36	1.31E-13

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
DHX30	H7BXY3	Putative ATP-dependent RNA helicase DHX30	389.83	6.32E-10	1041.95	1.40E-14	331.74	5.25E-09
DHX36	Q9H2U1	ATP-dependent RNA helicase DHX36	720.71	6.45E-11	2280.39	2.01E-13	697.88	5.10E-12
DHX37	Q8IY37	Probable ATP-dependent RNA helicase DHX37	12987763.12	3.62E-12	31441107.41	4.62E-11	9841609.99	1.12E-11
DHX57	Q6P158	Putative ATP-dependent RNA helicase DHX57	125.34	1.29E-08	296.53	2.69E-12	113.41	2.22E-11
DHX9	Q08211	ATP-dependent RNA helicase A	135.61	6.69E-09	571.86	1.27E-14	162.29	2.33E-10
DIDO1	Q9BTC0	Death-inducer obliterator 1	8598205.34	1.67E-14	14315963.67	4.87E-10	4978360.21	2.14E-08
DKC1	O60832	H/ACA ribonucleoprotein complex subunit 4	6297101.43	2.78E-08	17548727.79	2.96E-13	5121861.46	6.20E-06
DNAJA3	Q96EY1	DnaJ homolog subfamily A member 3, mitochondrial	13.04	1.04E-06	17.65	4.80E-09	6.26	2.98E-03
DNAJC9	Q8WXX5	DnaJ homolog subfamily C member 9	3871940.22	2.01E-08	6773458.13	1.67E-08	3106494.68	1.58E-08
DNTTIP2	Q5QJE6	Deoxynucleotidyltransferase terminal-interacting protein 2	3377173.51	3.98E-08	7993573.40	2.63E-08	2338817.24	2.14E-08
DSC1	Q08554	Desmocollin-1	0.13	2.13E-01	0.00	1.49E-01	0.00	1.49E-01
DSG1	Q02413	Desmoglein-1	0.10	1.84E-01	0.00	1.45E-01	0.01	1.49E-01
DSP	P15924	Desmoplakin	0.33	2.96E-01	0.05	1.46E-01	0.05	1.46E-01
DSTN	P60981	Destrin	3.28	6.93E-04	2.90	1.35E-02	1.35	5.06E-01
DUSP9	Q99956	Dual specificity protein phosphatase 9	2.25	1.57E-01	1.39	6.90E-01	0.97	9.51E-01
DYNC1H1	Q14204	Cytoplasmic dynein 1 heavy chain 1	1064462.03	1.66E-01	4929765.90	5.12E-02	1.00	N/A
EBNA1BP2	H7C2Q8	Probable rRNA-processing protein EBP2	7613.76	7.73E-10	17887.54	4.55E-14	4864.73	1.46E-08
EEF1A1P5	Q5VTE0	Putative elongation factor 1-alpha-like 3	4.03	2.47E-06	6.24	1.02E-09	2.88	7.66E-06
EEF1D	P29692	Elongation factor 1-delta	4.58	1.81E-05	5.20	3.81E-04	2.30	8.43E-02
EEF1G	P26641	Elongation factor 1-gamma	3.34	1.12E-05	4.51	1.62E-06	2.50	6.94E-07

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
EEF2	P13639	Elongation factor 2	3.62	5.63E-07	3.66	3.65E-06	2.37	6.16E-05
EFEMP1	Q12805	EGF-containing fibulin-like extracellular matrix protein 1	4.06	3.37E-06	4.84	3.42E-06	1.83	3.84E-02
EGFR	E9PFD7	Receptor protein-tyrosine kinase	3.79	5.51E-05	0.00	8.80E-03	0.42	1.92E-01
EIF2S1	P05198	Eukaryotic translation initiation factor 2 subunit 1	32.33	6.19E-10	73.07	4.06E-08	24.80	7.76E-08
EIF2S2	P20042	Eukaryotic translation initiation factor 2 subunit 2	40.43	1.39E-07	103.43	8.63E-09	30.23	6.29E-09
EIF2S3;EIF2S3L	P41091;F8W810	Eukaryotic translation initiation factor 2 subunit 3;Putative eukaryotic translation initiation factor 2 subunit 3-like protein	61.40	1.07E-10	128.85	9.58E-10	49.32	1.34E-08
EIF3A	Q14152	Eukaryotic translation initiation factor 3 subunit A	3.32	3.10E-04	3.58	8.64E-05	1.70	1.04E-01
EIF3C;EIF3CL	Q99613;B5ME19	Eukaryotic translation initiation factor 3 subunit C;Eukaryotic translation initiation factor 3 subunit C-like protein	3.75	1.18E-02	6.12	1.15E-05	0.56	3.28E-01
EIF3E	P60228	Eukaryotic translation initiation factor 3 subunit E	1.80	3.56E-01	4.27	4.90E-09	1.08	8.81E-01
EIF3I	Q13347	Eukaryotic translation initiation factor 3 subunit I	4.88	3.09E-06	5.17	2.36E-03	2.04	1.43E-01
EIF4A1;EIF4A2	P60842;Q14240	Eukaryotic initiation factor 4A-I;Eukaryotic initiation factor 4A-II	2.94	1.03E-01	2.00	4.49E-01	1.65	2.71E-01
EIF4G1	E7EX73	Eukaryotic translation initiation factor 4 gamma 1	6.75	1.06E-07	11.67	1.13E-05	3.64	7.20E-03
EIF4H	Q15056	Eukaryotic translation initiation factor 4H	1.05	9.47E-01	0.00	1.39E-03	2.67	1.45E-03
EIF5A;EIF5A2;EIF5AL1	P63241;Q9GZV4;Q6IS14	Eukaryotic translation initiation factor 5A;Eukaryotic translation initiation factor 5A-1;Eukaryotic translation	4.80	2.35E-06	4.75	1.97E-03	2.86	6.41E-05

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
		initiation factor 5A-2;Eukaryotic translation initiation factor 5A-1-like						
EIF6	P56537	Eukaryotic translation initiation factor 6	120.83	1.85E-09	280.55	5.77E-09	81.15	2.13E-09
ELAVL1	Q15717	ELAV-like protein 1	8.65	8.20E-07	20.83	1.40E-08	12.73	2.54E-09
ELMSAN1	Q6PJG2	ELM2 and SANT domain-containing protein 1	1476645.25	5.74E-04	1873797.40	1.14E-03	328068.23	5.15E-02
EMG1	Q92979	Ribosomal RNA small subunit methyltransferase NEP1	390025.74	5.07E-02	2174942.23	2.29E-07	982778.80	1.54E-03
ENO1	P06733	Alpha-enolase;Enolase	3.76	1.58E-05	3.83	2.61E-08	2.44	2.28E-04
EP300	Q09472	Histone acetyltransferase p300	2.20	2.93E-03	0.04	3.82E-06	0.02	2.96E-06
EPRS	P07814	Bifunctional glutamate/proline--tRNA ligase	5.64	4.15E-07	5.99	2.12E-03	0.45	2.73E-01
ERC1	Q8IUD2	ELKS/Rab6-interacting/CAST family member 1	4.00	3.02E-05	5.24	3.00E-09	2.29	3.75E-04
ERH	P84090	Enhancer of rudimentary homolog	59.19	2.24E-04	209.60	6.32E-08	116.52	1.23E-05
EWSR1	Q01844;	RNA-binding protein EWS	2.69	1.01E-01	4.28	8.77E-03	3.01	1.48E-01
EXOSC1	Q9Y3B2	Exosome complex component CSL4	3878612.67	8.65E-07	10018548.33	2.89E-08	1976025.50	2.04E-08
EXOSC10	Q0178	Exosome component 10	23562201.78	2.35E-10	53391911.16	1.01E-10	12181070.19	1.42E-07
EXOSC2	Q13868	Exosome complex component RRP4	7688695.63	1.68E-09	18011622.03	1.95E-11	4248476.92	3.41E-09
EXOSC3	Q9NQT5	Exosome complex component RRP40	384.49	2.59E-05	974.78	1.13E-09	192.08	1.03E-05
EXOSC4	Q9NPD3	Exosome complex component RRP41	10472455.30	6.56E-10	24248284.97	1.40E-13	4767964.71	1.00E-04
EXOSC5	Q9NQT4	Exosome complex component RRP46	876963.41	1.61E-03	3465910.24	5.25E-07	433674.70	1.22E-02
EXOSC6	Q5RKV6	Exosome complex component MTR3	14379220.22	5.44E-09	31515619.11	2.21E-11	7376530.95	2.16E-08
EXOSC7	Q15024	Exosome complex component RRP42	6266664.09	2.04E-06	15275343.50	3.45E-12	2898094.38	1.30E-06
EXOSC8	Q96B26	Exosome complex component RRP43	5970974.03	8.78E-08	14235455.49	1.27E-09	2781336.36	1.11E-05
EXOSC9	Q06265	Exosome complex component RRP45	5078282.54	1.52E-07	11354761.95	3.10E-11	2412094.90	1.66E-08

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
FABP5	Q01469	Fatty acid-binding protein, epidermal	0.40	3.58E-01	0.00	1.14E-01	0.00	1.14E-01
FAM111B	Q6SJ93	Protein FAM111B	1763366.18	3.91E-07	3902297.87	8.00E-08	1850788.57	8.91E-04
FAM120A	Q9NZB2	Constitutive coactivator of PPAR-gamma-like protein 1	30.19	1.11E-12	79.23	1.26E-12	40.44	4.10E-08
FAU	P62861	40S ribosomal protein S30	152.56	3.52E-11	301.48	3.87E-10	159.26	1.54E-08
FBL	P22087	rRNA 2-O-methyltransferase fibrillarin	408.43	1.57E-08	1439.75	6.67E-13	490.32	5.81E-10
FBLL1	A6NHQ2	rRNA/tRNA 2-O-methyltransferase fibrillarin-like protein 1	399.95	3.91E-09	1116.26	9.10E-10	428.17	7.36E-11
FBLN1	P23142	Fibulin-1	3.33	2.47E-02	2.43	2.39E-01	2.75	4.74E-02
FBN1	P35555	Fibrillin-1	0.80	4.86E-01	0.74	3.54E-01	0.35	3.27E-02
FBN2	P35556	Fibrillin-2	1.63	6.78E-02	1.17	6.46E-01	0.72	3.61E-01
FBXO5	Q9UKT4	F-box only protein 5	0.60	5.70E-01	0.00	1.11E-02	0.25	9.31E-02
FGF6	P10767	Fibroblast growth factor 6	1425299.54	1.09E-02	8960782.46	5.95E-02	1.00	N/A
FHL2	J3KNW4	Four and a half LIM domains protein 2	2.26	1.48E-01	0.00	1.74E-03	0.75	6.48E-01
FIP1L1	Q6UN15	Pre-mRNA 3-end-processing factor FIP1	6726560.39	1.67E-09	12325968.15	8.16E-11	5085534.07	8.87E-13
FKBP4	Q02790	Peptidyl-prolyl cis-trans isomerase FKBP4	3.73	3.53E-05	0.00	1.14E-03	0.75	6.38E-01
FMR1	Q06787	Fragile X mental retardation protein 1	497927.15	7.29E-04	946704.12	1.03E-03	564211.88	4.64E-07
FOXK1	P85037	Forkhead box protein K1	180862.73	5.06E-02	895961.60	4.33E-03	6787496.68	3.87E-06
FRAS1	Q86XX4	Extracellular matrix protein FRAS1	1.85	1.62E-02	1.81	1.05E-02	1.03	8.97E-01
FST	P19883	Follistatin	2.15	7.00E-02	1.21	8.09E-01	0.00	1.54E-02
FSTL1	Q12841	Follistatin-related protein 1	2.88	1.04E-01	0.00	1.16E-02	0.19	5.92E-02
FSTL3	O95633	Follistatin-related protein 3	2.28	2.79E-04	2.79	2.22E-06	1.49	4.21E-02
FTSJ3	Q8IY81	pre-rRNA processing protein FTSJ3	52379643.13	1.52E-11	143012501.03	2.65E-11	45317422.01	5.03E-09
FUBP1	Q96AE4	Far upstream element-binding protein 1	3.00	8.47E-04	3.34	1.85E-05	1.69	3.22E-02

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
FUS	P35637	RNA-binding protein FUS	5.43	2.17E-02	13.06	4.25E-04	6.44	4.10E-02
FXR1	P51114	Fragile X mental retardation syndrome-related protein 1	104.63	6.87E-04	364.16	6.90E-06	71.94	5.26E-02
FXR2	P51116	Fragile X mental retardation syndrome-related protein 2	98.39	2.50E-10	233.41	1.17E-09	87.49	6.48E-11
G3BP1	Q13283	Ras GTPase-activating protein-binding protein 1	113.58	8.36E-08	446.56	6.26E-12	146.08	9.50E-09
G3BP2	Q9UN86	Ras GTPase-activating protein-binding protein 2	65.24	1.22E-08	202.81	8.06E-13	80.42	5.59E-08
GADD45GIP1	Q8TAE8	Growth arrest and DNA damage-inducible proteins-interacting protein 1	109.90	3.88E-03	914.15	2.29E-08	447.77	5.17E-10
GAPDH	P04406	Glyceraldehyde-3-phosphate dehydrogenase	2.98	2.16E-04	3.09	1.94E-05	1.21	5.12E-01
GAR1	Q9NY12	H/ACA ribonucleoprotein complex subunit 1	1997153.27	4.05E-05	6034282.04	2.06E-11	1216867.31	1.28E-05
GBE1	Q04446	1,4-alpha-glucan-branching enzyme	808294.31	3.92E-03	3342038.70	1.83E-08	1390306.68	7.24E-03
GCN1L1	Q92616	Translational activator GCN1	0.63	6.46E-01	0.88	9.04E-01	0.49	4.63E-01
GEMIN5	Q8TEQ6	Gem-associated protein 5	60210.38	1.46E-01	87554.86	3.41E-01	174836.14	5.35E-02
GLTSCR2	Q9NZM5	Glioma tumor suppressor candidate region gene 2 protein	12715220.98	1.84E-10	23739029.12	7.08E-12	7488386.37	7.78E-08
GLYR1	Q49A26	Putative oxidoreductase GLYR1	2477752.45	1.64E-04	5062076.59	2.03E-03	1794308.31	1.07E-03
GNB2L1	P63244	Receptor of activated protein C kinase 1	4.78	1.25E-06	2.84	3.34E-01	1.28	6.49E-01
GNL2	Q13823	Nucleolar GTP-binding protein 2	18074569.03	6.40E-10	42342858.18	2.61E-13	11567603.34	5.21E-11
GNL3	Q9BVP2	Guanine nucleotide-binding protein-like 3	12721.33	1.37E-10	29545.12	4.72E-13	8909.59	5.66E-11
GOLGB1	Q14789	Golgin subfamily B member 1	1.79	3.91E-01	2.09	1.69E-01	1.31	4.16E-01
GPATCH4	Q5T3I0	G patch domain-containing protein 4	83782430.17	1.16E-14	174217469.64	1.30E-14	63726450.54	1.31E-10
GRN	P28799	Granulins	2.94	1.28E-04	3.66	1.03E-06	1.62	1.54E-02

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
GRSF1	Q12849	G-rich sequence factor 1	9877693.25	6.83E-09	24912338.61	1.59E-08	7419850.56	1.07E-09
GRWD1	Q9BQ67	Glutamate-rich WD repeat-containing protein 1	40.80	6.88E-07	121.76	2.52E-12	30.82	3.88E-06
GTF3C1	Q12789	General transcription factor 3C polypeptide 1	374.38	5.12E-13	704.79	3.91E-13	252.03	2.95E-11
GTF3C2	Q8WUA4	General transcription factor 3C polypeptide 2	2851123.99	1.30E-10	5806749.28	5.30E-08	2211590.98	1.03E-06
GTF3C3	Q9Y5Q9	General transcription factor 3C polypeptide 3	51.00	3.19E-05	109.28	5.81E-06	25.14	1.34E-03
GTF3C4	Q9UKN8	General transcription factor 3C polypeptide 4	3320640.50	1.78E-10	8861173.92	4.63E-09	1956093.65	7.19E-06
GTPBP4	Q9BZE4	Nucleolar GTP-binding protein 1	1864.62	1.99E-11	4962.49	7.68E-15	1520.42	3.09E-12
H1F0	P07305	Histone H1.0;Histone H1.0, N-terminally processed	941.15	2.59E-11	1905.35	5.51E-14	662.41	1.39E-10
H1FX	Q92522	Histone H1x	1843.96	1.63E-08	4762.80	1.62E-09	1403.42	8.29E-13
H2AFY;H2AFY2	Q5SQT3;Q9P0M6	Histone H2A;Core histone macro-H2A.1	1.00	N/A	1314399.34	1.83E-01	87520.12	3.41E-01
HBA1	P69905	Hemoglobin subunit alpha	1788171.45	1.47E-01	4768193.99	8.51E-04	1205616.12	1.47E-01
hCG_1984214;MRPS17	Q9Y2R5	28S ribosomal protein S17, mitochondrial	19691724.02	1.35E-07	70192579.19	1.37E-11	25829622.52	6.15E-07
HIST1H1A	Q02539	Histone H1.1	96820442.32	1.20E-10	253566882.08	2.07E-13	83398499.05	2.90E-13
HIST1H1B	P16401	Histone H1.5	430.85	1.00E-07	1141.45	3.14E-12	362.82	1.42E-09
HIST1H1C	P16403	Histone H1.2	753.90	2.72E-11	1684.43	2.34E-07	624.32	2.73E-11
HIST1H1E;HIST1H1D	P10412;P16402	Histone H1.4;Histone H1.3	519.75	1.28E-08	1237.83	1.10E-12	414.74	1.19E-10
HIST1H2AC;HIST3H2A;HIST1H2AB;HIST1H2AA;H2AFX	P04908;P16104;Q71UI9;P0C0S5;C9J386;	Histone H2A type 1-C;Histone H2A type 3;Histone H2A type 1-B/E;Histone H2A type 1-A;Histone H2AX	3453175.62	1.06E-03	17473938.47	5.53E-08	3456768.91	9.43E-04

