Analysis of RNA binding proteins using stable cell lines

Mahsa Kheirollahi Kouhestani

A thesis submitted for the degree of Doctor of Philosophy

Institute of Genetic Medicine

(IGM)

Newcastle University

December 2013

Declaration

I, Mahsa Kheirollahi Kouhestani, declare that no portion of the work compiled in this thesis has been submitted in support of another degree or qualification at this or any other University or Institute of Learning. This thesis includes nothing which is the work of others, nor the outcomes of work done in collaboration, except where otherwise stated.

…………………………………………

Mahsa Kheirollahi Kouhestani

Acknowledgements

stage without their tremendous support. I love them both and I am grateful to have these wonderful parents.

List of figures

List of tables

Abbreviations

ATRS	AT-rich sequence
BrdU	Bromodeoxyuridine
BRCA1	Breast cancer-specific tumor suppressor protein 1
BSA	Bovine Serum Albumin
CARM1	Coactivator-Associated Arginine Methyltransferase 1
CBP	CBP CREB-binding protein
CCAR1	Cell cycle and apoptosis regulatory protein 1
Cdna	Complementary DNA
CoCoA	Coiled-coil coactivator
CREB	cAMP response element-binding protein
CTE	Constitutive Transport Element
C-terminus	Carboxy-terminus
DAPI	4',6-diamidino-2-phenylindole
dH$_2$O	Millipore filtered water
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphate
dsRBDs	double stranded RNA binding domains
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EMT	Epithelial–mesenchymal transitions
ESE	Exon splicing enhancer
ESS	Exon splicing silencer
FACS	Fluorescence activated cell sorting
FBS	Foetal Bovine Serum
HEK293	Human Embryonic Kidney 293 cells
HIV	Human Immunodeficiency Virus
hnRNP	Heterogeneous nuclear ribonucleoprotein
IGM	Institute of Genetic Medicine

IP	Immunoprecipitation
ISE	Intron splicing enhancer
ISS	Intron splicing silencer
KCl	Potassium Chloride
LB	Luria Bertani
MDR1	Multidrug-resistant
mRNA	Messenger Ribonucleic Acid
NDH II	DNA helicase II
NES	Nuclear Export Sequence
NP-40	Nonylphenoxylpolyethoxylethanol-40
N-terminus	Amino-terminus
PARP	poly [ADP-ribose] polymerase
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
RHA	RNA helicase A
RNA	Ribonucleic Acid
rpm	Revolutions Per Minute
RRE	Rev Response Element
RRM	RNA recognition motif
RT-PCR	Reverse Transcriptase PCR
Sam68	Src-associated substrate during mitosis of 68 kDa
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SMN	Survival motor neuron gene
SNB	Sam68/SLM nuclear body
snRNPs	small nuclear ribonucleoprotein assembly
SR protein	Serine/arginine-rich protein
STAR	Signal Transduction and Activation of RNA
TAE	Tris-acetate-EDTA buffer
TBS	Tris Buffered Saline
TBS-T	TBS-Tween
TF	Transcription factors
Tra2	Transformer-2 gene

Abstract

Over 94% of human genes are alternatively spliced. Alternative splicing regulation is necessary for some developmental pathways. Many proteins including RNA binding proteins such as STAR family, hnRNPs, SR and SR-like proteins are involved in alternative splicing regulation. Some members of these proteins also have important roles in gene expression, transcription, signal transduction, RNA metabolism, cell cycle regulation and cancer. Although these proteins are known to be involved in alternative splicing control of specific targets, mechanisms of many of their actions and protein interaction partners are unknown. The aim of present study was the identification and characterization of protein interaction partners of some of these proteins including: STAR family (T-STAR, Sam68), hnRNP proteins (hnRNP G, hnRNP G-T) and SR-like proteins (Tra2β). The cDNAs of T-STAR, wild type Sam68 and P439R mutant Sam68 were cloned into a FLAG epitope encoding vector. The next step was generation of stable cell lines expressing these FLAG tagged proteins. To induce protein expression, stable cell lines were treated with tetracycline. The wild type Sam68-FLAG tagged protein was nuclear, while the P439R mutant Sam68-FLAG had cytoplasmic localization. Using immunoprecipitation and mass spectroscopy, both FLAG tagged proteins and co-purified proteins were purified and identified. RHA (RNA helicase A) was detected as an interacting partner for T-STAR, wild type and P439R mutant Sam68. Also, CCAR1 (cell division cycle and apoptosis regulator 1) was detected as a further partner for wild type and mutant Sam68. In addition, stable cell lines expressing FLAG tagged hnRNP G, hnRNP G-T and Tra2β were generated. These proteins and their candidate protein partners were pulled down and detected using immunoprecipitation and mass spectroscopy. Some of these detected proteins such as hnRNP C, hnRNP CL1 and RNA binding motif protein, X-Linked-Like 1 were common interacting candidates for both hnRNP G and hnRNP G-T. In addition to confirmation the roles of hnRNP G and hnRNP G-T in reduction of cell growth, the over-expression of hnRNP G-T led to remarkable cell morphological changes and alternative splicing of some target genes. In conclusion, in this project new interacting protein partners for T-STAR, Sam68, hnRNP G, hnRNP G-T were detected which their roles need to be tested within the cell.

<div align="center">**Chapter 1**</div>

1   Introduction

1.1   Human genes produce multiple mRNAs

From a molecular view, a gene is defined as a nucleic acid sequence which give rises to functional products such as proteins and RNAs (Lodish, 2004; Gerstein *et al.*, 2007). In eukaryotic cells there are different machineries for controlling gene expression. At first it seemed that selection and activation of gene promoters are most important factors for regulation of gene expression. This view was changed by understanding more about post-transcriptional mechanisms (Siomi and Dreyfuss, 1997). Although in prokaryotes transcription and translation take place at the same location, in eukaryotes transcription happens in the nucleus and translation occurs in the cytoplasm. As a result this leads to more post-transcriptional processing of pre-mRNAs (Glisovic et al., 2008).

For a long time there was a belief that the number of genes established the complexity of an organism. This idea was changed after discovering the fact that there is not very much difference in the number of genes between human and less complex creatures. Therefore, there must be other processes inside the cell which lead to human complexity. By post-transcriptional regulation diverse cell types can adjust patterns of their gene expression in different situations. Especially in some developmental processes such as pattern configuration and differentiation, post-transcriptional regulation has a very important role (Siomi and Dreyfuss, 1997). Also creation of different types of proteins from one single gene can be achieved by post-transcriptional regulation of gene expression. Ways for generating RNA sequence diversity include: RNA editing (Smith and Sowden, 1996), alternative polyadenylation site selection (Proudfoot, 1996), and alternative splicing of pre-mRNAs (Chabot, 1996) (figure1).
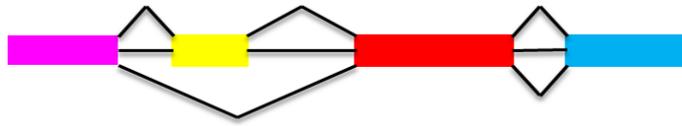
Figure 1: Some ways to produce RNA sequence diversity. Alternative splicing, RNA editing and alternative polyA site selection are some of the important ways to produce RNA sequence diversity. Alternative splicing can lead to different combination of exons from one pre-mRNA and to generate diverse mRNAs. Base RNA editing changes such as C→U and also alternative polyA site selection can lead to the production of RNA sequence diversity. The picture is modified from (Siomi and Dreyfuss, 1997).

1.2 Alternative splicing

Pre-mRNAs are produced by RNA Polymerase II. They also have a historical name, heterogeneous nuclear RNAs (hnRNAs), which describes the heterogeneity and localization of pre-mRNA (Dreyfuss et al., 2002). Production of mature mRNA requires some processes such as capping, polyadenylation and splicing (Beyer and Osheim,

1988; Bauren *et al.*, 1998). The splicing process was discovered for the first time in 1977 (Berget et al., 1977; Chow et al., 1977; Gelinas and Roberts, 1977). It has been shown by sequencing that in human only 1.2% of DNA bases are in exons (Lander *et al.*, 2001; Venter *et al.*, 2001). Mature RNA which is exported to the cytoplasm only contains exons (Elliott and Grellscheid, 2006). Therefore, mature transcripts in eukaryotes are much shorter than genes. For example, the average gene length in human is about 27000bp while mature transcript length is about 1300bp (Lander et al., 2001; Venter et al., 2001). This means that usually introns are much longer in comparison with exons (Elliott and Grellscheid, 2006).