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
HIST1H2AJ;HIST1H2AH;H2AFJ;HIST2H2AC;HIST2H2AA3;HIST1H2AD;HIST1H2AG	Q99878;Q96KK5;Q9BTM1;Q16777;Q6FI13;P20671;P0C0S8;	Histone H2A type 1-J;Histone H2A type 1-H;Histone H2A.J;Histone H2A type 2-C;Histone H2A type 2-A;Histone H2A type 1-D;Histone H2A type 1;Histone H2A	213.43	6.83E-07	856.96	1.65E-07	284.74	6.09E-11
HIST1H2BN;HIST1H2BM;HIST1H2BH;HIST2H2BF;HIST1H2BC;HIST1H2BD;HIST1H2BK;H2BFS	Q99879;Q99877;Q93079;Q5QNW6;P62807;P58876;O60814;P57053	Histone H2B;Histone H2B type 1-M;Histone H2B type 1-N;Histone H2B type 1-H;Histone H2B type 2-F;Histone H2B type 1-C/E/F/G/I;Histone H2B type 1-D;Histone H2B type 1-K;Histone H2B type F-S	46.57	1.92E-09	172.98	6.69E-09	59.48	3.23E-12
HIST1H4A	P62805	Histone H4	27.61	2.87E-09	107.43	7.21E-05	33.30	4.73E-05
HIST2H2BE;HIST1H2BB;HIST1H2BO;HIST1H2BJ	Q16778;P33778;P23527;P06899;Q6DRA6;Q6DN03	Histone H2B type 2-E;Histone H2B type 1-B;Histone H2B type 1-O;Histone H2B type 1-J	1011333.79	1.46E-01	5361678.38	6.14E-02	3577444.81	8.30E-10
HIST2H3A;HIST3H3;HIST1H3A;H3F3B;H3F3A	Q71DI3;Q16695;P68431;K7EMV3;B4DEB1;K7EK07;K7ES00;P84243;K7EP01;Q6NXT2	Histone H3.2;Histone H3.1t;Histone H3.1;Histone H3;Histone H3.3	38.78	5.86E-12	106.98	4.42E-08	42.01	2.79E-06
HIST2H3PS2	Q5TEC6	Histone H3	12.09	9.37E-02	98.49	3.51E-04	36.83	1.08E-01
HMGA1	P17096	High mobility group protein HMG-I/HMG-Y	512.92	4.72E-09	1476.73	6.21E-07	500.85	1.71E-06

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
HNRNPA0	Q13151	Heterogeneous nuclear ribonucleoprotein A0	37.19	1.81E-06	159.33	7.81E-16	52.37	1.39E-10
HNRNPA1;HNRNPA1L2	P09651;Q32P51	Heterogeneous nuclear ribonucleoprotein A1; Heterogeneous nuclear ribonucleoprotein A1-like 2	9.47	1.05E-12	25.01	1.90E-10	11.12	1.52E-09
HNRNPA2B1	P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	6.23	1.65E-07	18.83	1.75E-08	10.40	1.13E-07
HNRNPA3	P51991	Heterogeneous nuclear ribonucleoprotein A3	5.83	1.97E-11	14.70	1.10E-10	7.22	3.50E-10
HNRNPAB	Q99729	Heterogeneous nuclear ribonucleoprotein A/B	25.72	1.10E-09	115.00	9.42E-14	48.04	7.91E-10
HNRNPC;HNRNPCL4;HNRNPCL1;HNRNPCL3;HNRNPCL2	P07910;P0DMR1;Q86SE5	Heterogeneous nuclear ribonucleoproteins C1/C2;Heterogeneous nuclear ribonucleoprotein C-like 4;Heterogeneous nuclear ribonucleoprotein C-like 1;Heterogeneous nuclear ribonucleoprotein C-like 3;Heterogeneous nuclear ribonucleoprotein C-like 2	75.16	3.65E-05	284.21	9.73E-13	175.76	1.58E-08
HNRNPD	Q14103	Heterogeneous nuclear ribonucleoprotein D0	26.80	4.07E-06	130.81	8.26E-10	52.33	2.06E-08
HNRNPDL	O14979	Heterogeneous nuclear ribonucleoprotein D-like	7.28	3.15E-03	30.68	4.17E-07	13.33	1.71E-04
HNRNPDL	A0A087WU03	Heterogeneous nuclear ribonucleoprotein D-like	1.80	5.77E-01	17.57	6.84E-06	5.84	4.46E-02
HNRNPF	P52597	Heterogeneous nuclear ribonucleoprotein F	16.60	4.75E-09	37.39	5.38E-12	13.33	8.80E-11
HNRNPH1	P31943	Heterogeneous nuclear ribonucleoprotein H	9.31	2.59E-05	21.75	2.60E-09	7.94	6.71E-07

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
HNRNPH3	P31942	Heterogeneous nuclear ribonucleoprotein H3	2.42	1.64E-01	4.61	5.19E-02	3.22	1.32E-01
HNRNPK	P61978	Heterogeneous nuclear ribonucleoprotein K	5.05	1.40E-06	16.65	7.26E-12	7.23	2.40E-10
HNRNPL	P14866	Heterogeneous nuclear ribonucleoprotein L	7.12	1.07E-04	28.08	2.28E-09	10.71	1.95E-09
HNRNPM	P52272	Heterogeneous nuclear ribonucleoprotein M	94.46	2.72E-11	223.39	1.03E-12	76.19	8.36E-10
HNRNPM	M0QZM1	Heterogeneous nuclear ribonucleoprotein M	704.21	1.94E-12	1260.47	7.23E-11	401.45	3.65E-09
HNRNPR	O43390	Heterogeneous nuclear ribonucleoprotein R	238.44	5.58E-06	1021.16	7.04E-12	369.77	6.63E-09
HNRNPU	Q00839	Heterogeneous nuclear ribonucleoprotein U	104.13	1.78E-07	396.64	3.48E-16	147.77	7.66E-09
HNRNPUL1	Q9BUJ2	Heterogeneous nuclear ribonucleoprotein U-like protein 1	14.23	2.78E-06	42.02	7.14E-11	21.74	2.62E-08
HNRNPUL2; HNRNPUL2-BSCL2	Q1KMD3;H3BQZ7	Heterogeneous nuclear ribonucleoprotein U-like protein 2	78.45	1.54E-06	275.37	8.09E-12	121.33	8.22E-09
HP1BP3	Q5SSJ5;	Heterochromatin protein 1-binding protein 3	70.15	1.14E-05	234.53	2.21E-13	71.86	7.52E-10
HSP90AA1	P07900	Heat shock protein HSP 90-alpha	3.12	1.01E-05	3.99	9.18E-08	1.98	5.54E-08
HSP90AB1	P08238	Heat shock protein HSP 90-beta	2.92	1.12E-05	4.06	2.25E-10	1.94	2.42E-06
HSP90B1	P14625	Endoplasmin	2.30	5.98E-03	2.00	3.88E-03	1.18	4.10E-01
HSPA1B;HSPA1A	P0DMV9;P0DMV8	Heat shock 70 kDa protein 1B;Heat shock 70 kDa protein 1A	9.90	3.88E-09	20.70	1.84E-10	8.43	1.43E-08
HSPA5	P11021	78 kDa glucose-regulated protein	3.63	1.74E-05	5.28	2.23E-08	2.50	4.40E-05
HSPA6;HSPA7	P17066;P48741	Heat shock 70 kDa protein 6;Putative heat shock 70 kDa protein 7	5.71	6.88E-07	9.27	7.11E-10	3.74	1.93E-05
HSPA8	P11142	Heat shock cognate 71 kDa protein	6.32	1.70E-09	12.17	1.84E-11	5.09	9.89E-08

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
HSPA9	P38646	Stress-70 protein, mitochondrial	5.00	7.39E-06	9.78	1.33E-09	3.40	3.74E-03
HSPB1	P04792	Heat shock protein beta-1	3.08	7.47E-03	1.91	3.42E-01	1.11	8.18E-01
HSPD1	P10809	60 kDa heat shock protein, mitochondrial	3.88	7.04E-03	5.15	4.20E-02	2.52	2.19E-02
ICT1	Q14197	Peptidyl-tRNA hydrolase ICT1, mitochondrial	3490079.59	2.74E-05	22213315.92	3.83E-09	9447895.76	5.45E-08
IGF2BP1	Q9NZI8	Insulin-like growth factor 2 mRNA-binding protein 1	300.86	7.29E-12	672.97	5.49E-07	297.61	7.45E-06
IGF2BP2	Q9Y6M1	Insulin-like growth factor 2 mRNA-binding protein 2	1500531.08	1.11E-06	4326720.31	1.33E-06	1769976.02	5.91E-06
IGF2BP3	O00425	Insulin-like growth factor 2 mRNA-binding protein 3	495.48	5.24E-08	1646.54	7.08E-12	719.50	1.35E-08
IGFBP4	P22692	Insulin-like growth factor-binding protein 4	3.29	4.34E-06	4.37	1.24E-09	2.08	1.01E-03
IGFBP7	Q16270	Insulin-like growth factor-binding protein 7	2.43	7.78E-05	2.94	1.58E-07	1.55	4.62E-02
IGHMBP2	P38935	DNA-binding protein SMUBP-2	5739951.89	1.79E-08	9992008.40	9.58E-09	3279732.96	1.68E-10
IGKV2D-28;IGKV2D-29;IGKV2-40;IGKV2D-40;IGKV2D-26;IGKV2D-30;IGKV2-30	P01617;P01614;P06309;P06310	Ig kappa chain V-II region TEW;Ig kappa chain V-II region Cum;Ig kappa chain V-II region GM607;Ig kappa chain V-II region RPMI 6410	1.02	9.76E-01	2.72	8.11E-02	1.02	9.61E-01
ILF2	Q12905	Interleukin enhancer-binding factor 2	359.32	1.31E-08	1214.69	4.55E-11	427.59	7.56E-09
ILF3	Q12906	Interleukin enhancer-binding factor 3	1990.45	8.09E-11	6587.53	1.39E-13	2315.21	4.56E-12
IMP3	Q9NV31	U3 small nucleolar ribonucleoprotein protein IMP3	32.17	1.66E-02	56.72	1.10E-03	20.81	2.52E-03
IMP4	Q96G21	U3 small nucleolar ribonucleoprotein protein IMP4	4070834.06	1.57E-07	10950277.77	6.72E-12	3125522.36	1.01E-05

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
ISG20L2	Q9H9L3	Interferon-stimulated 20 kDa exonuclease-like 2	11527547.81	4.18E-08	19228905.64	6.90E-12	6396093.10	2.17E-07
ISYNA1	Q9NPH2	Inositol-3-phosphate synthase 1	6.89	1.68E-02	12.05	1.13E-06	7.51	1.65E-05
ITGB1	P05556	Integrin beta-1	2.30	1.09E-03	1.72	4.30E-02	1.61	8.30E-03
ITGB5	P18084	Integrin beta;Integrin beta-5	1.62	5.98E-01	0.00	6.56E-02	0.00	6.56E-02
ITGB6	P18564	Integrin beta;Integrin beta-6	4.28	1.20E-02	4.18	3.19E-02	1.77	3.05E-01
JUP	P14923	Junction plakoglobin	0.05	1.95E-01	0.01	1.78E-01	0.01	1.78E-01
KHDRBS1;KHDRBS2	Q07666;Q5VWX1	KH domain-containing, RNA-binding, signal transduction-associated protein 1;KH domain-containing, RNA-binding, signal transduction-associated protein 2	1.55	5.01E-01	5.17	8.10E-03	3.33	6.25E-03
KHSRP	Q92945	Far upstream element-binding protein 2	2.75	3.49E-05	2.91	1.62E-09	1.73	5.63E-06
KIAA0020	Q15397	Pumilio domain-containing protein KIAA0020	14908491.71	1.48E-10	31461595.35	1.78E-12	7354469.29	3.32E-06
KIF14	Q15058	Kinesin-like protein KIF14	1542857.72	7.89E-07	2939731.02	1.25E-08	979929.50	1.69E-07
KIF23	Q02241	Kinesin-like protein	2842897.68	2.41E-07	5402431.84	1.16E-03	1471293.70	2.05E-08
KNOP1	Q1ED39	Lysine-rich nucleolar protein 1	2794773.78	4.03E-07	6564636.58	2.01E-08	2057802.32	6.71E-12
KPNA2	P52292	Importin subunit alpha-1	1111.02	2.49E-13	2619.30	5.29E-09	1158.80	3.49E-09
KPNB1	Q14974	Importin subunit beta-1	3.16	1.05E-05	3.84	6.14E-03	1.86	2.15E-03
KRI1	Q8N9T8	Protein KRI1 homolog	13197168.36	6.11E-07	43808662.20	2.63E-12	13222227.83	1.99E-08
KRR1	Q13601	KRR1 small subunit processome component homolog	321.72	2.06E-08	1075.55	3.04E-14	355.21	5.48E-09
LAMA1	P25391	Laminin subunit alpha-1	3.36	2.47E-03	2.75	2.99E-03	1.75	1.04E-02
LAMA5	O15230	Laminin subunit alpha-5	3.88	1.57E-05	4.19	1.94E-02	1.65	1.39E-01
LAP3	P28838	Cytosol aminopeptidase	2.32	4.89E-04	5.74	9.49E-02	0.00	3.25E-07
LARP1	Q6PKG0	La-related protein 1	155.01	7.24E-10	303.46	8.79E-09	207.17	1.15E-10

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
LARP4	Q71RC2	La-related protein 4	33.93	1.26E-07	63.32	8.71E-09	30.52	4.62E-10
LARP7	Q4G0J3	La-related protein 7	2139267.64	5.49E-02	36644045.38	1.26E-09	22825579.93	2.41E-14
LDHA;LDHC;LDHB;LDHAL6A	P00338;P07864;P07195	L-lactate dehydrogenase A chain;L-lactate dehydrogenase;L-lactate dehydrogenase A-like 6A;L-lactate dehydrogenase C chain;L-lactate dehydrogenase B chain	2.35	1.37E-01	0.70	7.09E-01	2.04	2.18E-01
LDLR	P01130	Low-density lipoprotein receptor	1.95	5.36E-04	2.60	2.47E-06	1.14	4.67E-01
LGALS1	P09382	Galectin-1	4.14	2.22E-05	5.10	3.06E-08	2.85	6.68E-04
LGALS7	P47929	Galectin-7	0.00	5.03E-02	0.00	5.03E-02	0.00	5.03E-02
LLPH	Q9BRT6	Protein LLP homolog	6382746.43	5.06E-07	15252579.12	6.91E-08	3656428.77	1.77E-07
LMNA	P02545	Prelamin-A/C;Lamin-A/C	4.22	1.28E-06	4.87	1.54E-06	2.29	8.43E-04
LRP1	Q07954	Prolow-density lipoprotein receptor-related protein 1	1.85	5.24E-02	1.53	2.51E-01	1.09	8.21E-01
LRP1B	Q9NZR2	Low-density lipoprotein receptor-related protein 1B	1.50	2.66E-01	0.39	2.30E-01	0.00	5.06E-03
LRP8	Q14114	Low-density lipoprotein receptor-related protein 8	2.27	1.83E-03	2.37	5.10E-06	1.24	1.45E-01
LRP8	E9PP15	Low-density lipoprotein receptor-related protein 8	0.00	1.24E-02	1.09	9.38E-01	0.00	1.24E-02
LRRC16A	Q5VZK9	Leucine-rich repeat-containing protein 16A	2943921.48	1.41E-03	9367209.44	1.18E-05	3979839.04	2.97E-07
LTBP3	Q9NS15	Latent-transforming growth factor beta-binding protein 3	3.37	3.43E-01	1.17	8.99E-01	0.00	1.51E-01
LTV1	Q96GA3	Protein LTV1 homolog	10454494.52	2.03E-09	25146922.77	9.77E-12	10119309.63	5.86E-08
LUC7L2;C7orf55-LUC7L2	Q9Y383	Putative RNA-binding protein Luc7-like 2	51.89	1.05E-09	132.63	2.37E-09	62.26	3.11E-07
LUC7L3	O95232	Luc7-like protein 3	41.62	5.03E-07	118.79	6.49E-07	54.31	3.11E-09

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
LYAR	Q9NX58	Cell growth-regulating nucleolar protein	404.15	2.85E-11	985.80	5.69E-11	277.32	3.88E-09
MAGEB2	P43366	Melanoma-associated antigen B2	32650900.72	1.74E-11	105269713.40	1.71E-09	31334388.40	3.53E-08
MAGT1	Q9H0U3	Magnesium transporter protein 1	1727579.46	5.03E-05	4346758.51	2.15E-05	1292343.10	8.90E-06
MAK16	Q9BXY0	Protein MAK16 homolog	132138.19	3.41E-01	232952.63	3.41E-01	515754.59	1.33E-07
MAP4	P27816	Microtubule-associated protein 4	97.72	4.49E-08	233.96	1.10E-11	89.66	4.14E-06
MAP7	Q14244	Ensconsin	110.06	4.71E-07	265.45	5.75E-17	108.61	2.62E-05
MAP7D1	Q3KQU3	MAP7 domain-containing protein 1	242.30	7.33E-08	482.08	1.61E-12	224.17	5.43E-07
MAT2A	P31153	S-adenosylmethionine synthase isoform type-2	2.30	1.54E-01	0.00	2.25E-05	0.00	2.25E-05
MATR3	P43243	Matrin-3	17.95	3.27E-07	51.08	2.57E-12	21.19	1.60E-10
MCM3	P25205	DNA replication licensing factor MCM3	0.00	1.45E-01	0.00	1.45E-01	0.28	3.21E-01
MCM7	P33993	DNA replication licensing factor MCM7	6.93	9.25E-06	9.40	1.22E-06	3.53	1.36E-02
MEPCE	Q7L2J0	7SK snRNA methylphosphate capping enzyme	121.75	7.02E-07	583.88	1.90E-14	309.98	8.05E-11
MGP	P08493	Matrix Gla protein	2.16	4.82E-02	2.94	8.34E-02	2.32	3.29E-02
MKI67	P46013	Antigen KI-67	233.60	3.14E-05	525.15	6.16E-07	154.89	9.21E-10
MMP3	P08254	Stromelysin-1	0.83	8.45E-01	0.69	7.12E-01	1.55	6.20E-01
MMTAG2	Q9BU76	Multiple myeloma tumor-associated protein 2	93.05	5.27E-10	336.25	2.01E-12	73.58	4.31E-07
MNT	Q99583	Max-binding protein MNT	416201.06	1.36E-03	1044563.37	1.11E-02	1191572.75	5.66E-12
MOV10	Q5JR04	Putative helicase MOV-10	1910256.04	8.42E-07	7420466.24	3.31E-06	3668062.38	1.72E-07
MPHOSPH10	O00566	U3 small nucleolar ribonucleoprotein protein MPP10	5501343.26	8.32E-09	15260352.32	4.29E-12	4443435.28	3.79E-12
MPHOSPH6	Q99547	M-phase phosphoprotein 6	1319431.20	1.94E-05	3488208.23	3.50E-08	721151.05	1.03E-05

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
MRPL1	Q9BYD6	39S ribosomal protein L1, mitochondrial	6763441.67	2.90E-05	43532713.49	1.23E-08	18235227.66	1.27E-09
MRPL10	Q7Z7H8	39S ribosomal protein L10, mitochondrial	468568.87	1.63E-01	8277684.14	1.10E-03	2169336.44	5.73E-04
MRPL11	Q9Y3B7	39S ribosomal protein L11, mitochondrial	47463967.27	7.93E-13	122522337.67	1.16E-13	42587643.23	1.79E-10
MRPL12	P52815	39S ribosomal protein L12, mitochondrial	8448120.93	2.69E-04	80239444.20	4.90E-11	25699020.71	9.64E-06
MRPL13	Q9BYD1	39S ribosomal protein L13, mitochondrial	8.74	9.36E-04	63.95	1.14E-08	26.66	5.07E-09
MRPL14	Q6P1L8	39S ribosomal protein L14, mitochondrial	12.80	6.07E-05	67.03	7.70E-12	34.32	4.04E-11
MRPL15	Q9P015	39S ribosomal protein L15, mitochondrial	498.97	1.43E-06	2723.10	1.98E-13	1071.61	9.85E-07
MRPL16	Q9NX20	39S ribosomal protein L16, mitochondrial	65.54	4.93E-04	471.36	3.34E-10	233.15	1.25E-08
MRPL17	Q9NRX2	39S ribosomal protein L17, mitochondrial	4828522.70	3.61E-04	31204413.37	4.03E-10	13849176.80	1.42E-08
MRPL18	Q9H0U6	39S ribosomal protein L18, mitochondrial	6993823.99	7.48E-06	41749318.60	2.60E-09	18231008.22	2.86E-09
MRPL19	P49406	39S ribosomal protein L19, mitochondrial	11.60	7.51E-02	104.14	4.90E-07	95.60	5.44E-07
MRPL2	Q5T653	39S ribosomal protein L2, mitochondrial	4070971.16	3.30E-04	35308942.22	4.91E-09	17132546.52	1.14E-09
MRPL21	Q7Z2W9	39S ribosomal protein L21, mitochondrial	1.00	N/A	154720.89	3.41E-01	343661.75	4.77E-03
MRPL22	Q9NWU5	39S ribosomal protein L22, mitochondrial	11274863.06	2.56E-05	59318423.29	1.97E-09	20685920.41	3.97E-09
MRPL23	Q16540	39S ribosomal protein L23, mitochondrial	4743357.56	1.85E-05	30020110.36	2.07E-09	12802152.13	5.34E-08
MRPL24	Q96A35	39S ribosomal protein L24, mitochondrial	10315949.79	1.91E-05	52906641.19	2.41E-10	24312833.18	6.88E-09

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
MRPL27	Q9P0M9	39S ribosomal protein L27, mitochondrial	2812926.50	4.17E-03	31746372.20	2.33E-09	10459477.12	2.40E-07
MRPL28	Q13084	39S ribosomal protein L28, mitochondrial	336.74	1.66E-04	2026.75	1.11E-09	883.54	1.06E-08
MRPL3	P09001	39S ribosomal protein L3, mitochondrial	101.94	2.01E-04	625.01	2.44E-12	290.80	9.08E-10
MRPL32	Q9BYC8	39S ribosomal protein L32, mitochondrial	684984.58	8.24E-02	7906124.87	3.26E-10	3813231.31	1.96E-07
MRPL33	O75394	39S ribosomal protein L33, mitochondrial	803886.21	1.24E-04	5400432.85	1.31E-08	2413662.12	4.97E-05
MRPL34	Q9BQ48	39S ribosomal protein L34, mitochondrial	839442.31	3.22E-03	6836198.99	2.06E-11	2849590.87	1.83E-06
MRPL35	Q9NZE8	39S ribosomal protein L35, mitochondrial	1772815.17	3.66E-03	14706289.51	2.40E-10	5239664.30	3.07E-08
MRPL37	Q9BZE1	39S ribosomal protein L37, mitochondrial	2118354.37	1.78E-02	19881299.47	3.41E-10	8139506.73	1.35E-06
MRPL38	Q96DV4	39S ribosomal protein L38, mitochondrial	13.78	5.65E-03	122.02	4.75E-08	57.99	1.06E-06
MRPL39	Q9NYK5	39S ribosomal protein L39, mitochondrial	4546511.16	6.52E-04	32909854.13	1.97E-09	14352737.20	4.37E-11
MRPL4	Q9BYD3	39S ribosomal protein L4, mitochondrial	5392116.18	1.10E-04	28748192.24	1.50E-12	12632309.50	1.78E-07
MRPL41	Q8IXM3	39S ribosomal protein L41, mitochondrial	2894248.12	5.96E-04	18813761.07	2.50E-10	7905367.82	9.14E-07
MRPL43	Q8N983	39S ribosomal protein L43, mitochondrial	4.77	2.51E-01	19.01	6.66E-02	14.41	1.41E-03
MRPL45	Q9BRJ2	39S ribosomal protein L45, mitochondrial	50.43	7.64E-03	531.59	3.23E-11	208.50	1.35E-07
MRPL46	Q9H2W6	39S ribosomal protein L46, mitochondrial	1029753.53	4.57E-02	13480899.56	5.53E-13	4301219.96	1.13E-05
MRPL47	Q9HD33	39S ribosomal protein L47, mitochondrial	2157.21	3.12E-06	11444.00	3.51E-12	4480.80	7.34E-08

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
MRPL48	Q96GC5	39S ribosomal protein L48, mitochondrial	1093703.94	2.49E-02	15432917.43	1.72E-07	5192969.12	2.55E-08
MRPL49	Q13405	39S ribosomal protein L49, mitochondrial	2145603.85	3.41E-06	12063678.07	3.17E-10	6338146.08	3.49E-09
MRPL50	Q8N5N7	39S ribosomal protein L50, mitochondrial	1109546.60	5.23E-02	13766120.59	1.61E-07	4773677.03	2.56E-07
MRPL51	Q4U2R6	39S ribosomal protein L51, mitochondrial	404396.65	5.65E-02	4284861.61	3.29E-08	1613971.32	4.41E-03
MRPL52	Q86TS9	39S ribosomal protein L52, mitochondrial	382281.59	5.27E-02	5078089.22	1.32E-06	2576030.31	2.69E-06
MRPL53	Q96EL3	39S ribosomal protein L53, mitochondrial	324944.39	1.47E-01	6492234.75	4.73E-09	402977.74	3.41E-01
MRPL54	Q6P161	39S ribosomal protein L54, mitochondrial	1677040.39	2.17E-03	8193789.07	8.45E-04	1154038.29	6.37E-02
MRPL55	Q7Z7F7	39S ribosomal protein L55, mitochondrial	829776.40	1.25E-03	5160074.31	4.58E-08	6684281.07	2.55E-08
MRPL57	Q9BQC6	Ribosomal protein 63, mitochondrial	2055411.69	9.88E-04	13359359.98	1.16E-11	5863619.47	1.04E-08
MRPL9	Q9BYD2	39S ribosomal protein L9, mitochondrial	448.93	4.98E-04	3524.44	1.61E-09	1147.41	1.25E-06
MRPS11	P82912	28S ribosomal protein S11, mitochondrial	5009181.56	2.98E-06	20021494.74	1.89E-10	7209260.16	1.15E-08
MRPS12	O15235	28S ribosomal protein S12, mitochondrial	32.88	1.14E-09	96.19	7.25E-10	27.45	1.42E-06
MRPS14	O60783	28S ribosomal protein S14, mitochondrial	9958493.87	1.62E-07	32319479.36	2.65E-10	13975338.15	4.92E-11
MRPS15	P82914	28S ribosomal protein S15, mitochondrial	12618196.25	1.61E-07	48089476.19	1.59E-11	19399793.96	1.24E-09
MRPS16	Q9Y3D3	28S ribosomal protein S16, mitochondrial	437.30	9.05E-06	1674.82	3.08E-09	596.75	9.97E-11
MRPS18A	Q9NVS2	28S ribosomal protein S18a, mitochondrial	6575584.79	5.13E-05	41287867.96	9.53E-09	20359014.04	1.97E-08

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
MRPS18B	Q9Y676	28S ribosomal protein S18b, mitochondrial	176.47	1.55E-06	634.88	4.31E-12	226.21	6.99E-09
MRPS18C	Q9Y3D5	28S ribosomal protein S18c, mitochondrial	414032.85	7.09E-02	2319868.38	3.56E-02	1000860.11	1.58E-03
MRPS2	Q9Y399	28S ribosomal protein S2, mitochondrial	1063.16	7.54E-07	4503.59	1.06E-11	1355.31	2.08E-07
MRPS21	P82921	28S ribosomal protein S21, mitochondrial	4341890.90	1.87E-06	16447209.39	9.08E-11	5593575.11	2.68E-04
MRPS22	P82650	28S ribosomal protein S22, mitochondrial	1357.63	1.36E-06	4987.36	1.18E-13	1793.17	5.34E-08
MRPS23	Q9Y3D9	28S ribosomal protein S23, mitochondrial	19374325.65	1.92E-07	68633496.74	3.36E-11	27646796.03	1.82E-07
MRPS25	P82663	28S ribosomal protein S25, mitochondrial	122.07	9.66E-09	430.10	3.80E-12	154.45	7.55E-10
MRPS26	Q9BYN8	28S ribosomal protein S26, mitochondrial	1837.33	6.65E-07	6593.44	1.32E-08	2350.28	5.02E-09
MRPS27	Q92552	28S ribosomal protein S27, mitochondrial	6390314.63	1.47E-03	34568933.29	3.54E-09	12642978.66	1.05E-07
MRPS28	Q9Y2Q9	28S ribosomal protein S28, mitochondrial	426.84	1.32E-05	2042.62	1.47E-10	686.61	3.89E-09
MRPS30	Q9NP92	28S ribosomal protein S30, mitochondrial	691590.76	7.23E-02	8721089.69	9.54E-09	3856229.45	2.01E-07
MRPS31	Q92665	28S ribosomal protein S31, mitochondrial	7748473.46	2.04E-06	26542990.05	1.53E-10	12363410.41	3.26E-07
MRPS33	Q9Y291	28S ribosomal protein S33, mitochondrial	2574583.11	1.17E-03	11421167.31	1.88E-09	5065200.37	1.73E-06
MRPS34	P82930	28S ribosomal protein S34, mitochondrial	471.48	6.72E-07	1826.27	1.15E-12	707.50	2.29E-09
MRPS35	P82673	28S ribosomal protein S35, mitochondrial	6275347.41	1.44E-06	24196815.26	6.87E-11	9505260.10	6.21E-08
MRPS5	P82675	28S ribosomal protein S5, mitochondrial	17862284.59	4.75E-07	76773150.84	1.55E-12	28415739.31	6.35E-10

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
MRPS6	P82932	28S ribosomal protein S6, mitochondrial	5521589.12	2.80E-03	25247197.09	7.62E-08	6826616.95	3.95E-03
MRPS7	Q9Y2R9	28S ribosomal protein S7, mitochondrial	19977816.72	2.70E-06	93805904.52	1.13E-12	32417579.73	3.98E-09
MRPS9	P82933	28S ribosomal protein S9, mitochondrial	16678313.26	3.60E-06	69068796.12	1.48E-11	24528426.63	1.52E-08
MRTO4	Q9UKD2	mRNA turnover protein 4 homolog	1761366.73	2.72E-07	4386362.98	3.92E-08	917956.09	1.02E-03
MSN	P26038	Moesin	0.00	3.41E-01	41.50	1.71E-02	20.13	2.36E-02
MT2A	P02795	Metallothionein-2	2.40	6.07E-02	1.45	1.82E-01	0.91	7.76E-01
MTDH	Q86UE4	Protein LYRIC	1026.03	4.81E-14	2507.03	1.32E-14	894.05	6.12E-12
MTMR4	Q9NYA4	Myotubularin-related protein 4	3.16	3.91E-04	1.36	6.00E-01	0.65	4.03E-01
MYBBP1A	Q9BQG0	Myb-binding protein 1A	1373.66	3.51E-11	3532.57	7.26E-13	1069.15	7.02E-10
MYL6;MYL6B	P60660;P14649	Myosin light polypeptide 6;Myosin light chain 6B	10.62	5.65E-11	17.96	1.30E-07	6.04	3.08E-06
NAP1L1	P55209	Nucleosome assembly protein 1-like 1	2051778.24	4.16E-08	4233291.96	2.00E-06	938725.97	1.19E-02
NAT10	Q9H0A0	N-acetyltransferase 10	25744229.90	1.42E-10	75187417.41	3.99E-13	20769505.57	6.20E-11
NCBP1	Q09161	Nuclear cap-binding protein subunit 1	1838.27	1.67E-07	7036.28	1.15E-10	3085.13	3.13E-07
NCBP2	P52298	Nuclear cap-binding protein subunit 2	9709076.79	2.24E-09	37412259.94	7.33E-10	16774094.78	3.21E-11
NCL	P19338	Nucleolin	373.58	1.06E-14	952.91	1.06E-11	330.88	8.30E-13
NEMF	O60524	Nuclear export mediator factor NEMF	590793.02	1.18E-03	1798585.94	1.49E-08	392987.38	5.83E-02
NHP2	Q9NX24	H/ACA ribonucleoprotein complex subunit 2	1349869.04	5.40E-02	5112824.50	1.21E-03	1618879.94	9.27E-04
NHP2L1	P55769	NHP2-like protein 1	18.38	1.39E-07	58.17	2.83E-09	21.56	2.42E-07
NIFK	Q9BYG3	MKI67 FHA domain-interacting nucleolar phosphoprotein	11820701.42	1.65E-12	32181893.66	1.15E-12	11666478.96	6.37E-09
NIP7	Q9Y221	60S ribosome subunit biogenesis protein NIP7 homolog	137.82	1.04E-06	406.68	4.52E-10	108.24	1.83E-06