Splicing takes place in the nucleus using a complex called the spliceosome. This spliceosome complex is made of small nuclear RNAs and proteins (Gerstein et al., 2007). There are more than one hundred proteins and five small nuclear ribonuclear proteins (snRNPs) in this complex (U1,U2,U4/U6 and U5) (Wahl et al., 2009). Spliceosome complexes are assembled on transcripts (Gornemann et al., 2005; Lacadie and Rosbash, 2005; Azubel et al., 2006). Splicing occurs in two steps. Firstly, the branch point adenosine attacks the 5' splice site. This reaction leads to making a lariat intermediate. Secondly, the free 5' exon attacks the 3' splice site and exons will be attached to each other. Afterwards, the lariat will be discarded (figure 2). Spliceosome composition and assembly are still under analysis (Azubel *et al.*, 2004; Gornemann *et al.*, 2005; Lacadie and Rosbash, 2005; Chen and Moore, 2014; Matera and Wang, 2014). Conversely to older models with the idea that splicing occurs after completion of transcription, current data show splicing and transcription can take place at the same time (Bentley, 2005).
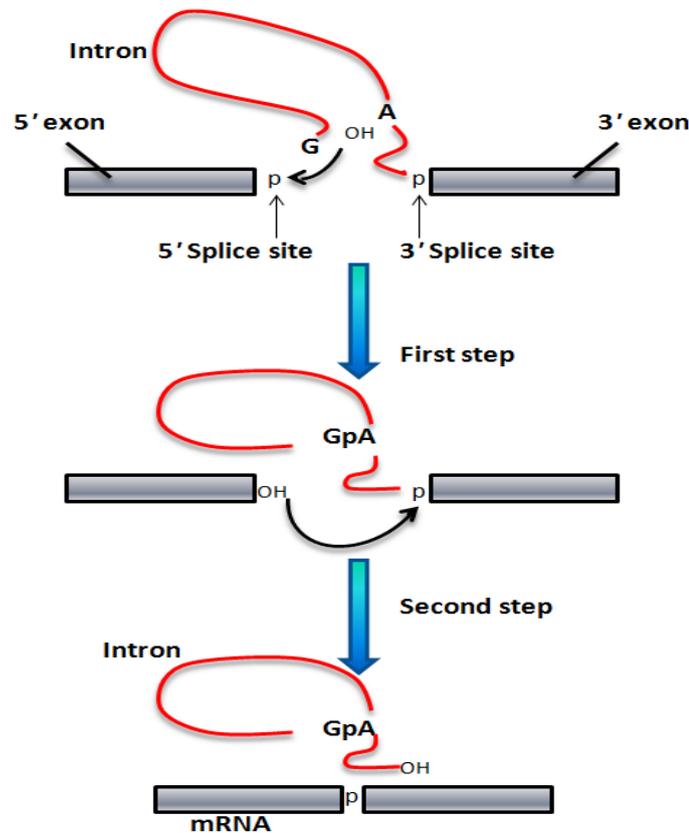
Figure 2: Splicing mechanism. This picture shows how an intron is removed and exons attach to each other. The picture is modified from (Newman, 2001). The branch point adenosine attacks the 5' splice site. This reaction creates a lariat intermediate. The free 5' exon attacks the 3' splice site and exons will be attached to each other.

An important property of alternative splicing is that it is a good process to produce different combination of exons from one pre-mRNA and to generate diverse mRNAs (Matlin et al., 2005) (figure 3). Therefore, alternative splicing can lead to an extension of the coding capacity in the genome and the production of many different kinds of proteins (Stamm, 2002). Additionally, alternative splicing can alter the reading frame of transcripts (Johnson et al., 2003). By alternative splicing, one genetic locus can be used for coding of diverse pre-mRNAs (Gerstein et al., 2007).

There are some sequences inside the pre-mRNA such as 'exonic splicing enhancers' (ESEs) and 'intronic splicing enhancers' (ISEs) which enhance splicing. On the other hand, silencing signals such as 'exonic splicing silencers' (ESSs) and 'intronic splicing silencers' (ISSs) repress the splicing process. In addition, secondary structures within the pre-mRNA have roles in splicing by hiding or displaying splicing signals and also placing specific sequences close to each other (Warf and Berglund, 2010). In this way splicing efficiency will be increased (Warf and Berglund, 2010). It is thought that correct splicing of target RNA depends on different expression levels of RNA splicing regulators (Fu, 2004; Matlin et al., 2005).

Specific patterns of tissue gene expression can also be achieved by alternative splicing (Pan et al., 2004). It has been shown that alternative splicing can influence about 90% of human genes (Wang et al., 2008). Selection of splice sites is regulated by factors such as splicing proteins which bind to RNA, the structure of pre-mRNA, and the presence of some specific sequences in pre-mRNAs. Testing of alternative splicing has been established by constructing minigenes which contain some exons and splicing signals as well as flanking introns (Stoss et al., 1999). A number of signalling pathways which act in response to extra cellular stimuli can also regulate alternative splicing (Lynch, 2007).
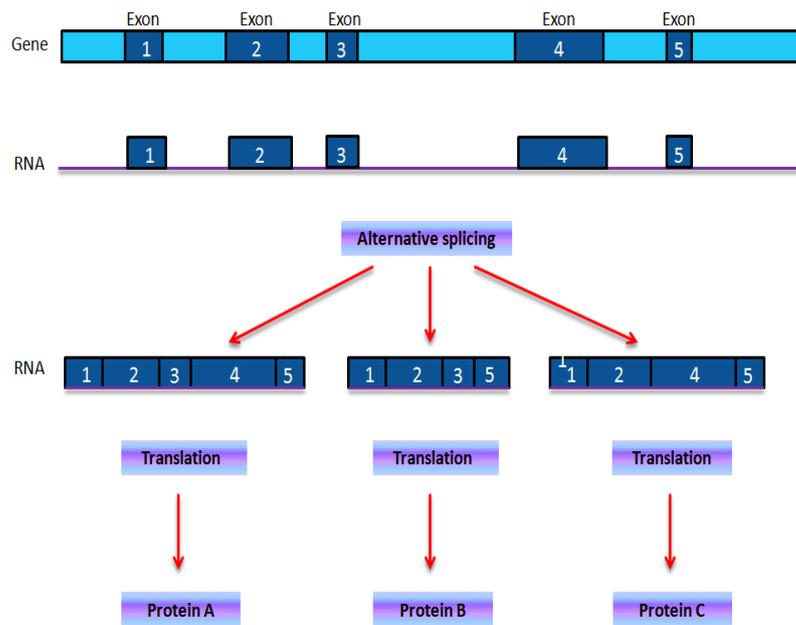
Figure 3: Alternative splicing can make different proteins from one gene. The picture is modified from (Guttmacher and Collins, 2002).

Especially during development, alternative splicing has a key role. Although our knowledge about this process is not complete yet, it seems that expression of cell or tissue-specific splicing factors and also extrinsic signals have important effects on this process (Maniatis and Tasic, 2002; Stamm, 2002; Gooding *et al.*, 2003). It has been shown that in some tissues there are high levels of alternative splicing (Liu and Elliott, 2010). For example, the highest levels of alternative splicing have been identified in brain, testis and liver respectively (Yeo et al., 2004). While high levels of evolutionarily conserved splicing have been seen when comparing between brains of humans and mice, in the testes there is more non-conserved splicing. Based on evolutionarily conservation, it is thought that conserved splicing between human and mouse in the brain has an important role (Kan et al., 2005; Yeo et al., 2005).

One of the most important sites of continuing development in the adult is in the testis. In this tissue male germ cells undergo essential developmental pathways. By differentiation of spermatogonia in prophase of Meiosis I, primary spermatocytes are

produced. First and second meiotic divisions of each spermatogonium lead to the production of two secondary spermatocytes and four spermatids respectively. Spermiogenic differentiation is a process which makes mature sperm from spermatids. In this process, some important events take place such as nuclear condensation and cytoplasmic ejection. Histone replacement (sequentially by transitional proteins and protamins) leads to chromatin packaging during spermiogenesis (Amann and Howards, 1980).