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
NKRF	O15226	NF-kappa-B-repressing factor	17399783.29	4.38E-14	37353572.25	5.39E-13	12136464.65	3.50E-11
NOA1	Q8NC60	Nitric oxide-associated protein 1	1255921.82	1.12E-06	4182693.96	4.98E-08	1276787.42	1.61E-06
NOC3L	Q8WTT2	Nucleolar complex protein 3 homolog	9405343.23	7.00E-08	28194740.18	2.92E-12	7719310.47	6.95E-08
NOL10	Q9BSC4	Nucleolar protein 10	5520583.68	2.73E-11	11866849.46	1.89E-10	3739129.86	9.64E-08
NOL12	Q9UGY1	Nucleolar protein 12	1063101.74	7.62E-06	3042187.39	1.95E-07	987789.72	1.59E-06
NOL6	Q9H6R4	Nucleolar protein 6	6662486.59	2.99E-12	15662893.19	4.67E-14	4523765.44	2.52E-10
NOL9	Q5SY16	Polynucleotide 5-hydroxyl-kinase NOL9	731693.84	1.34E-02	2203647.56	3.62E-09	120839.66	3.41E-01
NOM1	Q5C9Z4	Nucleolar MIF4G domain-containing protein 1	3186071.82	1.86E-09	5502095.24	3.20E-08	811244.70	7.80E-02
NONO	Q15233	Non-POU domain-containing octamer-binding protein	3.71	4.33E-08	4.62	3.72E-09	1.93	2.65E-05
NOP16	Q9Y3C1	Nucleolar protein 16	35983012.35	1.58E-09	90084761.15	1.10E-10	26402663.08	7.10E-10
NOP2	P46087	Probable 28S rRNA (cytosine(4447)-C(5))-methyltransferase	6839.28	1.11E-13	14689.22	3.61E-15	4375.95	8.18E-13
NOP56	O00567	Nucleolar protein 56	1577.07	1.95E-05	5180.18	9.32E-13	1597.12	6.08E-09
NOP58	Q9Y2X3	Nucleolar protein 58	572.36	5.67E-07	2140.14	1.57E-12	680.98	3.31E-11
NOTCH1	P46531	Neurogenic locus notch homolog protein 1	1.56	7.89E-02	0.98	9.40E-01	0.46	1.10E-01
NOTCH2	Q04721	Neurogenic locus notch homolog protein 2	2.84	1.81E-06	3.74	4.52E-08	1.51	1.68E-02
NOTCH3	Q9UM47	Neurogenic locus notch homolog protein 3	3.26	7.19E-07	3.85	1.72E-07	0.70	4.56E-01
NPM1	P06748	Nucleophosmin	943.61	5.61E-11	2291.40	3.40E-13	607.88	1.76E-10
NSUN2	Q08J23	tRNA (cytosine(34)-C(5))-methyltransferase	17.24	4.66E-12	32.10	1.21E-07	9.81	7.00E-07
NUFIP2	Q7Z417	Nuclear fragile X mental retardation-interacting protein 2	23.80	1.39E-08	45.56	5.97E-11	23.30	1.38E-07

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
NUMA1	Q14980	Nuclear mitotic apparatus protein 1	1585.40	1.23E-12	4037.41	9.84E-13	966.07	4.49E-09
NUPL2	O15504	Nucleoporin-like protein 2	3254670.08	8.09E-08	8450637.75	4.84E-11	3444572.54	2.35E-08
NUSAP1	Q9BXS6	Nucleolar and spindle-associated protein 1	1011118.33	3.01E-07	1758659.44	2.03E-08	1008529.29	7.34E-08
NXF1	Q9UBU9	Nuclear RNA export factor 1	2167.62	9.25E-08	7355.69	7.93E-13	2058.37	1.01E-07
OLFML3	Q9NRN5	Olfactomedin-like protein 3	3.28	1.63E-06	3.93	1.16E-05	1.67	2.52E-01
PA2G4	Q9UQ80	Proliferation-associated protein 2G4	2.99	2.97E-04	3.13	1.37E-05	1.68	8.22E-02
PABPC1;PABPC3	P11940;Q9H361	Polyadenylate-binding protein 1; Polyadenylate-binding protein 3	19.95	3.47E-10	40.15	7.66E-10	21.39	6.30E-10
PABPC4	P0CB38	Polyadenylate-binding protein 4	57.66	4.35E-11	140.13	7.20E-15	72.76	9.11E-14
PABPC4	H0Y5F5	Polyadenylate-binding protein	18086908.49	5.97E-10	33966732.14	2.10E-10	19751213.30	9.99E-11
PABPN1	Q86U42	Polyadenylate-binding protein 2	10.83	7.15E-02	74.91	6.22E-09	49.90	1.61E-06
PAK1IP1	Q9NWT1	p21-activated protein kinase-interacting protein 1	4679443.72	7.33E-10	11671131.31	4.60E-10	4340018.74	4.14E-09
PCBP1	P57723	Poly(rC)-binding protein 1	7.91	2.39E-08	11.11	1.32E-08	4.74	5.53E-06
PCBP2;PCBP3	Q15366;P57721	Poly(rC)-binding protein 2;Poly(rC)-binding protein 3	9.10	1.14E-07	15.04	2.80E-09	7.28	7.76E-08
PCNA	P12004	Proliferating cell nuclear antigen	3.63	2.72E-04	1.94	3.22E-01	1.66	1.34E-01
PDCD11	Q14690	Protein RRP5 homolog	982.87	5.02E-09	2440.43	3.26E-11	622.78	7.63E-06
PDIA3	H7BZJ3	Protein disulfide-isomerase A3	1.59	3.24E-01	0.54	4.44E-01	0.46	1.51E-01
PEG10	Q86TG7	Retrotransposon-derived protein PEG10	20.15	6.12E-02	109.14	3.12E-08	55.51	2.00E-07
PES1	O00541	Pescadillo homolog	21.52	4.80E-10	56.90	8.15E-10	18.63	3.59E-09
PGAM5	Q96HS1	Serine/threonine-protein phosphatase PGAM5, mitochondrial	53.62	8.81E-09	108.07	1.21E-11	32.95	9.20E-09
PGK1	P00558	Phosphoglycerate kinase 1	4.35	1.86E-02	2.84	1.08E-01	3.18	3.14E-04

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
PHAX	Q9H814	Phosphorylated adapter RNA export protein	1675196.32	6.47E-04	8080855.37	1.63E-10	5044103.15	3.96E-08
PHF5A	Q7RTV0	PHD finger-like domain-containing protein 5A	93.18	1.00E-05	204.43	5.91E-04	40.99	1.54E-01
PHF6	Q8IWS0	PHD finger protein 6	512.31	2.03E-07	1629.95	4.55E-13	549.15	3.89E-09
PHGDH	Q5SZU1	D-3-phosphoglycerate dehydrogenase	3.32	5.37E-06	3.95	2.22E-10	2.15	6.40E-06
PINX1	Q96BK5	PIN2/TERF1-interacting telomerase inhibitor 1	1620948.56	1.38E-02	7106484.36	2.47E-08	2963194.52	8.55E-04
PIP	P12273	Prolactin-inducible protein	0.00	1.47E-01	0.00	1.47E-01	0.00	1.47E-01
PKM	P14618;P30613	Pyruvate kinase	4.02	4.19E-07	5.62	1.92E-09	2.51	1.20E-07
PLAUR	Q03405	Urokinase plasminogen activator surface receptor	3.28	6.60E-05	4.24	1.94E-06	1.91	1.59E-02
PLRG1	O43660	Pleiotropic regulator 1	3218017.48	9.37E-04	12579374.57	1.32E-11	4007265.42	3.37E-07
POLR1A	O95602	DNA-directed RNA polymerase I subunit RPA1	9762227.88	3.27E-11	13890713.93	9.76E-13	4846030.44	1.48E-09
POLR1B	Q9H9Y6	DNA-directed RNA polymerase I subunit RPA2	377.54	3.32E-11	540.62	2.01E-11	151.77	1.98E-12
POLR1C	O15160	DNA-directed RNA polymerases I and III subunit RPAC1	7056915.20	4.79E-09	12680259.64	9.20E-12	3321847.02	4.72E-07
POLR1D	Q9Y2S0	DNA-directed RNA polymerases I and III subunit RPAC2	1539082.76	5.33E-09	2613073.27	8.40E-09	683528.02	1.40E-02
POLR1E	E7EX70	DNA-directed RNA polymerase I subunit RPA49	14935547.13	1.77E-16	31675525.01	7.63E-12	8772362.92	3.81E-09
POLR1E	Q9GZS1	DNA-directed RNA polymerase I subunit RPA49	4570262.51	1.53E-03	11136890.17	5.32E-09	3121142.18	1.60E-10
POLR2E	P19388	DNA-directed RNA polymerases I, II, and III subunit RPABC1	17.87	9.30E-11	29.15	1.07E-08	9.71	1.09E-07
POLR2H	P52434	DNA-directed RNA polymerases I, II, and III subunit RPABC3	95.71	8.44E-08	184.79	8.59E-09	39.76	6.24E-02

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
POLR3B	Q9NW08	DNA-directed RNA polymerase III subunit RPC2	6432470.22	1.20E-02	18708545.21	2.76E-07	6552694.72	4.33E-07
POLRMT	O00411	DNA-directed RNA polymerase, mitochondrial	4260084.84	3.30E-10	10798866.88	1.17E-11	3394622.37	1.75E-08
POM121;POM121C	Q96HA1	Nuclear envelope pore membrane protein POM 121;Nuclear envelope pore membrane protein POM 121C	1976644.63	1.67E-09	4007702.50	1.40E-07	558936.07	5.55E-02
POP1	Q99575	Ribonucleases P/MRP protein subunit POP1	46283492.12	2.30E-10	149845146.56	9.61E-14	68289075.05	1.50E-14
POP4	O95707	Ribonuclease P protein subunit p29	243170.16	6.13E-02	1744578.84	1.58E-09	1175386.16	9.87E-07
POP7	O75817	Ribonuclease P protein subunit p20	1213967.65	2.81E-06	3605877.87	7.13E-10	2670525.74	8.14E-09
PPAN-P2RY11;PPAN	Q9NQ55;C9J3W3	Suppressor of SWI4 1 homolog	10792915.68	3.28E-08	27260678.78	1.90E-10	8713089.79	5.08E-09
PPIA	P62937	Peptidyl-prolyl cis-trans isomerase A	3.14	2.60E-05	4.12	3.45E-06	2.26	2.36E-05
PPP1CC	P36873	Serine/threonine-protein phosphatase	1365430.30	1.06E-02	3653128.01	6.93E-04	446044.03	1.48E-01
PRDX1	P32119	Peroxiredoxin-1	3.91	9.54E-06	4.58	9.39E-07	2.53	5.96E-04
PRKCDBP	Q969G5	Protein kinase C delta-binding protein	3176977.18	6.66E-06	13833247.75	1.46E-06	2978113.10	8.79E-05
PRKRA	O75569	Interferon-inducible double-stranded RNA-dependent protein kinase activator A	1239905.66	5.62E-09	3323777.23	8.57E-07	1230823.62	1.35E-05
PRPF19	Q9UMS4	Pre-mRNA-processing factor 19	893.52	1.65E-10	2753.41	4.81E-09	745.42	1.67E-06
PRPF3	O43395	U4/U6 small nuclear ribonucleoprotein Prp3	4382.23	1.01E-11	10159.38	1.54E-14	3549.73	9.06E-15
PRPF4	O43172	U4/U6 small nuclear ribonucleoprotein Prp4	341.62	7.39E-13	1011.63	1.04E-10	327.56	3.30E-10
PRPF40A	O75400	Pre-mRNA-processing factor 40 homolog A	46.27	3.12E-11	98.30	2.30E-09	29.36	3.02E-06
PRPF4B	Q13523	Serine/threonine-protein kinase PRP4 homolog	65.40	8.48E-04	177.81	1.94E-05	103.28	2.32E-08

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
PRRC2A	P48634	Protein PRRC2A	16734137.90	1.86E-08	32227811.09	6.30E-12	16074270.00	1.39E-08
PRRC2B	Q5JSZ5	Protein PRRC2B	5.44	9.17E-04	4.22	1.25E-01	2.06	8.38E-02
PRRC2C	Q9Y520	Protein PRRC2C	571.20	1.75E-10	1385.71	5.55E-11	647.73	1.62E-06
PSMC2	P35998	26S protease regulatory subunit 7	1.43	5.87E-01	0.00	9.80E-07	0.00	9.80E-07
PSMC5	P62195	26S protease regulatory subunit 8	1.52	5.21E-01	0.44	2.72E-01	0.00	9.16E-04
PSMD14	O00487	26S proteasome non-ATPase regulatory subunit 14	1.45	7.84E-01	0.00	1.67E-01	0.00	1.67E-01
PSMD2	Q13200	26S proteasome non-ATPase regulatory subunit 2	0.97	9.76E-01	2.38	2.46E-01	1.19	7.64E-01
PTBP1	P26599	Polypyrimidine tract-binding protein 1	11.62	6.98E-08	36.09	8.88E-08	17.74	5.25E-10
PTCD3	Q96EY7	Pentatricopeptide repeat domain-containing protein 3, mitochondrial	5784282.55	1.54E-03	20933180.42	9.50E-09	6068383.02	2.48E-06
PTRF	Q6NZI2	Polymerase I and transcript release factor	1602.69	3.49E-06	5600.42	3.17E-11	1571.16	9.27E-12
PUF60	Q9UHX1	Poly(U)-binding-splicing factor PUF60	27.62	1.51E-07	77.55	3.75E-10	38.61	1.95E-06
PURA	Q00577	Transcriptional activator protein Pur-alpha	4885855.49	1.09E-06	11833757.83	5.71E-07	6232217.97	1.58E-10
PURB	Q96QR8	Transcriptional activator protein Pur-beta	10210241.74	2.22E-09	26165267.90	3.15E-11	11280600.84	5.18E-08
PWP1	Q13610	Periodic tryptophan protein 1 homolog	5577139.55	2.98E-07	14675751.92	3.43E-07	4081306.60	5.49E-11
RAB14	P61106	Ras-related protein Rab-14	2.41	1.77E-01	0.00	1.45E-03	0.18	1.83E-02
RAB7A	P51149	Ras-related protein Rab-7a	0.50	4.84E-01	0.00	5.39E-02	0.00	5.39E-02
RACGAP1	Q9H0H5	Rac GTPase-activating protein 1	3511033.82	1.44E-08	6362755.39	2.53E-12	1829831.84	1.13E-08
RALY	Q5QPM0	RNA-binding protein Raly	1390083.76	4.81E-08	6685676.46	4.75E-10	3062471.82	1.05E-07
RAN	P62826	GTP-binding nuclear protein Ran	4.79	3.34E-07	7.93	1.08E-12	3.29	4.29E-07
RBBP4;RBBP7	Q09028;C9JAJ9	Histone-binding protein RBBP4	4.40	3.80E-02	7.07	2.60E-02	2.42	2.25E-01

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
RBBP6	Q7Z6E9	E3 ubiquitin-protein ligase RBBP6	2044443.12	2.31E-11	3617204.54	9.27E-12	854815.31	1.36E-02
RBM14	Q96PK6	RNA-binding protein 14	7.17	7.70E-10	11.78	8.96E-11	4.31	8.07E-08
RBM15	Q96T37	Putative RNA-binding protein 15	544339.68	1.21E-02	1426794.14	5.22E-02	1.00	N/A
RBM17	Q96I25	Splicing factor 45	733.50	1.12E-08	1934.41	3.29E-11	716.14	6.74E-11
RBM19	Q9Y4C8	Probable RNA-binding protein 19	1315135.57	1.12E-02	5912409.32	3.09E-10	1735175.66	4.41E-08
RBM25	P49756	RNA-binding protein 25	35.40	3.70E-10	74.90	1.69E-11	26.35	9.64E-11
RBM27	Q9P2N5	RNA-binding protein 27	658.92	2.13E-08	1141.21	3.38E-09	565.83	3.23E-07
RBM28	Q9NW13	RNA-binding protein 28	57443392.52	3.36E-11	115210529.70	4.27E-17	33257650.76	1.65E-10
RBM34	Q5TCT4	RNA-binding protein 34	5995675.80	1.85E-09	11413505.15	5.67E-11	2824414.86	1.23E-04
RBM39	Q14498	RNA-binding protein 39	28.51	7.01E-11	68.92	2.25E-11	29.30	1.64E-10
RBM42	Q9BTD8	RNA-binding protein 42	4468598.89	6.61E-06	19195369.82	1.52E-15	6026228.77	5.70E-07
RBM7	Q9Y580	RNA-binding protein 7	243522.39	3.41E-01	714530.15	1.50E-01	628586.85	1.28E-02
RBMX;RBMXL1	P38159;Q96E39	RNA-binding motif protein, X chromosome; RNA binding motif protein, X-linked-like-1	219.44	5.22E-12	1585.01	4.46E-13	776.64	1.71E-11
RBMX2	Q9Y388	RNA-binding motif protein, X-linked 2	1746947.84	3.71E-09	3993222.13	1.44E-08	1265738.67	8.15E-04
REPIN1	Q9BWE0	Replication initiator 1	2383846.86	1.09E-05	3744301.96	1.67E-09	2295698.88	1.80E-02
REXO4	Q9GZR2	RNA exonuclease 4	22343108.69	1.33E-08	51936656.49	4.45E-15	15389916.51	2.17E-08
RFC1	P35251	Replication factor C subunit 1	648.98	3.35E-14	1546.42	1.49E-10	466.88	2.65E-10
RFC2	P35250	Replication factor C subunit 2	4206202.52	5.86E-07	12717348.93	1.15E-09	2990078.24	5.04E-06
RFC4	P35249	Replication factor C subunit 4	113.22	2.74E-06	313.00	1.92E-10	91.21	1.23E-07
RFC5	P40937	Replication factor C subunit 5	1158703.02	1.18E-03	1384346.90	6.89E-02	591487.90	2.70E-02
RNMTL1	Q9HC36	rRNA methyltransferase 3, mitochondrial	4175925.14	7.36E-09	12716516.08	4.34E-06	2966811.02	1.11E-06
RPF2	Q9H7B2	Ribosome production factor 2 homolog	12481913.75	1.72E-08	42921406.59	3.81E-11	13838166.06	9.57E-12