There are particular programmes of gene expression in the testes. These programmes lead to producing mature spermatozoa from spermatogonia. One of these programmes is alternative splicing of pre-mRNAs (Venables, 2002). In a human testis 10^8 sperm are produced per day, and during the spermatogenesis process division and differentiation of an enormous number of germ cells occurs (Gilbert, 2000). It seems that the high level of alternative splicing in the testes is due to the requirement of increased transcript isoforms (Liu et al., 2009). In the testis the high level of alternative splicing requires distinct expression patterns of splicing regulators in spermatogenesis (Kan *et al.*, 2005; Schmid *et al.*, 2013).

1.3 Regulators of alternative splicing

There are different kinds of alternative splicing mechanisms including: alternative 5' splice sites, alternative 3' splice sites, retained introns, exon skipping and mutually exclusive exons. Alternative mRNA isoforms can also be made using alternative promoters and alternative polyadenylation (Keren *et al.*, 2010). Important *cis*-elements include ESS (Exonic Splicing Silencer), ESE (Exonic Splicing Enhancer), ISS (Intronic Splicing Silencer) and ISE (Intronic Splicing Enhancer). *Trans*-acting proteins are involved in regulation of alternative splicing (Heyd and Lynch, 2011). The spliceosome plus approximately one hundred regulatory factors are involved in selection of splicing sites in the pre-mRNA (Tsuda *et al.*, 2011). Some alternative

splicing regulators which are tissue-specific have been recognized such as Nova-1, ESRP1 and ESRP2 (Grabowski, 2000; Warzecha et al., 2009).

Among the factors regulating alternative splicing, serine/arginine-rich (SR) proteins, Signal Transduction and Activation of RNA (STAR) family and heterogeneous ribonucleoproteins (hnRNPs) have been identified as important alternative splicing regulators (Bourgeois *et al.*, 2004; Martinez-Contreras *et al.*, 2007; Long and Caceres, 2009). In the next sections more information regarding these groups of RNA binding proteins has been discussed.

1.4   RNA binding proteins

RNA binding proteins have important roles in processing, transportation, localization, translation and stabilization of mRNAs. 2% of genes in *Caenorhabditis elegans* and *Drosophila melanogaster* encode RNA binding proteins. There are three types of modification for RNA binding proteins: phoshorylation, arginine methylation and SUMO (small ubiquitin-like modifier) (Glisovic et al., 2008). In the SUMOylation process SUMO proteins are added to the target proteins. There are three isoforms of SUMO proteins (Matunis *et al.*, 1996; Chen *et al.*, 1998; Kamitani *et al.*, 1998). SUMOylation can compete with ubiquitination and acetylation as lysine residues are the targets for these post-translational modifications as well. SUMOylation can block or create interaction surfaces (Kerscher, 2007). Studies show lack of RNA binding proteins can lead to genomic instability, defects in cell cycle regulation and cancer (Lukong et al., 2008). RNA binding proteins may have some roles in cell attachments and cancer metastasis as well (de Hoog et al., 2004; Lukong et al., 2008).

1.4.1 *Some important groups of RNA binding proteins*

Important kinds of RNA binding proteins include STAR (signal transduction and activation of RNA), hnRNPs (heterogeneous nuclear ribonucleoproteins) and SR (Serine/arginine-rich) families (Sergeant *et al.*, 2007; Long and Caceres, 2009; Ehrmann and Elliott, 2010; Busch and Hertel, 2012). These groups of RNA binding proteins have fundamental roles in alternative splicing (Lukong and Richard, 2003) (figure 4).
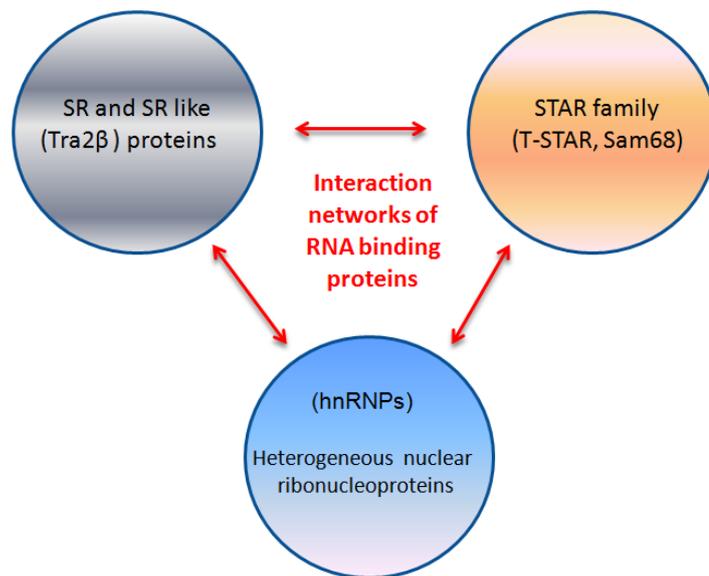
Figure 4: Some important groups of RNA binding proteins. The STAR (Signal Transduction and Activation of RNA) family (including T-STAR, Sam68 and SLM-1), SR (serine/arginine-rich protein) and SR-like proteins such as Tra2β and hnRNP (heterogeneous nuclear ribonucleoproteins) are some of important groups of RNA binding proteins. These proteins interact with each other.

1.4.2 *Roles of RNA binding proteins in alternative splicing*

Each RNA binding protein has specific RNA target sequences (Black, 2003; Matlin *et al.*, 2005). It has been confirmed that binding of some RNA-binding proteins to splicing enhancers and repressors within pre-mRNAs can control the splicing pathway (Black,

2003; Matlin et al., 2005). Modulations of nuclear RNA binding protein concentration in different cell types and tissues can control pre-mRNA splicing (Elliott *et al.*, 1997; Ehrmann *et al.*, 2008). For example, SR proteins usually promote and hnRNP proteins usually repress exon inclusion (Dreumont *et al.*, 2010). It has been shown that considerable changes in the expression of RNA binding proteins occur during male germ cell development (Elliott and Grellscheid, 2006). RNA binding protein activity can be controlled by some post-transcriptional modifications. For example, RNA binding protein activity is influenced by acetylation or phosphorylation of their specific sites (Zhang *et al.*, 2003; Lu *et al.*, 2005; Najib *et al.*, 2005).

1.4.3   Structure of RNA binding proteins

RNA binding proteins contain one or multiple RNA binding domains such as KH (hnRNP K homology) domain, cold-shock domain, RNA Recognition Motif (RRM), Sm domain, RGG (Arg-Gly-Gly) box, DEAD/DEAH box, dsRBD (double stranded RNA binding domain) and zinc finger (Glisovic et al., 2008). Generally, these domains are important for interactions between RNA binding proteins and their target RNA and also with other proteins. When misregulation of these interactions happens, different kinds of disease including cancer can occur (Huranova *et al.*, 2009). KH domain and RRMs bind single stranded RNA (Lunde *et al.*, 2007).

The RRM (RNA recognition motif) has an important role in mediating the binding of proteins to specific RNA sequences (Nagai *et al.*, 1995; Perez-Canadillas and Varani, 2001; Clery *et al.*, 2008). RNP1 and RNP2 are two conserved sequences within the RRM motif (Kielkopf et al., 2004). The RRM motif contains 4 β sheets (β1-β4) and 2 α helixes (α1 and α2) (Birney et al., 1993). RNP1 is located in β3 and RNP2 is present on β1 (Han et al.). The structure and function of the RRM is more characterized than the KH domain and the double stranded RNA binding domain (dsRBD) (Chen and Varani, 2005).

The KH domain has an important role in RNA binding activity of many RNA binding proteins. This domain was first characterized in a heterogeneous ribonucleoprotein

named hnRNP K (Siomi et al., 1993; Najib et al., 2005). There are three α helixes and β sheets in the structure of KH domain (Grishin, 2001).

1.5   STAR Family

The STAR family is a family of RNA binding proteins which link RNA processing to signalling transduction (Vernet and Artzt, 1997; Lukong and Richard, 2003). It has been shown that there are homologues for STAR proteins in different species such as GLD-1, HOW and QKI in *C.elegans*, *D. Melanogaster* and various mammals respectively (Baehrecke, 1997; Lo and Frasch, 1997; Zaffran et al., 1997; Larocque and Richard, 2005; Hansen and Schedl, 2006; Kimble and Crittenden, 2007). T-STAR (testis signal transduction and activation of RNA), Sam68 (Src-associated in mitosis 68 kDa) and SLM1 (Sam68-like mammalian protein 1) are members of the STAR family (Sergeant et al., 2007). The genes encoding these three proteins are called *KHDRBS3*, *KHDRBS1* and *KHDRBS2* respectively. These genes evolved from a common ancestral gene around 500 million years ago (Ehrmann et al., 2013). A STAR domain (including KH RNA binding domain, QUA1 and QUA2 domains) is present within the structure of these proteins. This STAR domain has important roles in protein interactions and also RNA processing.