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
RPL10	P27635	60S ribosomal protein L10	185.59	5.47E-12	453.42	1.14E-14	167.48	8.10E-09
RPL10A	P62906	60S ribosomal protein L10a	364.81	7.58E-09	812.06	1.65E-13	265.52	3.49E-10
RPL11	P62913	60S ribosomal protein L11	33.36	1.53E-12	99.21	1.71E-14	31.00	1.02E-08
RPL12	P30050	60S ribosomal protein L12	56.64	5.02E-11	113.08	8.62E-11	35.84	5.40E-08
RPL13	P26373	60S ribosomal protein L13	464.22	8.43E-10	981.77	1.42E-14	358.81	2.10E-08
RPL13A	P40429	60S ribosomal protein L13a	753.02	6.96E-13	1670.90	9.18E-15	615.47	3.59E-08
RPL14	P50914	60S ribosomal protein L14	944.50	2.40E-11	1950.92	1.48E-13	662.78	1.81E-06
RPL15	P61313	60S ribosomal protein L15	477.95	5.21E-11	1083.22	2.91E-11	338.64	7.66E-10
RPL17	P18621	60S ribosomal protein L17	386.29	6.43E-12	940.61	8.63E-14	332.40	5.52E-11
RPL18	Q07020	60S ribosomal protein L18	371.92	3.56E-08	698.99	2.81E-13	236.63	5.36E-10
RPL18A	Q02543	60S ribosomal protein L18a	546.92	5.45E-11	1188.16	6.34E-12	380.34	3.09E-11
RPL19	P84098	60S ribosomal protein L19	210.99	1.22E-09	587.19	6.41E-13	238.41	1.64E-09
RPL21	P46778	60S ribosomal protein L21	804.26	1.03E-10	1759.98	1.03E-12	664.08	1.12E-06
RPL22	P35268	60S ribosomal protein L22	86.75	3.08E-07	396.31	2.33E-07	111.57	1.32E-10
RPL23	P62829	60S ribosomal protein L23	58.32	1.20E-13	165.99	1.69E-10	58.76	2.17E-09
RPL23A	P62750	60S ribosomal protein L23a	272.77	1.64E-12	812.30	1.26E-11	298.94	1.03E-08
RPL24	P83731	60S ribosomal protein L24	282.23	3.43E-09	588.35	5.37E-15	259.31	1.79E-10
RPL26	P61254	60S ribosomal protein L26	175.48	2.77E-12	451.07	3.17E-12	187.86	6.82E-09
RPL26L1	Q9UNX3	60S ribosomal protein L26-like 1	226.18	1.97E-08	679.54	6.15E-10	219.18	3.64E-10
RPL27	P61353	60S ribosomal protein L27	2099.40	1.04E-13	4585.31	2.03E-12	1367.92	8.47E-09
RPL27A	P46776	60S ribosomal protein L27a	254.37	5.46E-10	606.99	1.21E-14	215.90	7.66E-09
RPL28	P46779	60S ribosomal protein L28	338.54	5.78E-10	639.99	4.84E-13	239.42	8.10E-09
RPL29	P47914	60S ribosomal protein L29	420.14	1.24E-08	792.41	1.35E-14	319.65	4.14E-09
RPL3	P39023	60S ribosomal protein L3	464.67	3.25E-13	1042.76	1.50E-13	350.88	1.51E-08

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
RPL30	P62888	60S ribosomal protein L30	305.09	1.09E-08	637.13	7.00E-14	182.64	7.99E-11
RPL31	P62899	60S ribosomal protein L31	259.20	3.59E-11	736.90	1.45E-14	322.97	2.57E-09
RPL32	P62910	60S ribosomal protein L32	606.69	1.34E-12	1533.21	5.19E-13	435.23	7.74E-07
RPL34	P49207	60S ribosomal protein L34	420.44	7.95E-10	798.74	2.06E-13	273.87	8.30E-09
RPL35	P42766	60S ribosomal protein L35	319.71	1.06E-09	859.22	3.39E-13	331.90	2.06E-08
RPL35A	P18077	60S ribosomal protein L35a	768.86	5.64E-09	1681.73	8.14E-13	606.02	6.86E-10
RPL36	Q9Y3U8	60S ribosomal protein L36	1870.19	5.71E-10	3978.09	1.13E-10	1408.09	6.28E-07
RPL36A;RPL36A-HNRNPH2	J3KQN4;P83881	60S ribosomal protein L36a	254.08	3.49E-10	657.61	1.86E-14	217.55	5.62E-10
RPL36AL	Q969Q0	60S ribosomal protein L36a-like	328.21	2.54E-10	873.70	2.14E-11	306.22	4.44E-10
RPL37	P61927	60S ribosomal protein L37	198.47	2.27E-07	431.68	8.94E-08	197.90	1.86E-11
RPL37A	P61513	60S ribosomal protein L37a	3009.23	3.99E-09	7033.06	1.50E-15	2303.99	9.22E-10
RPL38	P63173	60S ribosomal protein L38	9.33	1.85E-08	18.63	2.64E-07	8.83	3.17E-09
RPL39P5;RPL39	Q59GN2;P62891	60S ribosomal protein L39	560.81	1.73E-07	1276.89	1.47E-09	604.28	1.77E-07
RPL4	P36578	60S ribosomal protein L4	678.42	9.68E-14	1413.07	1.43E-14	458.36	5.72E-09
RPL5	P46777	60S ribosomal protein L5	458.01	3.81E-13	1957.36	1.29E-14	436.25	1.81E-07
RPL6	Q02878	60S ribosomal protein L6	1314.50	9.30E-09	2710.91	2.56E-19	898.99	1.62E-08
RPL7	P18124	60S ribosomal protein L7	620.36	6.96E-10	1308.88	1.27E-19	449.74	3.14E-12
RPL7A	P62424	60S ribosomal protein L7a	612.10	1.75E-11	1201.97	5.71E-16	426.88	8.10E-10
RPL8	P62917	60S ribosomal protein L8	451.95	1.30E-10	877.18	1.81E-18	310.72	1.01E-08
RPL9	P32969	60S ribosomal protein L9	71.77	7.32E-09	266.38	3.29E-06	75.35	5.66E-07
RPLP0;RPLP0P6	P05388;Q8NHW5	60S acidic ribosomal protein P0;60S acidic ribosomal protein P0-like	114.12	3.23E-10	211.85	3.58E-07	79.02	2.37E-08
RPLP1	P05386	60S acidic ribosomal protein P1	18987689.19	2.55E-10	36454046.09	3.58E-11	14205036.60	1.80E-08

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
RPLP2	P05387	60S acidic ribosomal protein P2	432.52	1.40E-09	786.55	3.16E-08	281.79	1.98E-07
RPN1	P04843	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	714.25	3.84E-09	2061.21	1.55E-11	562.99	6.68E-09
RPP25	Q9BUL9	Ribonuclease P protein subunit p25	1205567.68	1.49E-07	2442240.47	2.26E-11	1546064.41	9.68E-04
RPP30	P78346	Ribonuclease P protein subunit p30	1338493.38	3.28E-14	3520217.17	6.32E-08	2811676.16	9.30E-08
RPP38	P78345	Ribonuclease P protein subunit p38	40.35	7.92E-09	101.02	7.42E-04	122.77	8.20E-06
RPP40	O75818	Ribonuclease P protein subunit p40	1.00	N/A	480379.07	1.68E-01	1.00	N/A
RPS10;RPS10-NUDT3	P46783	40S ribosomal protein S10	67.03	1.12E-12	132.23	2.21E-14	78.50	8.23E-08
RPS11	P62280	40S ribosomal protein S11	92.07	7.50E-12	245.35	1.83E-12	83.96	5.11E-09
RPS12	P25398	40S ribosomal protein S12	349.64	3.89E-06	1255.17	1.08E-07	437.50	5.61E-07
RPS13	P62277	40S ribosomal protein S13	227.67	9.78E-14	679.19	3.20E-11	245.01	1.29E-12
RPS14	P62263	40S ribosomal protein S14	143.12	2.83E-11	460.03	7.58E-11	157.33	7.47E-10
RPS15	P62841	40S ribosomal protein S15	224.01	2.43E-12	685.52	1.91E-14	241.61	4.02E-07
RPS15A	P62244	40S ribosomal protein S15a	18.18	1.51E-11	57.71	1.01E-12	14.94	2.12E-12
RPS16	P62249;	40S ribosomal protein S16	20.15	6.31E-10	44.28	1.26E-12	19.33	3.97E-08
RPS17L;RPS17	P0CW22;P08708	40S ribosomal protein S17-like;40S ribosomal protein S17	68.56	1.22E-10	190.11	1.81E-08	54.94	3.89E-05
RPS18	P62269	40S ribosomal protein S18	12.30	6.36E-09	24.08	2.44E-10	11.90	9.35E-07
RPS19	P39019	40S ribosomal protein S19	160.74	3.20E-12	372.40	1.10E-09	190.35	3.60E-07
RPS19BP1	Q86WX3	Active regulator of SIRT1	1843718.92	1.66E-06	4760920.49	7.82E-04	1353390.40	1.20E-08
RPS2	P15880	40S ribosomal protein S2	178.46	5.09E-12	496.60	4.82E-13	178.26	6.98E-16
RPS20	P60866	40S ribosomal protein S20	19.28	2.56E-13	48.24	1.59E-07	17.30	1.25E-07
RPS23	P62266	40S ribosomal protein S23	226.80	1.91E-09	592.36	1.12E-13	192.19	2.94E-11
RPS24	P62847	40S ribosomal protein S24	214.94	6.11E-11	533.19	3.68E-12	233.46	1.04E-08
RPS25	P62851	40S ribosomal protein S25	72.54	2.58E-13	171.71	3.97E-09	64.27	1.34E-10

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
RPS26;RPS26P11	P62854	40S ribosomal protein S26;Putative 40S ribosomal protein S26-like 1	177.70	1.03E-08	443.30	2.52E-12	159.34	8.35E-11
RPS27	P42677	40S ribosomal protein S27	25.75	1.87E-09	62.94	1.21E-10	24.00	6.76E-08
RPS27A	P62979	40S ribosomal protein S27a	221.24	4.54E-10	704.31	1.26E-10	273.37	5.13E-12
RPS27L	Q71UM5	40S ribosomal protein S27-like	47.93	5.31E-05	95.75	9.55E-04	37.68	1.45E-03
RPS28	P62857	40S ribosomal protein S28	19.17	3.44E-11	37.26	2.87E-10	17.81	1.42E-06
RPS29	P62273	40S ribosomal protein S29	175.91	2.40E-06	456.19	7.06E-08	177.30	3.35E-05
RPS3	P23396	40S ribosomal protein S3	11.75	3.88E-12	25.46	1.56E-11	10.16	2.77E-11
RPS3A	P61247	40S ribosomal protein S3a	214.07	2.16E-11	629.81	4.08E-13	239.83	1.41E-09
RPS4X	P62701;	40S ribosomal protein S4, X isoform	138.71	1.65E-11	406.07	1.85E-12	112.10	1.60E-11
RPS5	P46782	40S ribosomal protein S5	53.54	3.54E-13	141.80	1.14E-08	51.85	2.51E-07
RPS6	P62753	40S ribosomal protein S6	325.12	2.01E-13	834.97	1.93E-16	319.57	8.45E-10
RPS7	P62081	40S ribosomal protein S7	394.24	3.30E-08	1255.27	5.58E-11	412.25	5.51E-06
RPS8	P62241	40S ribosomal protein S8	286.76	1.77E-12	725.89	8.87E-15	266.01	1.21E-11
RPS9	P46781	40S ribosomal protein S9	196.09	7.18E-10	546.65	2.87E-16	191.94	2.53E-08
RPSA;RPSAP58	C9J9K3;P08865	40S ribosomal protein SA	4.41	1.42E-07	6.86	2.12E-09	2.41	2.84E-07
RRBP1	Q9P2E9	Ribosome-binding protein 1	1940.34	5.75E-11	4350.98	3.14E-10	1348.23	3.65E-08
RRP1	P56182	Ribosomal RNA processing protein 1 homolog A	11750617.85	3.48E-11	30014946.73	3.00E-10	11921855.73	1.25E-11
RRP12	Q5JTH9	RRP12-like protein	44077013.59	2.94E-12	119691226.06	7.17E-15	39283618.33	5.41E-14
RRP15	Q9Y3B9	RRP15-like protein	12169382.61	3.68E-11	34035191.85	7.12E-12	10322686.79	3.85E-08
RRP1B	Q14684	Ribosomal RNA processing protein 1 homolog B	6420.49	8.29E-14	14374.53	1.20E-12	4633.29	1.52E-13
RRP7BP;RRP7A	Q9NSQ0;Q9Y3A4	Putative ribosomal RNA-processing protein 7 homolog B;Ribosomal RNA-processing protein 7 homolog A	660897.16	1.49E-03	2328124.30	3.04E-10	696700.72	2.39E-02

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
RRP8	O43159	Ribosomal RNA-processing protein 8	6628507.37	8.37E-07	17334298.43	6.53E-11	3569809.63	3.52E-04
RRP9	O43818	U3 small nucleolar RNA-interacting protein 2	1.28	8.65E-01	40.05	1.21E-10	35.97	2.52E-09
RRS1	Q15050	Ribosome biogenesis regulatory protein homolog	14047037.34	4.09E-09	43088530.45	1.89E-10	16249075.50	1.30E-12
RSBN1	Q5VWQ0	Round spermatid basic protein 1	15241063.05	4.55E-10	35315049.74	2.19E-13	13310896.41	1.80E-12
RSBN1L	Q6PCB5	Round spermatid basic protein 1-like protein	4823244.15	1.37E-14	11682680.07	4.40E-10	3436013.10	1.12E-09
RSL1D1	O76021	Ribosomal L1 domain-containing protein 1	1514.87	2.13E-09	3633.23	1.82E-13	1053.49	5.94E-07
RTCB	Q9Y3I0	tRNA-splicing ligase RtcB homolog	16.31	4.76E-09	42.19	1.74E-09	13.34	7.42E-08
RUVBL1	Q9Y265	RuvB-like 1	4.37	1.13E-06	5.97	4.23E-05	1.95	2.58E-03
RUVBL2	Q9Y230	RuvB-like 2	3.96	1.05E-08	6.12	4.99E-05	2.19	1.80E-05
SAFB	Q15424	Scaffold attachment factor B1	79542.37	3.41E-01	3426294.67	9.49E-09	1796779.53	2.37E-10
SARS	P49591	Serine--tRNA ligase, cytoplasmic	3.29	4.60E-04	4.58	6.33E-02	0.87	8.23E-01
SART1	O43290	U4/U6.U5 tri-snRNP-associated protein 1	5549992.73	6.57E-09	12937996.85	7.98E-11	4169231.76	7.45E-11
SART3	Q15020	Squamous cell carcinoma antigen recognized by T-cells 3	4474814.32	1.42E-08	14389142.82	1.55E-12	5123910.91	1.18E-07
SCFD1	Q8WVM8	Sec1 family domain-containing protein 1	2.95	1.91E-04	3.71	1.93E-06	1.53	1.66E-01
SEC23A	Q15436	Protein transport protein Sec23A	3.65	7.15E-03	5.00	3.72E-02	2.47	1.04E-03
SEC23B	Q15437	Protein transport protein Sec23B	2.44	3.40E-05	2.66	1.51E-02	1.71	2.63E-03
SEC24C	P53992	Protein transport protein Sec24C	2.55	2.35E-03	0.47	1.63E-01	1.41	2.54E-01
SEC24D	O94855	Protein transport protein Sec24D	2.00	6.77E-03	1.99	5.49E-02	1.05	8.93E-01
SEC31A	O94979	Protein transport protein Sec31A	0.80	8.14E-01	0.00	3.07E-03	0.00	3.07E-03
SERBP1	Q8NC51	Plasminogen activator inhibitor 1 RNA-binding protein	22.97	7.17E-09	49.02	5.27E-09	17.21	2.59E-06