In addition to the STAR domain there are other domains in the structure of STAR family members. For example, the tyrosine-rich C-terminal domain (containing candidate SH2 binding domains) and proline rich sequences (SH3 binding sites) are possibly involved in cell signalling pathways (Lukong and Richard, 2003; Ehrmann and Elliott, 2010; Sette, 2010; Sette et al., 2010). Phosphorylation of tyrosine residues can enable these STAR family members as a dock for signalling proteins which contain SH2 domains (Sanchez-Jimenez *et al.*, 2011).

1.6   SLM-1

SLM-1 (Sam68-like mammalian protein 1) is a STAR family protein. In mouse *SLM1* is located on chromosome 1 (Ehrmann and Elliott, 2010). This protein is also called KHDRBS2 (Di Fruscio et al., 1999). In addition to interaction with itself, SLM1 has interactions with Sam68, T-STAR and hnRNP G (Stoss et al., 2004). SLM-1 is only expressed in the brain and, in smaller amounts, in the testis (Stoss et al., 2004). In addition to a KH domain, the SLM-1 structure contains proline and tyrosine rich residues within its structure (Di Fruscio *et al.*, 1999).

1.7   T-STAR

T-STAR protein has similarities with Sam68. There are other names such as SLM-2 and KHDRBS3 for this protein. The *T-STAR* gene is located on chromosome 8 and 15 in human and mouse respectively (Venables *et al.*, 2004; Ehrmann and Elliott, 2010). The STAR family has important roles in splicing, localization and translation of mRNAs (Li *et al.*, 2000; Matter *et al.*, 2002; Wu *et al.*, 2002; Mootz *et al.*, 2004; Schumacher *et al.*, 2005; Edenfeld *et al.*, 2006; Paronetto *et al.*, 2007). T-STAR has important roles in signal transduction and RNA binding (Elliott, 2004). On the other hand, it seems that T-STAR may have some roles in cell division. It has been shown that ectopic expression of T-STAR in fibroblasts from chicken embryo increases the time of cell division (Lee and Burr, 1999). In addition, it has been shown that in immortalized cells, T-STAR is down regulated (Kool et al., 2001).

Similar to other members of the STAR family, T-STAR has a KH domain (Lukong and Richard, 2003). In addition, the KH domain is flanked by two other conserved domains called QUA1 and QUA2 (Volk et al., 2008) (figure 5). These three domains (QUA1, QUA2 and KH) make up the STAR domain (Maxi-KH region in figure 5). The KH domain of Sam68 and T-STAR are used for binding of these two proteins to RNA (Elliott, 2004). A difference with Sam68 is that T-STAR does not have an N-terminal domain which is present in Sam68. Amongst the similarities, there is an arginine and glycine rich region

(RG rich region) in both T-STAR and Sam68 proteins (Elliott, 2004). In T-STAR, within the arginine and glycine rich region there is one SH3-binding site (Ehrmann and Elliott, 2010). The C-terminal region of both T-STAR and Sam68 proteins contain conserved tyrosine residues (Elliott, 2004). These tyrosine residues can be phosphorylated by protein kinases and then act like a dock for signalling proteins which contain SH2 domains. There are proline rich sites in T-STAR and Sam68 which act like docking sites for proteins with SH3 domains (Ehrmann and Elliott, 2010).
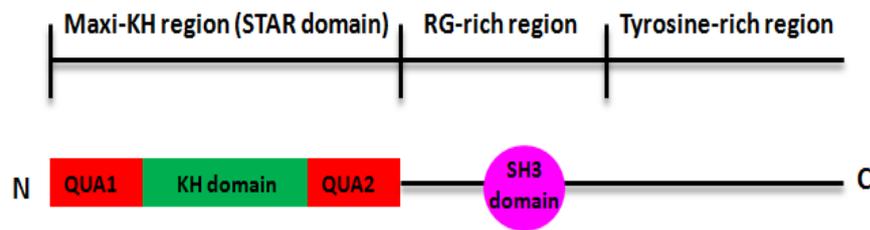
Figure 5: T-STAR structure. There is a STAR domain (Maxi-KH region) which contains QUA1, KH and QUA2 domains. Also T-STAR contains RG-rich and Tyrosine-rich regions in its structure. The picture is modified from (Elliott, 2004).

Unlike Sam68 which has ubiquitous expression, T-STAR is expressed only in the developing brain, adult testis and kidney (Venables *et al.*, 1999; Venables *et al.*, 2004). T-STAR protein is amplified in the medulloblastoma (Lu *et al.*, 2009). It has been proven that T-STAR interacts with hnRNP G and nuclear proteins which have effects on splicing activity such as SAFB (Hartmann et al., 1999; Stoss et al., 2001). It seems that both T-STAR and Sam68 interacts with hnRNP G-T and RBMY (Elliott et al., 1997; Elliott et al., 2000b; Venables et al., 2000). Interestingly, Sam68 and T-STAR have interactions with each other (Venables et al., 2000) and hnRNP G (Stoss et al., 2001). Interestingly, hnRNP G inhibits T-STAR splicing activity after co-expression (Stoss et al., 2001).

1.8   Sam68

Another member of the STAR family is Sam68 (Src-associated in mitosis 68 kDa) (Stoss *et al.*, 2001; Paronetto *et al.*, 2007; Chawla *et al.*, 2009). The *KHDRBS1* gene is located on chromosome 1 (Lukong and Richard, 2003) and 4 (Ehrmann and Elliott, 2010) in human and mouse respectively. *KHDRBS1* has 9 exons in human and mouse (Lukong and Richard, 2003). There are conserved sequences within *KHDRBS1* which are also present in other members of the STAR family (Najib et al., 2005). *KHDRBS1* has been cloned from different species. The sequences of important domains are highly conserved among these different species (Taylor and Shalloway, 1994; Richard *et al.*,

1995; Di Fruscio *et al.*, 1998; Fung *et al.*, 1998). Sam68 protein has different roles including contribution to the growth and invasion of cancer cells (Busa et al., 2007; Richard et al., 2008) plus regulation of apoptosis and progression of cell cycle (Taylor *et al.*, 2004; Paronetto *et al.*, 2007). Sam68 protein also has some roles in signal transduction , alternative splicing and viral RNAs transportation as well (Lukong and Richard, 2003).

1.8.1   Structure of Sam68

There are 443 amino acids in the Sam68 protein. The sequences of these amino acids are highly conserved among different species (Lukong and Richard, 2003). The KH domain is one of the important domains in Sam68 (Fumagalli et al., 1994). The KH domain plus the flanking regions (QUA1 and QUA2) make up the STAR domain (also called the Maxi KH region) (figure 6) (Elliott, 2004). Also, there are proline-rich motifs in Sam68 protein which facilitate binding of Sam68 to other proteins that contain an SH3 domain (figure 7) (Reddy, 2000). In addition, in Sam68 protein there are SH3-binding sites (figure 7) (Ehrmann and Elliott, 2010). There are 24 amino acids in the C-terminus of Sam68 called the NLS (nuclear localization signal) which have a role in the localization of this protein (Ishidate et al., 1997) (figure 6).
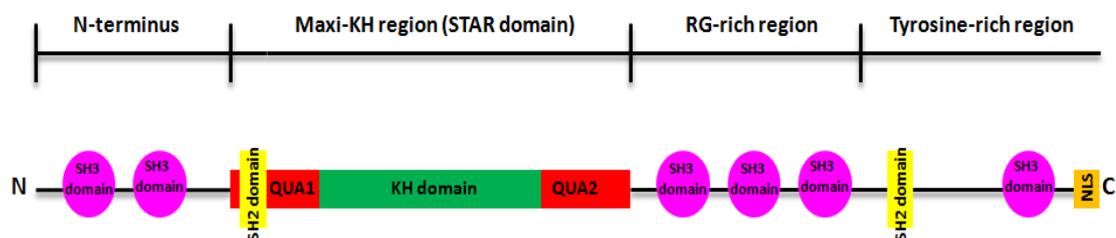
Figure 6*: Sam68 domains. As well as the maxi-KH region (STAR domain), there are RG-rich and Tyrosine rich regions in the structure of Sam68 proteins. Six SH3 and two SH2 domains respectively are present in this protein. This picture is modified from (Elliott, 2004).