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
SF3B1	O75533	Splicing factor 3B subunit 1	12.70	2.61E-10	32.65	7.12E-11	13.58	1.77E-08
SF3B2	Q13435	Splicing factor 3B subunit 2	182.06	5.73E-07	447.07	1.57E-08	168.31	7.69E-09
SF3B3	Q15393	Splicing factor 3B subunit 3	2.62	9.34E-04	3.80	1.32E-02	1.74	2.73E-01
SF3B4	Q15427	Splicing factor 3B subunit 4	13.76	7.13E-02	37.28	5.54E-02	22.93	1.14E-05
SF3B6	Q9Y3B4	Splicing factor 3B subunit 6	15.63	1.20E-08	36.30	6.14E-07	7.41	9.11E-02
SFPQ	P23246	Splicing factor, proline- and glutamine-rich	3.66	1.48E-06	4.66	2.60E-11	2.26	2.08E-06
SHMT1	P34896	Serine hydroxymethyltransferase, cytosolic	2.41	8.09E-04	2.75	3.75E-05	1.35	1.43E-02
SKIV2L2	P42285	Superkiller viralicidic activity 2-like 2	2141445.00	8.81E-08	4199911.60	2.57E-09	1596110.86	8.52E-06
SLC25A3	Q00325	Phosphate carrier protein, mitochondrial	1.28	7.47E-01	0.00	7.26E-05	0.00	7.26E-05
SLC25A5;SLC25A4	P05141;P12235	ADP/ATP translocase 2;ADP/ATP translocase 2, N-terminally processed;ADP/ATP translocase 1	5.04	1.49E-08	8.52	9.74E-12	3.13	6.56E-08
SLC25A6	P12236	ADP/ATP translocase 3	8.32	5.79E-06	9.91	2.09E-07	6.26	1.46E-05
SLPI	P03973	Antileukoproteinase	2.71	1.06E-01	2.29	4.08E-01	1.60	3.97E-01
SLTM	Q9NWH9	SAFB-like transcription modulator	4033198.70	4.67E-08	16206850.84	1.34E-10	7463089.81	1.60E-09
SMARCA4	P51532;P51531	Transcription activator BRG1	5.86	2.39E-07	7.94	1.16E-07	3.20	8.05E-05
SMARCA5	O60264	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	9796207.10	2.42E-13	25141037.34	7.22E-11	8362855.92	2.37E-11
SMARCC2;SMARCC1	Q8TAQ2;Q92922	SWI/SNF complex subunit SMARCC2;SWI/SNF complex subunit SMARCC1	4.17	6.06E-07	7.47	3.21E-03	2.43	1.30E-04
SMARCE1	Q969G3	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1	3.75	1.05E-06	4.90	2.96E-07	2.13	1.30E-04

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
SND1	Q7KZF4	Staphylococcal nuclease domain-containing protein 1	2.94	2.87E-05	3.99	5.32E-08	1.96	1.58E-05
SNRNP70	P08621	U1 small nuclear ribonucleoprotein 70 kDa	84.62	2.08E-09	290.58	3.84E-13	120.28	4.94E-11
SNRPA	P09012;P08579	U1 small nuclear ribonucleoprotein A	5264082.84	8.96E-11	19638942.10	2.99E-10	8650662.12	1.95E-10
SNRPC	P09234	U1 small nuclear ribonucleoprotein C	5705972.67	6.10E-06	17694308.78	2.60E-10	5277157.06	3.23E-07
SNRPD2	P62316	Small nuclear ribonucleoprotein Sm D2	18.22	1.56E-11	49.19	1.82E-09	18.67	7.85E-12
SNRPD3	P62318	Small nuclear ribonucleoprotein Sm D3	22.99	2.83E-08	56.00	1.40E-08	24.84	2.93E-09
SNRPE	P62304	Small nuclear ribonucleoprotein E	3.18	9.78E-03	3.44	7.92E-03	2.74	1.99E-04
SNRPF	P62306	Small nuclear ribonucleoprotein F	5.52	4.98E-03	13.88	7.21E-08	5.74	4.20E-06
SNRPG	Q49AN9	Small nuclear ribonucleoprotein G;Putative small nuclear ribonucleoprotein G-like protein 15	8.06	3.70E-03	10.72	9.82E-02	6.92	3.71E-03
SP110	Q9HB58	Sp110 nuclear body protein	1175989.53	7.56E-08	3064779.65	1.27E-07	966877.52	1.08E-03
SPATS2	Q86XZ4	Spermatogenesis-associated serine-rich protein 2	8758549.79	1.80E-09	18667957.99	1.23E-11	6983861.53	1.22E-10
SPATS2L	Q9NUQ6	SPATS2-like protein	9581477.62	2.54E-07	18644138.67	1.23E-14	7004295.69	1.40E-06
SPTY2D1	Q68D10	Protein SPT2 homolog	568.58	1.97E-09	846.29	4.19E-12	299.33	4.41E-11
SRFBP1	Q8NEF9	Serum response factor-binding protein 1	10704871.34	1.29E-07	27153797.24	1.77E-13	11828465.39	2.46E-11
SRP14	P37108	Signal recognition particle 14 kDa protein	1604.34	1.58E-09	4430.29	2.84E-13	1537.14	5.96E-12
SRP68	Q9UHB9	Signal recognition particle subunit SRP68	1335065.92	1.14E-03	5056968.32	4.55E-09	1372625.04	1.43E-04
SRP72	O76094	Signal recognition particle subunit SRP72	2157234.94	6.56E-09	8036992.30	2.92E-08	1160467.75	1.14E-02
SRP9	P49458	Signal recognition particle 9 kDa protein	33439483.72	1.46E-09	100432560.16	1.47E-11	32570082.46	6.76E-11

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
SRPK1	Q96SB4	SRSF protein kinase 1	718.79	2.24E-08	1596.87	2.46E-14	841.64	2.45E-12
SRPK2	P78362	SRSF protein kinase 2	1183.55	3.95E-10	1950.17	3.02E-08	974.98	1.53E-11
SRRM1	Q8IYB3	Serine/arginine repetitive matrix protein 1	1497.68	2.62E-08	3229.70	1.59E-12	1531.38	1.10E-11
SRRM2	Q9UQ35	Serine/arginine repetitive matrix protein 2	24.65	2.81E-09	49.38	2.44E-15	22.14	9.02E-14
SRSF1	Q07955	Serine/arginine-rich splicing factor 1	4313.31	1.60E-08	31890.95	7.59E-11	47.78	1.17E-02
SRSF10	Q8WXF0	Serine/arginine-rich splicing factor 10	584474.92	1.43E-02	10401100.97	6.55E-08	4287717.74	1.18E-09
SRSF11	Q05519	Serine/arginine-rich splicing factor 11	278072.55	5.41E-02	1966376.08	8.77E-08	687182.86	1.06E-02
SRSF2	Q01130	Serine/arginine-rich splicing factor 2	0.61	5.46E-01	1.07	9.13E-01	1788.83	1.01E-10
SRSF3	P84103	Serine/arginine-rich splicing factor 3	682.37	3.06E-07	3280.19	3.45E-14	905.57	5.31E-09
SRSF4	Q08170	Serine/arginine-rich splicing factor 4	3883068.59	5.75E-10	10741012.04	3.75E-08	5148243.03	2.19E-09
SRSF5	Q13243	Serine/arginine-rich splicing factor 5	8179006.41	2.84E-07	20715697.30	2.34E-08	8669549.15	4.85E-09
SRSF6	Q13247	Serine/arginine-rich splicing factor 6	537.07	1.01E-08	1395.31	7.49E-19	575.44	5.34E-12
SRSF7	Q16629	Serine/arginine-rich splicing factor 7	21122141.82	1.53E-06	99865116.33	8.49E-10	36022591.57	9.16E-07
SRSF9	Q13242	Serine/arginine-rich splicing factor 9	1589.55	1.81E-04	5018.78	2.48E-11	895.74	1.20E-04
SSB	P05455	Lupus La protein	210.20	1.41E-05	1011.94	3.55E-12	394.41	4.94E-08
SSR4	P51571	Translocon-associated protein subunit delta	3.88	4.74E-04	1.01	9.90E-01	2.26	2.19E-01
STAU1	O95793	Double-stranded RNA-binding protein Staufen homolog 1	1188852.61	1.33E-02	5560338.74	5.66E-07	1077324.63	1.11E-02
STAU2	Q9NUL3	Double-stranded RNA-binding protein Staufen homolog 2	12062303.51	1.27E-09	28142604.20	8.63E-11	10480574.33	1.49E-08
STT3B	Q8TCJ2	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B	32528877.16	1.29E-09	88483481.31	4.00E-11	32990017.51	1.36E-13
SUGP2	Q8IX01	SURP and G-patch domain-containing protein 2	508744.57	1.56E-07	1398204.91	3.34E-09	823561.12	1.65E-03

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
SUPT16H	Q9Y5B9	FACT complex subunit SPT16	1902265.43	1.46E-02	6083863.97	2.31E-07	528779.10	6.73E-02
SURF6	O75683	Surfeit locus protein 6	4434.71	9.01E-12	8570.25	1.24E-12	3100.23	1.50E-09
SUZ12	Q15022	Polycomb protein SUZ12	1327429.17	6.96E-04	2463489.08	4.72E-09	779367.48	6.60E-04
SYNCRIP	O60506	Heterogeneous nuclear ribonucleoprotein Q	118.70	1.46E-07	512.33	3.41E-17	170.27	5.67E-09
TAF15	Q92804	TATA-binding protein-associated factor 2N	2.78	4.00E-02	5.29	1.14E-02	3.67	1.43E-01
TAGLN2	P37802	Transgelin-2	2.63	3.34E-02	2.40	1.33E-01	2.34	1.07E-02
TARDBP	Q13148	TAR DNA-binding protein 43	3.70	1.84E-05	4.44	2.36E-06	2.66	5.76E-05
TCEB1	Q15369	Transcription elongation factor B polypeptide 1	30.37	3.75E-07	85.58	4.95E-07	17.83	3.09E-02
TCEB3	Q14241	Transcription elongation factor B polypeptide 3	10345290.29	2.22E-10	26563255.90	6.42E-10	9372990.60	4.70E-09
TCP1	P17987	T-complex protein 1 subunit alpha	2.85	5.31E-05	3.38	9.31E-09	1.94	4.91E-05
TDG	Q13569	G/T mismatch-specific thymine DNA glycosylase	4.27	1.57E-02	7.49	7.78E-07	4.98	1.79E-04
TENM4	Q6N022	Teneurin-4	4.81	4.73E-02	9.05	1.55E-06	3.74	2.38E-02
THRAP3	Q9Y2W1	Thyroid hormone receptor-associated protein 3	257.27	8.80E-08	618.45	4.16E-12	470.32	1.33E-14
THUMPD1	Q9NXG2	THUMP domain-containing protein 1	138706.73	2.64E-01	3176164.10	1.50E-02	30415956.28	6.74E-06
TIMP1	P01033	Metalloproteinase inhibitor 1	0.52	4.50E-01	0.00	1.31E-02	2.08	3.03E-01
TKT	P29401	Transketolase	1.79	2.73E-01	2.21	2.41E-01	1.02	9.59E-01
TLN2	Q9Y4G6	Talin-2	445.47	8.40E-04	803.94	1.03E-02	301.24	5.30E-02
TOP1	P11387	DNA topoisomerase 1	1013.18	3.60E-09	2642.80	1.83E-11	962.68	1.23E-10
TRA2A	Q13595	Transformer-2 protein homolog alpha	4394790.63	9.52E-08	18552636.97	1.29E-10	9099389.85	5.46E-08
TRA2B	P62995	Transformer-2 protein homolog beta	345.92	2.02E-11	1633.37	5.87E-16	842.84	4.89E-11
TRAF2	Q12933	TNF receptor-associated factor 2	2.94	1.33E-04	3.38	2.67E-08	1.40	1.96E-02

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
TRIM28	Q13263	Transcription intermediary factor 1-beta	6.05	1.77E-10	9.95	5.18E-15	4.14	1.02E-08
TRIM56	Q9BRZ2	E3 ubiquitin-protein ligase TRIM56	2748523.65	1.92E-10	6397124.06	3.10E-11	1930536.05	1.14E-03
TRIP12	Q14669	E3 ubiquitin-protein ligase TRIP12	21945732.32	6.93E-11	46619230.34	9.74E-11	16610139.28	3.37E-12
TRMT6	Q9UJA5	tRNA (adenine(58)-N(1))-methyltransferase non-catalytic subunit TRM6	10520419.24	3.17E-09	25031990.41	5.55E-08	9124505.59	2.20E-10
TRMT61A	Q96FX7	tRNA (adenine(58)-N(1))-methyltransferase catalytic subunit TRMT61A	1297729.74	6.72E-06	3401759.11	3.88E-11	1348956.14	9.73E-04
TSR1	Q2NL82	Pre-rRNA-processing protein TSR1 homolog	47.51	5.12E-11	144.25	3.80E-13	38.43	2.71E-11
TUBA1B;TUBA1A;TUBA1C;TUBA3C;TUBA3E;TUBA8	P68363;Q71U36;Q9BQE3;Q13748Q6PEY2	Tubulin alpha-1B chain;Tubulin alpha-1A chain;Tubulin alpha-1C chain;Tubulin alpha-3C/D chain;Tubulin alpha-3E chain;Tubulin alpha-8 chain	3.70	8.63E-06	4.32	5.72E-09	2.31	5.30E-05
TUBA4A	P68366	Tubulin alpha-4A chain	3.64	1.16E-04	4.91	4.76E-05	2.32	1.94E-02
TUBB	P07437	Tubulin beta chain	3.23	1.78E-05	3.48	1.05E-08	2.18	4.24E-05
TUBB4B;TUBB4A	P68371;P04350	Tubulin beta-4B chain;Tubulin beta-4A chain	4.34	4.78E-06	5.07	3.55E-10	2.76	3.28E-06
TUBB6	Q9BUF5	Tubulin beta-6 chain	1.00	N/A	221303.93	3.41E-01	1.00	N/A
TUBB8	Q3ZCM7	Tubulin beta-8 chain	13.68	1.17E-06	18.71	4.66E-06	5.56	1.56E-03
TUFM	P49411	Elongation factor Tu, mitochondrial	4.49	1.73E-04	5.40	1.89E-05	2.75	8.44E-06
U2AF1;U2AF1L4	Q01081	Splicing factor U2AF 35 kDa subunit;Splicing factor U2AF 26 kDa subunit	53.78	7.44E-08	167.68	2.44E-07	76.76	2.09E-05
U2AF2	P26368	Splicing factor U2AF 65 kDa subunit	35.66	1.27E-05	74.30	2.23E-06	54.28	1.32E-08

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
U2SURP	O15042	U2 snRNP-associated SURP motif-containing protein	3646402.28	3.71E-06	11626157.78	1.00E-06	3348409.42	2.39E-07
UBA1	P22314	Ubiquitin-like modifier-activating enzyme 1	2.68	6.93E-05	2.80	8.66E-05	1.37	1.79E-01
UBA52	P62987	Ubiquitin-60S ribosomal protein L40	2318561.97	5.26E-10	4953085.01	2.03E-05	1418999.06	1.42E-02
UPF1	Q92900	Regulator of nonsense transcripts 1	38.16	1.31E-07	74.90	2.17E-09	42.63	5.81E-07
UPF3B	Q9BZI7	Regulator of nonsense transcripts 3B	2659610.98	4.27E-11	6368862.91	6.84E-13	3330948.96	2.84E-09
UTP14A	Q9BVJ6	U3 small nucleolar RNA-associated protein 14 homolog A	12712364.62	1.48E-12	29661488.08	5.96E-11	10528233.95	3.16E-10
UTP18	Q9Y5J1	U3 small nucleolar RNA-associated protein 18 homolog	1488230.46	3.72E-12	3703342.73	1.49E-07	814605.86	7.72E-04
UTP23	Q9BRU9	rRNA-processing protein UTP23 homolog	2.64	1.31E-05	2.84	2.03E-05	1.44	2.49E-02
UTP3	Q9NQZ2	Something about silencing protein 10	3423172.28	3.22E-07	8982602.45	1.90E-09	3293918.33	1.14E-10
VARS	P26640	Valine--tRNA ligase	3.31	3.16E-07	2.37	2.85E-03	1.59	9.49E-02
VIM	P08670	Vimentin	4.25	1.75E-07	6.94	3.49E-09	2.60	2.37E-04
VLDLR	P98155	Very low-density lipoprotein receptor	2.19	5.12E-04	2.27	2.22E-04	0.92	6.55E-01
WDR5	P61964	WD repeat-containing protein 5	17.16	4.84E-07	26.51	2.16E-07	7.64	1.78E-03
WHSC1	O96028	Histone-lysine N-methyltransferase NSD2	1104039.21	5.69E-04	2615810.54	1.05E-08	817288.05	2.97E-07
WIBG	Q9BRP8	Partner of Y14 and mago	14498547.38	2.78E-10	56167499.89	8.77E-13	23780781.49	5.47E-08
WNK1	Q9H4A3	Serine/threonine-protein kinase WNK1	2.03	1.02E-02	0.06	6.71E-04	0.02	4.58E-04
XRCC5	P13010	X-ray repair cross-complementing protein 5	78.06	2.70E-06	351.21	7.61E-07	203.22	1.33E-09
XRCC6	P12956	X-ray repair cross-complementing protein 6	79.45	9.46E-09	348.78	8.80E-07	202.30	2.13E-10
XRN2	Q9H0D6	5-3 exoribonuclease 2	156.48	4.51E-09	274.45	3.49E-09	92.75	5.51E-09

Gene Name	Protein ID	Protein Name	T7-SRSF1		T7-SRSF1-NRS		T7-SRSF2	
			Fold Enrichment	p-value	Fold Enrichment	p-value	Fold Enrichment	p-value
YBX1	P67809	Nuclease-sensitive element-binding protein 1	114.21	9.45E-11	292.94	3.37E-09	148.23	6.71E-09
YBX3	P16989	Y-box-binding protein 3	271.32	1.48E-09	595.68	8.83E-09	269.94	8.61E-08
YTHDC1	Q96MU7	YTH domain-containing protein 1	619838.91	1.31E-02	3789564.21	8.52E-06	2084746.48	5.67E-08
YTHDC2	Q9H6S0	Probable ATP-dependent RNA helicase YTHDC2	99.62	2.23E-13	218.76	1.77E-07	90.71	9.16E-07
YWHAE	P62258	14-3-3 protein epsilon	117.07	5.82E-07	31.82	1.13E-03	7.97	7.14E-03
YWHAG	P61981	14-3-3 protein gamma;14-3-3 protein gamma, N-terminally processed	11.57	1.82E-06	4.61	2.61E-06	2.91	4.70E-02
YWHAQ	P27348	14-3-3 protein theta	5.45	6.07E-06	4.88	3.37E-07	0.73	6.22E-01
YWHAZ	P63104	14-3-3 protein zeta/delta	6.79	8.76E-08	4.82	7.95E-07	1.96	7.64E-02
ZC2HC1A	Q96GY0	Zinc finger C2HC domain-containing protein 1A	1339005.24	6.23E-07	2968808.35	1.63E-08	1345277.02	1.79E-05
ZC3H15	Q8WU90	Zinc finger CCCH domain-containing protein 15	1120530.51	2.85E-07	2116948.75	7.29E-04	107668.97	3.41E-01
ZC3H8	Q8N5P1	Zinc finger CCCH domain-containing protein 8	7317483.49	3.60E-09	20762392.04	1.21E-11	8680178.13	1.17E-09
ZC3HAV1	Q7Z2W4	Zinc finger CCCH-type antiviral protein 1	494.30	1.55E-13	1302.90	6.13E-17	533.91	6.05E-12
ZCCHC8	Q6NZY4	Zinc finger CCHC domain-containing protein 8	1924270.47	1.63E-10	4058716.92	1.96E-08	1608574.62	6.44E-02
ZFR	Q96KR1	Zinc finger RNA-binding protein	476.44	2.12E-05	1865.09	2.54E-12	632.49	9.12E-09
ZNF326	Q5BKZ1	DBIRD complex subunit ZNF326	409.31	1.31E-03	1488.52	5.88E-10	519.38	6.17E-10
ZNF512	Q96ME7	Zinc finger protein 512	709263.44	1.05E-02	1286471.16	5.90E-02	456909.93	4.95E-02
ZNF638	Q14966	Zinc finger protein 638	13620000.57	3.05E-11	19943651.66	2.04E-11	8369828.04	4.61E-12

Supplementary Table 2: IP-MS dataset for the endogenous mESC interactome (described in Section 4.3.3). Fold enrichment over wt (untagged cells) and p-values for all hits obtained in the IP-MS experiments using with SRSF1-T7 or SRSF1-NRS-T7 (Clone 1) mESCs. Only processed data values are shown, raw intensity values and LFQ values omitted for ease. For processing methods used, see Section 4.3.3 for details. Where N/A is shown for p-value and 1.00 is shown for fold enrichment, this represents a scenario where there was no binding in either condition tested. All values are given to two decimal places. Hits are in alphabetical order according to Gene Name as imputed from the UniProt database. Protein ID refers to the given UniProt ID for each protein identified. Where multiple proteins are listed, this indicates that the interacting peptide identified could map to multiple proteins or isoforms of a protein. Database search to map the raw data hits was carried out by Jimi Wills of the IGMM Mass Spectrometry facility.

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
5430403G16Rik	D3Z5L4	RIKEN cDNA 5430403G16 gene	11.65	4.23E-02	8.53	9.76E-02
Aatf	Q9JKX4	Protein AATF	916412.75	5.73E-02	808975.25	2.44E-02
Abcf1	Q6P542	ATP-binding cassette sub-family F member 1	4526675.00	4.53E-02	4254300.25	1.61E-01
Abt1	Q9QYL7	Activator of basal transcription 1	2951625.00	9.10E-04	2383050.00	2.85E-05
Acin1	Q9JIX8	Apoptotic chromatin condensation inducer in the nucleus	117.51	2.19E-02	218.02	9.31E-02
Adgrl1	Q80TR1	Adhesion G protein-coupled receptor L1	10.80	1.80E-02	13.68	9.22E-02
Aebp2	Q9Z248	Zinc finger protein AEBP2	607775.50	1.35E-01	1717105.00	1.84E-02
Ap2a2	P17427	AP-2 complex subunit alpha-2	12.41	1.91E-01	16.75	2.20E-01
Apobec3	Q99J72	DNA dC->dU-editing enzyme APOBEC-3	578875.75	3.56E-01	1.00	N/A
Arpc1b	Q9WV32	Actin-related protein 2/3 complex subunit 1B	2128050.50	1.49E-01	2089825.50	1.34E-01
AW146154	Q3TPQ7	Expressed sequence AW146154	659172.50	8.63E-03	336425.75	3.56E-01

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Bclaf1	Q8K019	Bcl-2-associated transcription factor 1	166.76	4.79E-02	300.82	9.83E-02
Blm	O88700	Bloom syndrome protein homolog	1001125.00	9.81E-05	303775.50	1.40E-01
Bms1	Q6PGF5	BMS1 homolog, ribosome assembly protein (Yeast)	13.58	6.51E-03	5.70	8.42E-02
Brd7	O88665	Bromodomain-containing protein 7	573560.25	4.55E-02	838987.75	1.07E-01
Brix1	Q9DCA5	Ribosome biogenesis protein BRX1 homolog	10.63	3.17E-03	8.59	3.04E-03
Bud13	Q8R149	BUD13 homolog	3577350.00	7.18E-04	3614550.00	3.10E-03
C12orf66	Q6P1I3	KICSTOR complex protein C12orf66 homolog	325930.00	9.81E-04	531550.50	2.97E-01
Cactin	Q9CS00	Cactin	166.77	3.39E-03	156.10	2.23E-02
Ccar1	Q8CH18	Cell division cycle and apoptosis regulator protein 1	1.00	N/A	402875.75	3.56E-01
Ccdc137	Q8R0K4	Coiled-coil domain-containing protein 137	2178925.00	2.15E-03	1153067.50	3.51E-04
Ccdc59	Q8R2N0	Thyroid transcription factor 1-associated protein 26	16.14	1.35E-03	13.13	4.82E-03
Ccnk	O88874	Cyclin-K	2263575.50	1.48E-01	1572700.50	1.34E-01
Cd3eap	Q76KJ5	DNA-directed RNA polymerase I subunit RPA34	27.34	1.50E-02	21.34	4.64E-04
Cd40lg	P27548	CD40 ligand;CD40 ligand, membrane form;CD40 ligand, soluble form	1236100.75	3.56E-01	5004250.50	1.57E-01

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Cdk11b	P24788	Cyclin-dependent kinase 11B	20.92	1.03E-03	23.49	5.45E-03
Cdk13	Q69ZA1	Cyclin-dependent kinase 13	825300.50	1.34E-01	1338577.50	1.77E-02
Cfap20	Q8BTU1	Cilia- and flagella-associated protein 20	2317225.00	2.72E-02	2662117.75	1.27E-01
Chchd2	Q9D1L0	Coiled-coil-helix-coiled-coil-helix domain-containing protein 2	26.80	1.68E-01	48.45	1.44E-01
Chtop	Q9CY57	Chromatin target of PRMT1 protein	10.20	1.25E-03	14.85	5.30E-02
Clasrp	Q8CFC7	CLK4-associating serine/arginine rich protein	7774550.00	1.97E-02	6790150.00	7.50E-02
Cpsf1	Q9EPU4	Cleavage and polyadenylation specificity factor subunit 1	9.30	1.31E-02	10.09	1.85E-02
Cwc25	Q9DBF7	Pre-mRNA-splicing factor CWC25 homolog	11.73	5.48E-03	8.70	1.82E-03
Cxxc1	Q9CWW7	CXXC-type zinc finger protein 1	2273767.50	5.56E-02	1780975.50	1.46E-01
D2Wsu81e	Q3UHX9	Putative methyltransferase C9orf114 homolog	411005.25	5.73E-02	377647.75	7.19E-02
Ddx18	Q8K363	ATP-dependent RNA helicase DDX18	12.54	2.86E-03	7.35	5.21E-04
Ddx24	Q9ESV0	ATP-dependent RNA helicase DDX24	22.36	5.85E-05	17.09	1.92E-03
Ddx27	Q921N6	Probable ATP-dependent RNA helicase DDX27	33.39	1.42E-02	18.37	2.43E-02
Ddx31	Q6NZQ2	Probable ATP-dependent RNA helicase DDX31	1998502.50	4.60E-03	1515625.00	1.38E-05
Ddx50	Q99MJ9	ATP-dependent RNA helicase DDX50	22.34	1.79E-03	23.88	1.65E-03
Ddx51	Q6P9R1	ATP-dependent RNA helicase DDX51	10.01	4.05E-03	10.58	2.36E-04

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Ddx54	Q8K4L0	ATP-dependent RNA helicase DDX54	44.85	6.91E-03	30.80	1.25E-02
Dhx15	O35286	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	41.48	3.91E-02	34.51	6.49E-02
Dhx30	Q99PU8	Putative ATP-dependent RNA helicase DHX30	86.48	6.80E-03	52.43	2.52E-02
Dhx37	Q6NZL1	DEAH (Asp-Glu-Ala-His) box polypeptide 37	14.13	1.29E-03	10.46	8.66E-04
Dhx8	A2A4P0	ATP-dependent RNA helicase DHX8	1156425.25	3.06E-02	1125875.00	1.65E-03
Dido1	Q8C9B9	Death-inducer obliterator 1	19.72	1.81E-03	20.04	1.91E-02
Dkc1	Q9ESX5	H/ACA ribonucleoprotein complex subunit 4	1078325.00	1.10E-02	1059115.00	9.48E-02
Dlgap5	Q8K4R9	Disks large-associated protein 5	1792900.50	2.10E-01	572025.75	3.56E-01
Dnajc9	Q91WN1	DnaJ homolog subfamily C member 9	973387.50	6.58E-04	520645.00	3.84E-03
Dus3l	Q91XI1	tRNA-dihydrouridine(47) synthase [NAD(P)(+)]-like	37.08	1.96E-01	38.63	1.76E-01
Enox1	Q8BHR2	Ecto-NOX disulfide-thiol exchanger 1;Hydroquinone [NADH] oxidase;Protein disulfide-thiol oxidoreductase	2332350.50	1.48E-01	1865725.50	1.61E-01
Erbb3	Q61526	Receptor tyrosine-protein kinase erbB-3	1072425.50	2.11E-01	1478900.50	1.37E-01
Erh	P84089	Enhancer of rudimentary homolog	26.61	8.05E-02	33.18	1.14E-01
Ero1l	Q8R180	ERO1-like protein alpha	2429060.25	1.28E-01	2840900.50	1.35E-01

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Exosc10	P56960	Exosome component 10	10.65	1.66E-01	7.11	1.95E-01
Exosc4	Q921I9	Exosome complex component RRP41	14.91	8.06E-03	11.83	1.89E-04
Ezh1	P70351	Histone-lysine N-methyltransferase EZH1	813782.75	6.37E-02	1158225.00	9.27E-05
Fam136a	Q9CR98	Protein FAM136A	6.53	2.16E-01	14.65	1.66E-01
Fam60a	Q8C8M1	Protein FAM60A	1321975.50	1.50E-01	1900075.50	1.93E-01
Fam76a	Q922G2	Protein FAM76A	1076900.50	1.52E-01	1652325.50	1.68E-01
Fkbp9	Q9Z247	Peptidyl-prolyl cis-trans isomerase FKBP9	932705.50	2.15E-01	805725.50	1.39E-01
Fmr1	P35922	Fragile X mental retardation protein 1 homolog	26.60	2.32E-02	24.02	1.74E-01
Ftsj3	Q9DBE9	pre-rRNA processing protein FTSJ3	19.29	5.58E-04	12.05	3.91E-03
Gab1	Q9QYY0	GRB2-associated-binding protein 1	6.53	5.57E-02	10.76	6.32E-02
Gatad2a	Q8CHY6	Transcriptional repressor p66 alpha	6.94	1.18E-01	10.46	9.60E-02
Gm10130	F6ZHM5	Predicted gene 10130	1909900.75	3.56E-01	1011275.75	3.56E-01
Gnai2	P08752	Guanine nucleotide-binding protein G(i) subunit alpha-2	5454425.50	1.90E-01	6789000.25	1.07E-01
Gnl2	Q99LH1	Nucleolar GTP-binding protein 2	14.49	1.30E-03	18.89	5.49E-02
Gpatch1	Q9DBM1	G patch domain-containing protein 1	1406382.75	8.97E-02	1.00	N/A
Gpatch4	Q3TFK5	G patch domain-containing protein 4	252275.75	3.56E-01	215088.00	1.34E-01

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Gpc4	P51655	Glypican-4;Secreted glypican-4	1370150.50	1.88E-01	1877950.50	1.40E-01
Gtf3c1	Q8K284	General transcription factor 3C polypeptide 1	262960.50	1.58E-01	1.00	N/A
H1f0	P10922	Histone H1.0;Histone H1.0, N-terminally processed	10.58	6.13E-04	7.70	1.16E-01
H1fx	Q80ZM5	H1 histone family, member X	13.73	2.21E-03	12.18	1.47E-02
H2afj;Hist1h2aa;Hist1h2ah;Hist1h2ak;Hist1h2af;Hist3h2a;Hist1h2ab	Q8R1M2; Q8CGP4; Q8CGP6; Q8CGP7; Q8CGP5; Q8BFU2;P22752	Histone H2A.J;Histone H2A;Histone H2A type 1-H;Histone H2A type 1-K;Histone H2A type 1-F;Histone H2A type 3;Histone H2A type 1	3343400.00	3.11E-02	261725.75	3.56E-01
H2afx	P27661	Histone H2AX	95.42	1.33E-01	21.47	1.14E-01
Hdac6	Q9Z2V5	Histone deacetylase 6	808328.00	1.85E-01	1154650.50	1.69E-01
Hic2	Q9JLZ6	Hypermethylated in cancer 2 protein	1.00	N/A	1118050.50	1.64E-01
Hmga1	P17095	High mobility group protein HMG-I/HMG-Y	6443410.25	2.63E-01	1583775.25	1.96E-01
Hp1bp3	Q3TEA8	Heterochromatin protein 1-binding protein 3	1078150.25	2.90E-02	1321605.00	8.95E-03
Igf2bp2	Q5SF07	Insulin-like growth factor 2 mRNA-binding protein 2	13.68	1.03E-03	8.09	1.61E-03

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Ighv5-15	A0A0A6YWC7	Immunoglobulin heavy variable 5-15	1.00	N/A	275750.75	3.56E-01
Isg20l2	Q3U1G5	Interferon-stimulated 20 kDa exonuclease-like 2	3935400.00	1.91E-03	2421975.00	1.40E-04
Itga3	Q62470	Integrin alpha-3;Integrin alpha-3 heavy chain;Integrin alpha-3 light chain	239.34	1.52E-01	320.71	1.50E-01
Itgb4	A2A863	Integrin beta-4	1.00	N/A	128090.75	3.56E-01
Jarid2	Q62315	Protein Jumonji	9.08	7.75E-04	11.87	5.08E-03
Kiaa0020	Q8BKS9	Pumilio domain-containing protein KIAA0020	66.87	1.47E-02	37.10	5.35E-03
Kiaa1429	A2AIV2	Protein virilizer homolog	2584155.25	9.99E-02	2588625.25	1.29E-01
Kif23	E9Q5G3	Kinesin-like protein KIF23	37.45	1.55E-03	33.16	3.44E-02
Llph	Q9D945	Protein LLP homolog	9.69	9.53E-02	15.36	1.94E-03
Luc7l3	Q5SUF2	Luc7-like protein 3	11.41	2.11E-02	17.00	1.73E-01
Magt1	Q9CQY5	Magnesium transporter protein 1	445290.50	1.36E-01	887402.75	1.29E-01
Mak16	Q8BGS0	Protein MAK16 homolog	2177650.00	2.27E-03	1246850.25	2.50E-02
Mcm2	P97310	DNA replication licensing factor MCM2	17.81	1.87E-01	12.76	1.73E-01
Mina	Q8CD15	Bifunctional lysine-specific demethylase and histidyl-hydroxylase MINA	390700.75	3.56E-01	392550.50	2.02E-01
Mki67	E9PVX6	Proliferation marker protein Ki-67	1630575.25	2.59E-02	3043575.00	2.38E-02