Figure 7: Sam68 amino acid sequence and domains. Different domains are shown in different colours. Dark green: extended N-terminal domain of Sam68; green: proline rich sequences (SH3 binding sites); red: N-terminus; **bold and underlined**: Qua1 domain; **bold, blue and underlined**: KH Domain; blue font color with under line: Qua2 domain; yellow: RG-rich domain; purple: tyrosine-rich C-terminal domain; gray: Tyrosine residues (candidate SH2 binding sites). Picture is modified from (Ehrmann and Elliott, 2010).

1.8.2 Sam68 localization

Sam68 protein has a mostly nuclear localization, but initially was identified in the cytoplasm during mitosis (Rajan et al., 2008a). It has been shown that while methylated Sam68 is localized in the nucleus, hypomethylation of this protein leads Sam68 to be localized in the cytoplasm (Babic et al., 2004). In cancer cells, Sam68 protein has a general nucleoplasmic distribution and is also concentrated in nuclear SNB (SLM/Sam68 Nuclear Bodies) structures (Chen et al., 1999b).

Although SNBs function has not been determined yet, it seems that these structures contain nucleic acids, some splicing regulators, signalling components (Rajan et al., 2009) and hnRNP A1 interacting protein (HAP) which has roles in nuclear structure and transcription (Denegri et al., 2001). Sam68 seems to be exclusively nuclear in most somatic cells. Sam68 protein has been observed in the cytoplasm of secondary spermatocytes. In addition, Sam68 has some interactions with signalling molecules which are cytoplasmic (Lukong and Richard, 2003; Paronetto *et al.*, 2006; Paronetto *et al.*, 2009). It has been shown that after mutation of the Sam68 NLS (P439R, changing proline to arginine), the resulting protein will be localized in the cytoplasm (Reddy, 2000).

1.8.3 Sam68 modifications

There are different types of posttranslational modifications known for Sam68 such as: acetylation (Babic et al., 2004), methylation (Cote et al., 2003), tyrosine phosphorylation (Fumagalli et al., 1994; Taylor and Shalloway, 1994) and serine/threonine phosphorylation (Matter et al., 2002; Paronetto et al., 2006).

Methylation of arginine residues prevents interactions of Sam68 with proteins which have SH3 domains (Najib et al., 2005). Phosphorylation of Sam68 regulates some roles of Sam68 protein such as RNA binding ability and splicing activity (Haegebarth et al., 2004; Lukong et al., 2005). There is a positive correlation between acetylation of

Sam68 lysine residues and RNA binding activity of this protein (Babic et al., 2004). There are some proline rich domains in Sam68 that interact with proteins which contain SH3 domains. Furthermore, there is a tyrosine rich C-terminal domain in Sam68. This domain is a substrate for tyrosine kinase enzymes. Therefore, tyrosine phosphorylation of Sam68 can lead to the interaction of this protein with proteins which contain SH2 domains (Najib et al., 2005). The tyrosine-rich site can be phosphorylated by tyrosine kinases. Phosphorylation of Sam68 is important for its roles in signal transduction within the cell (Vernet and Artzt, 1997).

1.8.4 Sam68 interactions

There are some interactions between Sam68 and other RNA binding proteins (Venables *et al.*, 1999; Elliott *et al.*, 2000a). Sam68 protein is able to have protein-protein and protein-RNA interactions through some of its consensus motifs (Lukong and Richard, 2003). Sam68 RNA binding activity has been shown to be negatively regulated by binding of Sam68 with SH3 domain containing proteins (Chen *et al.*, 1997; Fusaki *et al.*, 1997; Derry *et al.*, 2000).

Diverse proteins which have important roles in cell signalling and RNA processing interact with Sam68. It has been reported that nuclear transcriptional regulators and hnRNP A1 which has a splicing repressive role, interact with Sam68 (Hong et al., 2002; Paronetto et al., 2007; Rajan et al., 2008b) . It has been shown that Sam68 and hnRNP L have both a spatial and physical association within SNBs, but they do not help or antagonize each other's activity. Conversely, hnRNP G which is generally nucleoplasmic rather than SNB in location, can inhibit some of Sam68 splicing activity (Rajan et al., 2009). Although some proteins which interact with Sam68 have been identified, it is likely that there are other interacting proteins still to be identified (Rajan et al., 2009). Identification of these undetected interacting protein partners can lead to better understanding of Sam68 roles within the cells.

1.8.5   Roles of Sam68 in development, cell division and cancer

Sam68 plays a key role in regulation of pre-mRNA splicing during neural development and neurogenesis (Chawla et al., 2009). Over expression of Sam68 with a deletion of its KH domain leads to inhibition of DNA synthesis. This variant is particularly expressed at growth arrest (Barlat et al., 1997). It has been shown that in NIH3T3 cells, after ectopic expression of Sam68 with KH deletion domain, *cyclin D1/CCND1* expression is down regulated and cells will be stopped in the transit from G1 to S phase (Barlat et al., 1997) These results suggest a role for Sam68 in the G1/S transition stage of the cell cycle (Reddy, 2000). STAR family members including Sam68 have important roles in cell cycle regulation. While expression of ΔKH Sam68 leads to cell growth arrest, expression of the full length of this protein in cell lines with limited level of expression of Sam68 can promote S phase (Barlat *et al.*, 1997). At first it was thought that Sam68 has a tumour suppressor role, but recently it has been found that this protein plays a pro-oncogenic function (Lukong and Richard, 2007). Some RNA binding proteins have altered expression in cancer. SRSF1 (Karni *et al.*, 2007) and Tra2β (Best *et al.*, 2013) are up-regulated in different kinds of cancers, and Fox2 shows down regulation in ovarian cancer (Venables *et al.*, 2009). It has been reported that in prostate and breast tumours Sam68 is over expressed (Elliott and Rajan; Busa *et al.*, 2007; Rajan *et al.*, 2008b; Elliott and Rajan, 2010). Sam68 protein is up-regulated in thyroid and liver cancers as well (Busa and Sette, 2010).

Sam68 knockout mice, in addition to bone metabolism problems, have defects in locomotion and mesenchymal cell differentiation which imply developmental roles of this protein (Richard *et al.*, 2005; Lukong *et al.*, 2008; Richard *et al.*, 2008). Unlike T-STAR which is not essential for male infertility (Ehrmann *et al.*, 2013), Sam68 has important roles in male germ cell development. It has been shown that male Sam68 knockout mice are infertile (Ehrmann and Elliott, 2010; Sette *et al.*, 2010). In addition, being a splicing regulator is another role for Sam68 which is important in neural development (Chawla et al., 2009).

1.8.6   Roles of Sam68 in gene expression, cell signalling and RNA processing

Sam68 has an essential role in linking transcriptional and post-transcriptional regulation of gene expression to signalling pathways (Rajan et al., 2008a). Sam68 participates in signal transduction pathways controlled by tyrosine kinases. Sam68 protein acts like a dock for recruiting other signalling molecules, acting like an adaptor in signalling pathways (Najib et al., 2005). Cytoplasmic Sam68 acts to link proteins to each other by using SH2 (phosphotyrosine) and SH3 (polyproline) binding domains. By acting as an adaptor, Sam68 has important roles in signal transduction (Elliott, 2004). For example, after interaction of insulin with its receptor, the insulin receptor plus Sam68 will be tyrosine phosphorylated. These proteins are then able to interact and activate the PI3K (phosphatidylinositol-3-kinase) p85 subunit. SH2 domains of p85 subunit interact with phosphorylated Sam68. Signalling complexes will be then formed by PI3K pathway. Also, phosphorylated Sam68 has association with p120GAP to link the phosphatidylinositol-3-kinase pathway to other signalling pathways. In summary, insulin stimulation leads to interaction of Sam68 as a dock with PI3K, initiating an intracellular signalling cascade to control glucose levels (Kahn, 1994; Sanchez-Margalet and Najib, 2001; Niswender and Schwartz, 2003).

One of the most important roles of Sam68 is in regulation of RNA processing. SELEX experiments (systematic evolution of ligands by exponential enrichment) have shown Sam68 protein binds specifically to UAAA or UUUA as high–affinity binding sequences (Lin et al., 1997). Also, Sam68 has some roles in nuclear export and cytoplasmic processing of viral mRNAs (Reddy *et al.*, 1999; Cote *et al.*, 2003). Sam68 also is involved in alternative splicing. A role of Sam68 protein has been shown in activating inclusion of exon V5 in CD44 mRNA (Matter et al., 2002).

Different isoforms of CD44 can be produced by inclusion or exclusion of variant exons. Alternative splicing of CD44 is important in cancer and development (Cooper and Dougherty, 1995; Sherman *et al.*, 1996; Sherman *et al.*, 1998). The Ras signalling pathway is involved in regulation of exon V5 inclusion. Sam68 which is a target in this

signalling pathway can bind to exonic splicing enhancers within the V5 exon. Sam68 phosphorylation can lead to inclusion of exon V5 in CD44 mRNA (Matter *et al.*, 2002). In addition, Sam68 protein is necessary for alternative splicing of mRNAs in neuronal cell lines (Chawla et al., 2009).

1.9   HnRNPs

Another group of RNA binding proteins are called HnRNPs (heterogeneous nuclear ribonucleoproteins). These proteins have key roles in the regulation of transcript packaging, alternative splicing and translation (Han et al., 2010). Also, members of this family are involved in RNA metabolism (Weighardt et al., 1996). HnRNP G has been reported as a spliceosome component (Rappsilber *et al.*, 2002; Heinrich *et al.*, 2009). Initially, hnRNPs A, B and C with their associated RNAs were separated and identified using sucrose density gradients (Beyer et al., 1977).

The number of identified hnRNPs were increased using UV cross-linking experiments (van Eekelen et al., 1981). Using immunopurification accompanied with two-dimensional gel electrophoresis led to a complete list of these kinds of RNA binding proteins (Pinol-Roma et al., 1988) comprising hnRNPs A–U (figure 8). More information about the sequences and structure of hnRNPs was achieved by using monoclonal antibodies. Furthermore, using immunostaining has proved that contrary to SR proteins and snRNPs (Small nuclear ribonucleoproteins) which have a speckled distribution in the nucleus, hnRNPs have diffuse pattern. HnRNPs have ubiquitous expression in different tissues (Kamma *et al.*, 1995).

1.9.1   Structure of hnRNPs

In structure, hnRNP proteins contain at least one RNA binding motif and other domains involved in their localisations and protein interactions (Dreyfuss *et al.*, 2002). There are

a number of distinct RNA binding motifs in hnRNP proteins including the KH domain, RGG box and RRM (figure 8). In addition, there are some domains to mediate protein-protein interactions (Dreyfuss et al., 2002). Each of these motifs, including the RRM, is not present in all hnRNPs (figure 8) (Han *et al.*, 2010). In hnRNP G there is an N-terminal RRM (figure 9) (Soulard et al., 1993). There is a motif in human RBMY called SRGY which is enriched in serine, arginine, glycine and tyrosine amino acids (figure 9). This sequence in Human RBMY motif is longer in comparison with mouse (Elliott et al., 1996; Mahadevaiah et al., 1998; Venables et al., 1999). There is also an RRM in RBMY (figure 9). The identity of this motif between human and mouse is around 74% (Mahadevaiah et al., 1998). Between human and mouse RRM there are 22 amino acid differences (Skrisovska et al., 2007). Most of the active copies of *RBMY* have been mapped to the AZFb region (Elliott *et al.*, 1997; Elliott, 2000). Some genes involved in spermatogenesis which are located on long arm of Y chromosome. AZFb is one of the regions on Y chromosome which is important for spermatogenesis and microdeletion of this region can lead to male infertility (Lahn and Page, 1997; McElreavey and Krausz, 1999). In this AZFb region six functional *RBMY* copies have been found (Skaletsky et al., 2003).
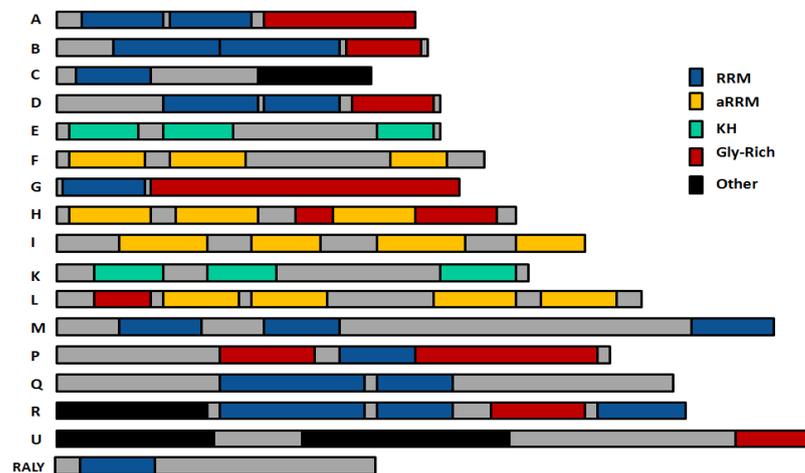
Figure 8: HnRNP structures. Most of these proteins contain an RRM (RNA recognition motif). Also KH and Gly-Rich domains are some of the common domains in the structure of hnRNPs. (Han *et al.*, 2010).
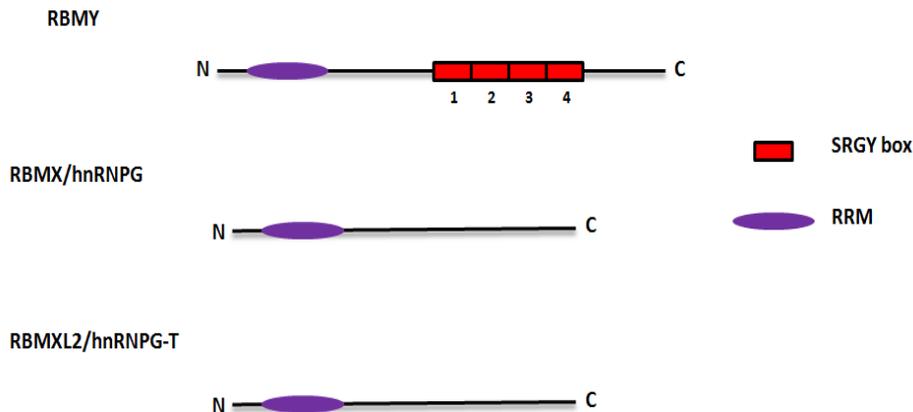
Figure 9: Structures of RBMY, hnRNP G and hnRNP G-T proteins. All of these three proteins contain an RRM. However only RBMY contains an SRGY box. (Elliott, 2004).

1.9.2 Role of hnRNPs in alternative splicing, mRNA transportation and localization

The hnRNP proteins have important roles in splicing (Min *et al.*, 1995; Chen *et al.*, 1999a; Chou *et al.*, 1999; Tange *et al.*, 2001), and transport of RNA from the nucleus to the cytoplasm (Liu and Mertz, 1995; Lee *et al.*, 1996; Izaurralde *et al.*, 1997; Gallouzi and Steitz, 2001), translation, and localization of mRNA (Dreyfuss et al., 2002). It has been shown that some hnRNP proteins shuttle between the nucleus and the cytoplasm (Alzhanova-Ericsson et al., 1996; Caceres et al., 1998). While RBMX has roles in selection of splicing sites in somatic cells, it seems that RBMY and hnRNP G-T have the same roles in germ cells (Elliott, 2004). For example, human RBMY is present in locations where gene expression is active. Also, the RBMY protein can interact with and affect SR protein activity or by preventing their access to their target mRNA, has some roles in splicing and can influence on this process (Dreumont et al., 2010).

1.9.3   Some interactions of hnRNP family members

The binding of hnRNP proteins (to target RNA) has some effects on interactions and the fate of transcripts (Dreyfuss et al., 2002). The RBMY and hnRNP G-T proteins have interactions with other RNA binding proteins such as SR and SR-related families and STAR family (Elliott *et al.*, 2000a; Elliott *et al.*, 2000b; Venables *et al.*, 2000; Venables *et al.*, 2005). For example, RBMY has direct interaction with Sam68 and T-STAR (Skrisovska et al., 2007). Also, Tra2β (an RNA binding protein, SR-like protein) and hnRNP G directly interact with each other (Venables et al., 2000). The hnRNP G protein action is synergistically or antagonistically with Tra2β. This action depends on the mRNA substrate (Hofmann and Wirth, 2002; Nasim et al., 2003). For example while hnRNP G and Tra2β have cooperation in splicing of *SMN2* transcript (Elliott, 2004), they usually show opposite effects on splicing of dystrophin transcripts and slow skeletal α-tropomyosin (Nasim *et al.*, 2003).