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Mreg	Q6NVG5	Melanoregulin	3275300.50	1.93E-01	3504175.50	1.45E-01
Mri1	Q9CQT1	Methylthioribose-1-phosphate isomerase	8.89	2.70E-01	38.93	2.07E-01
Mrpl11	Q9CQF0	39S ribosomal protein L11, mitochondrial	132.98	3.39E-02	108.50	5.51E-05
Mrpl15	Q9CPR5	39S ribosomal protein L15, mitochondrial	1877010.00	8.74E-02	1716820.25	1.18E-01
Mrpl2	Q9D773	39S ribosomal protein L2, mitochondrial	6529147.75	1.41E-01	6272077.75	1.48E-01
Mrpl23	O35972	39S ribosomal protein L23, mitochondrial	12.91	1.16E-01	15.44	1.59E-01
Mrpl24	Q9CQ06	39S ribosomal protein L24, mitochondrial	8.27	1.44E-01	12.66	1.76E-01
Mrpl32	Q9DCI9	39S ribosomal protein L32, mitochondrial	785747.75	1.64E-01	896550.50	1.66E-01
Mrpl41	Q9CQN7	39S ribosomal protein L41, mitochondrial	2337067.75	1.30E-01	2635050.50	1.34E-01
Mrpl42	Q9CPV3	39S ribosomal protein L42, mitochondrial	24.79	7.38E-02	19.81	1.71E-01
Mrpl43	Q99N89	39S ribosomal protein L43, mitochondrial	965795.50	2.98E-01	1759657.75	1.19E-01
Mrpl47	Q8K2Y7	39S ribosomal protein L47, mitochondrial	11.17	1.35E-01	4.33	2.64E-01
Mrpl49	Q9CQ40	39S ribosomal protein L49, mitochondrial	307475.75	3.56E-01	351800.75	3.56E-01
Mrpl9	Q99N94	39S ribosomal protein L9, mitochondrial	2136035.00	1.39E-01	958487.75	4.59E-02
Mrps16	Q9CPX7	28S ribosomal protein S16, mitochondrial	15.41	1.42E-01	19.66	1.53E-01
Mrps17	Q9CQE3	28S ribosomal protein S17, mitochondrial	412225.50	1.41E-01	459905.50	1.34E-01
Mrps18b	Q99N84	28S ribosomal protein S18b, mitochondrial	2080300.25	1.07E-01	3092825.50	1.86E-01

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Mrps26	Q80ZS3	28S ribosomal protein S26, mitochondrial	67.28	3.44E-02	77.91	1.10E-01
Mrps30	Q9D0G0	28S ribosomal protein S30, mitochondrial	863260.25	3.60E-02	629837.50	1.03E-04
Mrps33	Q9D2R8	28S ribosomal protein S33, mitochondrial	1119980.25	7.38E-02	1087260.00	1.01E-02
Mrps34	Q9JIK9	28S ribosomal protein S34, mitochondrial	33.53	3.48E-02	31.31	6.08E-02
Mrps5	Q99N87	28S ribosomal protein S5, mitochondrial	19.66	3.39E-02	20.69	9.90E-02
Mrps6	P58064	28S ribosomal protein S6, mitochondrial	2387257.50	7.17E-02	1734365.00	6.31E-02
Mrto4	Q9D0I8	mRNA turnover protein 4 homolog	32.31	2.64E-02	23.40	4.02E-03
Mtf2	Q02395	Metal-response element-binding transcription factor 2	11.94	2.16E-04	12.16	1.99E-03
Nav3	Q80TN7	Neuron navigator 3	1946450.75	3.56E-01	1.00	N/A
Ncbp1	Q3UYV9	Nuclear cap-binding protein subunit 1	841325.50	1.40E-01	1.00	N/A
Nepro	Q8R2U2	Nucleolus and neural progenitor protein	14.79	1.92E-03	13.98	5.03E-04
Ngdn	Q9DB96	Neuroguidin	27.80	7.11E-04	26.47	3.66E-03
Nifk	Q91VE6	MKI67 FHA domain-interacting nucleolar phosphoprotein	10712400.00	4.47E-04	8981575.00	2.70E-03
Noc3l	Q8VI84	Nucleolar complex protein 3 homolog	566750.50	1.37E-01	574110.25	3.18E-02
Nol10	Q5RJG1	Nucleolar protein 10	13.02	2.73E-03	15.21	3.89E-03
Nol6	Q8R5K4	Nucleolar protein 6	1575550.25	4.71E-02	889650.50	1.44E-01

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Nop2	Q922K7	Probable 28S rRNA (cytosine-C(5))-methyltransferase	17.07	2.50E-03	12.52	4.80E-04
Nop53	Q8BK35	Ribosome biogenesis protein NOP53	35.19	2.14E-03	34.90	3.68E-03
Nsa2	Q9CR47	Ribosome biogenesis protein NSA2 homolog	12.93	9.28E-05	13.24	1.36E-02
Numa1	E9Q7G0	Nuclear mitotic apparatus protein 1	6.06	3.40E-03	13.59	5.63E-02
Oxa1l	Q8BGA9	Mitochondrial inner membrane protein OXA1L	251250.75	3.56E-01	1318620.00	2.33E-02
P2rx7	Q9Z1M0	P2X purinoceptor 7	11.70	1.92E-01	32.51	1.88E-01
Pbrm1	Q8BSQ9	Protein polybromo-1	66030.75	3.56E-01	491980.25	4.68E-02
Pdcd11	Q6NS46	Protein RRP5 homolog	65.49	8.94E-04	50.02	1.31E-04
Pinx1	Q9CZX5	PIN2/TERF1-interacting telomerase inhibitor 1	78638.25	3.56E-01	92603.00	1.39E-01
Pnn	O35691	Pinin	8.63	7.19E-02	25.04	1.63E-01
Polr1e	Q8K202	DNA-directed RNA polymerase I subunit RPA49	16.07	1.88E-02	12.79	6.17E-04
Polr2h	Q923G2	DNA-directed RNA polymerases I, II, and III subunit RPABC3	668892.75	3.22E-02	973932.75	1.31E-01
Popdc3	Q9ES81	Popeye domain-containing protein 3	809500.75	3.56E-01	1.00	N/A
Ppan	Q91YU8	Suppressor of SWI4 1 homolog	16.74	4.60E-03	16.44	1.66E-02
Ppig	A2AR02	Peptidyl-prolyl cis-trans isomerase G	54.78	1.05E-02	96.69	8.96E-02
Prkcsh	O08795	Glucosidase 2 subunit beta	29.34	1.51E-01	84.66	1.47E-01

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Prpf38a	Q4FK66	Pre-mRNA-splicing factor 38A	1116725.50	1.98E-01	790032.75	7.86E-02
Prpf38b	Q80SY5	Pre-mRNA-splicing factor 38B	5102825.00	5.43E-03	4397275.25	7.55E-02
Prpf4b	Q61136	Serine/threonine-protein kinase PRP4 homolog	62.72	3.20E-04	71.84	2.33E-02
Ptcd3	Q14C51	Pentatricopeptide repeat domain-containing protein 3, mitochondrial	1.40	8.25E-01	12.01	6.65E-02
Pwp1	Q99LL5	Periodic tryptophan protein 1 homolog	3768475.00	1.08E-02	2001522.50	1.72E-02
Raly	Q64012	RNA-binding protein Raly	940842.75	1.08E-01	2558957.50	9.77E-02
Rbm15	Q0VBL3	RNA-binding motif protein 15	101.35	1.10E-03	87.69	2.59E-03
Rbm19	Q8R3C6	Probable RNA-binding protein 19	33.68	1.19E-02	23.95	8.18E-05
Rbm28	Q8CGC6	RNA-binding protein 28	99.42	2.10E-03	66.91	2.89E-03
Rbm34	Q8C5L7	RNA-binding protein 34	37.41	9.22E-03	29.50	1.35E-02
Rbmxl1;Rbmx	Q91VM5; Q9WV02	RNA binding motif protein, X-linked-like-1;RNA-binding motif protein, X chromosome;RNA-binding motif protein, X chromosome, N-terminally processed	43.35	6.45E-02	56.12	1.26E-01
Rbmxl2	Q9DAE2	RNA-binding motif protein, X-linked-like 2	87.01	1.62E-02	101.12	1.93E-01
Rcl1	Q9JJT0	RNA 3-terminal phosphate cyclase-like protein	622475.50	1.34E-01	443325.50	1.68E-01
Rexo4	Q6PAQ4	RNA exonuclease 4	4953950.00	1.98E-03	1440870.00	1.98E-03

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Rnf20	Q5DTM8	E3 ubiquitin-protein ligase BRE1A	4.88	3.11E-01	12.31	1.86E-01
Rnf40	Q3U319	E3 ubiquitin-protein ligase BRE1B	12.77	2.15E-01	18.37	2.30E-01
Rnps1	Q99M28	RNA-binding protein with serine-rich domain 1	47.06	6.29E-03	65.06	1.24E-01
Rpf2	Q9JJ80	Ribosome production factor 2 homolog	16.30	1.96E-03	10.98	3.38E-03
Rpp38	Q80UU2	Ribonuclease P protein subunit p38	8.22	1.60E-03	10.68	1.11E-02
Rrp15	Q9CYX7	RRP15-like protein	13.89	1.12E-04	9.85	7.05E-03
Rrp7a	Q9D1C9	Ribosomal RNA-processing protein 7 homolog A	34.47	2.16E-03	22.44	2.95E-04
Rrp8	Q9DB85	Ribosomal RNA-processing protein 8	3984750.00	1.35E-03	3337575.00	2.25E-03
Rrs1	Q9CYH6	Ribosome biogenesis regulatory protein homolog	16.33	4.18E-03	13.57	1.68E-03
Rsl24d1	Q99L28	Probable ribosome biogenesis protein RLP24	1666225.25	1.15E-01	1520512.75	9.74E-02
Safb	D3YXK2	Scaffold attachment factor B1	1522100.25	6.88E-02	1383902.75	1.25E-01
Sap18	O55128	Histone deacetylase complex subunit SAP18	178.06	5.55E-02	315.13	1.23E-01
Sap30bp	Q02614	SAP30-binding protein	2514482.50	5.04E-03	2573825.00	1.59E-04
Setd5	Q5XJV7	SET domain-containing protein 5	65090000.75	3.56E-01	1.00	N/A
Sf3a1	Q8K4Z5	Splicing factor 3A subunit 1	3216377.50	3.62E-02	489538.00	1.93E-01
Sgol2	Q7TSY8	Shugoshin-like 2	2917825.00	7.49E-04	6217700.00	1.85E-02
Sin3a	Q60520	Paired amphipathic helix protein Sin3a	17.22	1.31E-01	27.58	1.76E-01

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Snap23	O09044	Synaptosomal-associated protein 23	11.13	1.95E-01	23.16	1.91E-01
Snrnp27	Q8K194	U4/U6.U5 small nuclear ribonucleoprotein 27 kDa protein	607050.50	1.54E-01	258975.75	3.56E-01
Snrnp70	Q62376	U1 small nuclear ribonucleoprotein 70 kDa	16.21	8.26E-02	24.98	2.04E-01
Snrpa	Q62189	U1 small nuclear ribonucleoprotein A	15.81	8.41E-02	19.38	1.57E-01
Snrpb2	Q9CQI7	U2 small nuclear ribonucleoprotein B	743075.50	1.53E-01	1.00	N/A
Snrpc	Q62241	U1 small nuclear ribonucleoprotein C	14.47	1.38E-01	17.59	2.14E-01
Snrpd1	P62315	Small nuclear ribonucleoprotein Sm D1	35.87	1.52E-01	45.57	1.27E-01
Snrpd2	P62317	Small nuclear ribonucleoprotein Sm D2	9.48	9.50E-02	11.67	1.82E-01
Snrpd3	P62320	Small nuclear ribonucleoprotein Sm D3	10.34	1.54E-01	11.77	1.96E-01
Snrpf	P62307	Small nuclear ribonucleoprotein F	7.49	2.14E-01	12.08	2.59E-01
Snrpg	P62309	Small nuclear ribonucleoprotein G	18.07	1.26E-01	22.03	2.02E-01
Son	Q9QX47	Protein SON	64.29	1.48E-02	81.69	5.83E-02
Sparc	P07214	SPARC	13.38	1.58E-01	17.30	2.09E-01
Srp72	F8VQC1	Signal recognition particle subunit SRP72	1.00	N/A	268875.75	3.56E-01
Srpk1	O70551	SRSF protein kinase 1	13.88	6.97E-04	8.82	3.43E-03
Srrm1	Q52KI8	Serine/arginine repetitive matrix protein 1	30.55	1.36E-04	41.30	6.82E-02

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Srrm2	Q8BTI8	Serine/arginine repetitive matrix protein 2	12.17	2.10E-02	24.80	7.37E-02
Srsf1	Q6PDM2	Serine/arginine-rich splicing factor 1	210.59	1.01E-02	158.24	4.72E-02
Srsf10	Q9R0U0	Serine/arginine-rich splicing factor 10	8126150.00	5.85E-03	13184675.00	4.27E-02
Srsf3	P84104	Serine/arginine-rich splicing factor 3	8.87	5.69E-03	12.00	9.79E-02
Srsf4	Q8VE97	Serine/arginine-rich splicing factor 4	36.95	2.63E-03	57.14	8.96E-02
Srsf5	O35326	Serine/arginine-rich splicing factor 5	13.99	5.33E-05	17.52	4.67E-02
Srsf6	Q3TWW8	Serine/arginine-rich splicing factor 6	13.04	2.83E-04	21.17	7.16E-02
Srsf7	Q8BL97	Serine/arginine-rich splicing factor 7	15.17	1.26E-03	22.34	7.28E-02
Srsf9	Q9D0B0	Serine/arginine-rich splicing factor 9	13.78	2.55E-02	9.80	1.02E-01
Stau1	Q9Z108	Double-stranded RNA-binding protein Staufen homolog 1	1210307.50	3.41E-03	1.00	N/A
Suz12	Q80U70	Polycomb protein Suz12	25.69	3.70E-03	36.84	4.74E-04
Tbc1d31	Q6NXY1	TBC1 domain family member 31	957900.75	3.56E-01	5563250.75	3.56E-01
Tes	P47226	Testin	876475.50	1.64E-01	955125.50	1.74E-01
Thrap3	Q569Z6	Thyroid hormone receptor-associated protein 3	38.29	3.20E-02	53.00	9.59E-02
Tle1;Tle4;Tle2	Q62440;Q62441;Q9WVB2	Transducin-like enhancer protein 1;Transducin-like enhancer protein 4;Transducin-like enhancer protein 2	13.66	1.72E-01	12.25	1.73E-01

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Tle3	Q08122	Transducin-like enhancer protein 3	1133825.50	1.61E-01	1305950.50	1.71E-01
Tra2a	Q6PFR5	Transformer-2 protein homolog alpha	67.95	5.51E-03	96.71	6.60E-02
Tra2b	P62996	Transformer-2 protein homolog beta	28.42	3.28E-02	41.73	9.38E-02
Trim33	Q99PP7	E3 ubiquitin-protein ligase TRIM33	13.83	2.27E-01	23.21	1.50E-01
Ubr7	Q8BU04	Putative E3 ubiquitin-protein ligase UBR7	4313525.50	1.46E-01	5467950.50	1.95E-01
Utp18	Q5SSI6	U3 small nucleolar RNA-associated protein 18 homolog	2730850.00	3.81E-05	1956535.00	1.16E-02
Utp23	Q9CX11	rRNA-processing protein UTP23 homolog	2070550.25	3.90E-02	283490.50	2.10E-01
Wtap	Q9ER69	Pre-mRNA-splicing regulator WTAP	8.17	1.64E-01	17.82	2.72E-01
Ythdc1	E9Q5K9	YTH domain-containing protein 1	1977110.00	1.03E-02	3699105.00	1.46E-01
Zc3h13	E9Q784	Zinc finger CCCH domain-containing protein 13	52.16	7.25E-03	114.45	7.02E-02
Zcchc8	Q9CYA6	Zinc finger CCHC domain-containing protein 8	1073400.50	1.44E-01	2033350.25	3.82E-02
Zfp715	G3X9T1	Zinc finger protein 715	672300.50	1.45E-01	976300.50	1.50E-01
Zfp869	Q9DC47	Zinc finger protein 869	23.97	1.83E-01	18.70	1.63E-01
Zfyve1	Q810J8	Zinc finger FYVE domain-containing protein 1	125810.50	1.46E-01	96740.75	3.56E-01
Znf319	Q497V9	Zinc finger protein 319	669530.50	2.20E-01	512928.00	1.68E-01
Znf354c	Q571J5	Zinc finger protein 354C	262295.25	1.03E-01	1.00	N/A

Gene Name	Protein ID	Protein Name	SRSF1-T7		SRSF1-NRS-T7 (Clone 1)	
			Fold Enrichment	p-value	Fold Enrichment	p-value
Znf48	Q3US17	Zinc finger protein 48	9.93	1.26E-03	12.88	5.04E-02
Znf569	Q80W31	Zinc finger protein 569	11.18	1.41E-02	9.60	1.11E-01
Znf638	Q61464	Zinc finger protein 638	68.73623526	1.01E-04	84.93609984	3.25E-02