Human RBMY interactions with some SR or SR-like proteins and its localization at the periphery of nuclear speckles have been recognized (Dreumont *et al.*, 2010). It has been shown that hRBMY interactions with SRSF3 or Tra2β can influence splicing decisions (Elliott et al., 2000a; Venables et al., 2000).

1.9.4   HnRNP G (RBMX), hnRNP G-T (RBMXL2) and RBMY

The hnRNP G (encoded by *RBMX*), hnRNP G-T (encoded by *RBMXL2*) and RBMY proteins play important roles in nuclear RNA processing (Elliott, 2004). *RBMX* is the X-homologue of *RBMY* (RNA-binding motif gene on Y chromosome). Both *RBMX* and *RBMY* were originally a pair of autosomal genes (Delbridge et al., 1997; Mazeyrat et al., 1999) and encode proteins which have similar structures. The hnRNP G protein has an N-terminal RRM and domains rich in glycine, arginine and serine (Elliott, 2004). The *RBMX* gene has ubiquitous expression except during X chromosome inactivation in

meiosis. It is thought that the *RBMX* acts as a housekeeping gene (Elliott and Grellscheid, 2006). The expression level of *RBMX* gene is different based on a tissue specific manner (Nasim et al., 2003).

The hnRNP G protein has a nuclear localisation, and shows a speckled pattern in the nucleus. Using the electron microscope hnRNP G was found to be localised within the ICGCs (interchromatin granule clusters) (Soulard et al., 1991). The ICGCs seem to be storage sites which contain different proteins including splicing factors (Lewis and Tollervey, 2000; Spector, 2001; Sutherland et al., 2001). HnRNP G was implicated as a splicing factor using purification and proteomic analysis of spliceosome (Rappsilber et al., 2002). The hnRNP G protein also interacts with other proteins involved in splicing such as Tra2β (Venables *et al.*, 2000).

Another member of the hnRNP family is hnRNP G-T protein. HnRNP G-T is encoded by an autosomal retrogene, called *RBMXL2*, which is located on human chromosome 11p15. The hnRNP G-T protein is expressed only in the testis. It has been shown that the *RBMXL2* gene was derived from transcript retrotransposition of the *RBMX* gene (Elliott et al., 2000b). In human, there are 9 other retrogene copies of *RBMX* as well as *RBMXL2*. These include *RBMXL1* and *RBMXL9* which are located on chromosome 1 and 9 respectively (Lingenfelter et al., 2001). It has been proven by RT-PCR that, while *RBMXL1* expression is common in all tissues including the testis, *RBMXL9* is only expressed in the brain and testis. The presence of different retrogenes of *RBMX* gene may imply their roles in replacement of the *RBMX* gene function during meiosis (Ehrmann et al., 2008).

Although the proteins encoded by the *RBMX* retrogene copies are very similar to hnRNP G, the hnRNP G-T protein is only 73% identical to the hnRNP G protein. This means that *hnRNP G-T* evolved earlier or has been under less selective constraint. The hnRNP G-T protein is more similar to the hnRNP G protein in comparison to the protein encoded by *RBMY*. The RBMY protein is around 50% identical to the hnRNP G protein

(Elliott, 2004). In addition to interacting with Sam68, T-STAR, both hnRNP G-T and RBMY proteins interact with other RNA binding proteins including: SR proteins (SRSF3) and SR-like proteins (Tra2β) (Elliott *et al.*, 2000a; Elliott *et al.*, 2000b; Venables *et al.*, 2000; Venables *et al.*, 2005).

1.9.5   Roles of hnRNP G, hnRNP G-T and RBMY

The hnRNP G protein cooperates with other RNA binding proteins to perform important functions in splicing regulation of some target genes including the *SMN*2 gene which can be considered as a target for treatment (gene therapy) of the SMA (proximal spinal muscular atrophy) disease (Elliott, 2004). It also regulates splicing of *Tau* gene. Splicing misregulation and abnormal ratios of *Tau* protein isoforms can lead to frontotemporal dementia with Parkinsonism (FTDP-17) (Wang *et al.*, 2011). In frogs and zebrafish, the hnRNP G proteins play important roles in neural development pathway (Tsend-Ayush *et al.*, 2005; Dichmann *et al.*, 2008).

In mammals, during male meiosis a process called Meiotic Sex Chromosome Inactivation happens in which sex chromosomes become inactivated. During this time, autosomal retrogenes of X-linked genes play an important role (Turner *et al.*, 2005; Wang *et al.*, 2005; Turner *et al.*, 2006). It has been shown that human hnRNP G-T protein is highly expressed in the nuclei of pachytene spermatoctes and round spermatids (Elliott et al., 2000b). This suggests the hnRNP G-T protein is replacing the RBMX function when RBMX is inactivated during and after XY body formation. While hnRNP G plays a role in regulation of splice site selection in somatic cells, the same role can be considered for hnRNP G-T and RBMY proteins in germ cells (Elliott, 2004).

Specific expression of hnRNP G-T only in male germ cells suggests the role of this gene as a candidate gene in male infertility (Elliott, 2004). In mice, haploinsufficiency of hnRNP G-T protein prevents normal spermatogenesis suggesting a fundamental role in

meiotic and post-meiotic germ cell development (Ehrmann et al., 2008). By analogy, the *HNRNP G-T* gene might have an important role in human infertility. It has been shown that in men with impaired fertility there are some variations in the sequence of this gene (Westerveld et al., 2004). In addition, in some patients with incomplete spermatogenesis, hnRNP G-T is not expressed (Maymon et al., 2002; Westerveld et al., 2004).

Expression of human RBMY (hRBMY) is particularly found in nuclei of male germ cells during spermatogenesis. It has been shown that deletion of the *RBMY* copies leads to arrest of meiotic division I in spermatogenesis (Elliott et al., 1997). In mouse, reduction in numbers of *RBMY* gene copies leads to sperm abnormalities (Mahadevaiah et al., 1998). Therefore, conservation of *RBMY* in all mammals implies an important role in development of male germ cells (Delbridge et al., 1997). Unlike in human, mice with deficiency of *RBMY* expression have abnormal sperm development but they are still fertile (Mahadevaiah et al., 1998).

1.10 SR proteins and Tra2β

SR proteins are splicing regulators (Manley and Tacke, 1996; Valcarcel and Green, 1996). SR proteins contain one domain rich in serine and arginine dipeptides and normally one or two RNA binding domains (Fu, 1995; Matlin *et al.*, 2005; Long and Caceres, 2009) (figure 10). Unlike hnRNP proteins which most of the time bind to splicing silencers to inhibit splicing, SR proteins usually activate splicing by binding to the splicing enhancer sequences (Matlin et al., 2005). It has been shown that in the splicing process, SR proteins are antagonistic partners of hnRNP A1 (Graveley, 2000; Long and Caceres, 2009). After binding of SR proteins to RNA, binding of other components of the spliceosome are facilitated. It has been shown that SR proteins bind to exonic splicing enhancers (Zheng, 2004). Serine and arginine rich domains have important role in protein-protein interactions (Bourgeois et al., 2004).

In flies there is a prototypical alternative splicing regulator called Tra2 which has important role in sexual differentiation and spermatogenesis. There is a homologue gene of *Tra2* in human called *TRA2B* (encoding Tra2β protein) (Black, 2003). There are two protein isoforms for Tra2β called Tra2β-1 and Tra2β-3 (Stoilov et al., 2004). While most SR proteins contain 1-2 RRM domains and only one RS domain, Tra2β (SR-like protein) contains two RS domains in addition to one RRM. Tra2β contains RS-rich region which acts like a splicing activator and binds to purine-rich (GAA) repeats within the exons (Tacke *et al.*, 1998). Tra2β binds to several RNA sequences (Tsuda *et al.*, 2011).

SR proteins such as ASF/SF2, SRSF3, SRp30 and the SR-like protein Tra2β act as important splicing factors (Venables et al., 2000) and have ubiquitous expression (Elliott, 2004). Their presence influences on the selection of splice sites (Matlin et al., 2005).
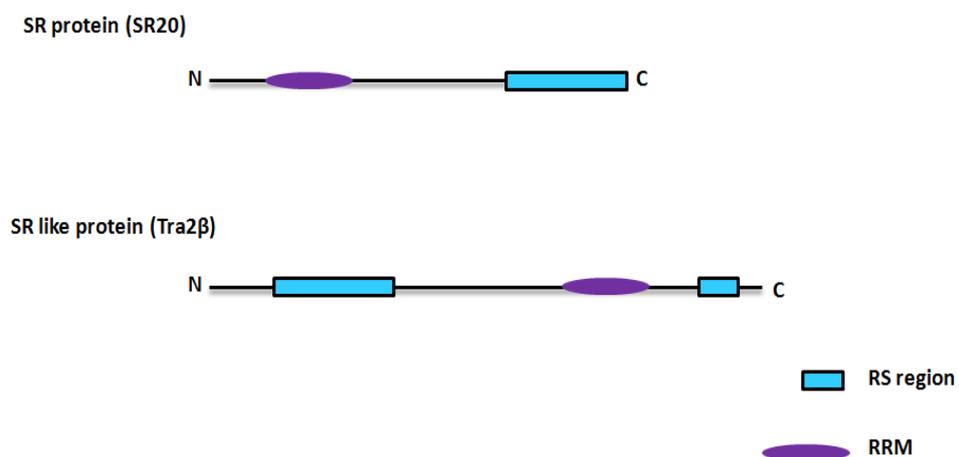
Figure 10: Structure of SR and SR-like proteins (Elliott, 2004). In addition to an RRM motif, RS domains (which contain arginine and serine residues) are present at the N-terminal and C-terminal of Tra2β (Sciabica and Hertel, 2006).

1.10.1 Phosphorylation of SR proteins by splice factor kinases (SRPKs and CLK1 families)

The SR proteins can be phosphorylated by splice factor kinases such as SRPK (SR protein kinase) and CLK (Cdc2-like kinase) families. While SRPK1 and SRPK2 belong to SRPK family, CLK1, CLK2, CLK3 and CLK4 are members of CLK family (Stojdl and Bell, 1999). The modification of protein sequences and the extent of phosphorylation carried out by members of SRPK and CLK families can be different (Colwill *et al.*, 1996a; Ngo *et al.*, 2005; Velazquez-Dones *et al.*, 2005; Bullock *et al.*, 2009).

Serines in the RS domain of Serine/arginine-rich proteins are the target for these phosphorylations. RS domains phosphorylation affects the function of serine/arginine-rich proteins in the spliceosome (Wu and Maniatis, 1993; Kohtz *et al.*, 1994). Phosphorylation of RS domains also plays important roles in alternative splicing regulation of pre-mRNA (Zhu and Krainer, 2000; Massiello and Chalfant, 2006), cellular distribution of SR proteins (Gui *et al.*, 1994; Duncan *et al.*, 1995; Colwill *et al.*, 1996b), Interaction of the nuclear import receptor with SR proteins and also transportation of serine/arginine-rich proteins (Caceres *et al.*, 1997; Kataoka *et al.*, 1999; Lai *et al.*, 2000). The nuclear import and nuclear distribution of SR proteins can be regulated by SRPKs and CLK families respectively (Ghosh and Adams, 2011).

1.10.2 Similarities between SR proteins and hnRNPs

There are many similarities between SR proteins and hnRNPs (Han *et al.*, 2010). For example, there are multiple RNA binding domains such as RRM and auxiliary domains in both of these proteins (figure 11). Furthermore, both groups of proteins can act like shuttles to export mRNAs to the cytoplasm (Gorlach *et al.*, 1993; Good, 1995; Caceres *et al.*, 1998; Borgeson and Samson, 2005). Both groups of SR and hnRNP proteins also

have important roles in the regulation of post-transcriptional events (Singh and Valcarcel, 2005; Samson, 2008). It has been shown that there is an overlap in localization of hnRNP G with SR proteins (Venables et al., 2000).
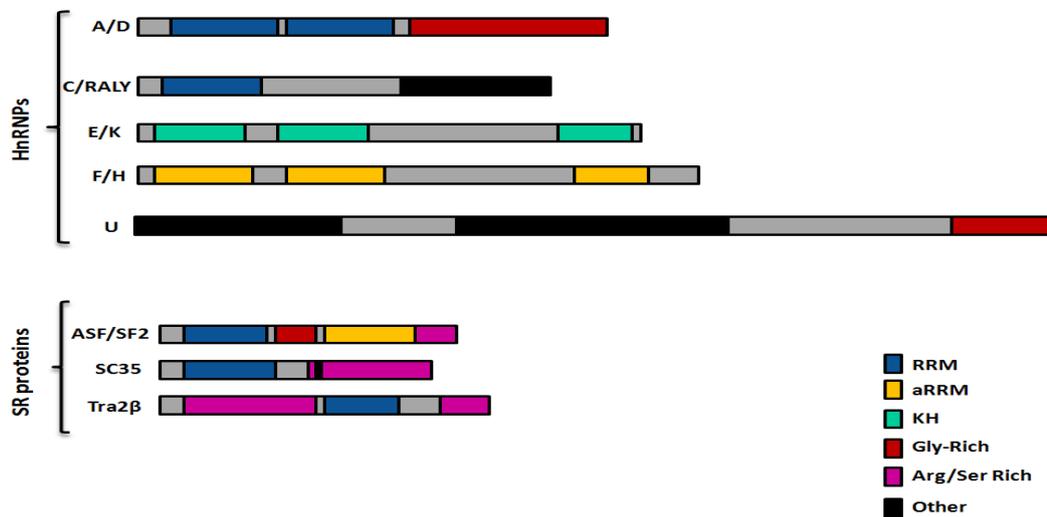
Figure 11: Structural similarities between hnRNPs and SR proteins. RRM and Gly-rich domains are present in the structure of most members of hnRNPs and SR proteins. (Han *et al.*, 2010)

1.10.3 Some roles of Tra2β

Tra2β acts as a splicing activator (Elliott, 2004). The exon inclusion percentage of target mRNA will be higher when the concentration of this protein is increased (Grellscheid *et al.*, 2011; Elliott *et al.*, 2012). The *TRA2A* and *TRA2B* genes are two homologues genes of *Tra2* (Tacke et al., 1998). Tra2α and Tra2β proteins have more than 60% amino acid identity (Elliott et al., 2012). It has been shown that if there is *Tra2B* deletion during embryogenesis, Tra2α is not able to compensate the vital roles of Tra2β and *Tra2B* $^{-/-}$ mouse dies during gestation (Mende et al., 2010).

It seems that the lack of Tra2β protein and its role in alternative splicing regulation is involved in the death of *Tra2B* $^{-/-}$ mice embryos (Elliott *et al.*, 2012). Tra2β plays a

critical role in embryonic brain development as well (Grellscheid et al., 2011). Also, Tra2β protein plays a role as a splicing activator of *SMN1* and *SMN2* mRNAs. Deletion of *SMN1* leads to SMA. *SMN2* exon 7 is mostly excluded from the mRNA which leads to production of unstable protein. Over-expression of Tra2β activates inclusion of this exon which makes functional SMN2 protein. Afterwards, SMN2 protein can compensate roles of SMN1 protein. Rescue of SMN2 splicing could be considered as a potential gene therapy in the future (Hofmann *et al.*, 2000; Elliott, 2004).

1.10.4 *Expression and interactions of SR proteins and Tra2β*

Tra2β and SR proteins are expressed ubiquitously (Elliott et al., 2000a; Venables, 2002). Although *TRA2B* mRNAs can be found ubiquitously in the cells, the highest expression of this protein is in the testis and developing brain (Beil *et al.*, 1997; Nayler *et al.*, 1998; Venables *et al.*, 2000; Chen *et al.*, 2003). The *TRA2B* transcript is amplified in different kinds of human cancer including: stomach, uterus, ovary and lung cancers (Best *et al.*, 2013). Also, Tra2β is up-regulated in ovarian, cervical, breast and colon cancers (Fischer *et al.*, 2004; Watermann *et al.*, 2006; Gabriel *et al.*, 2009; Kajita *et al.*, 2013). By knocking down *TRA2B*, the percentage of cell apoptosis is increased in colon cancer (Kajita *et al.*, 2013).

The phosphorylation of serine residues in the RS domain has an important role in controlling protein interactions of SR proteins with other proteins and RNAs (Xiao and Manley, 1997; Yeakley et al., 1999). The phosphorylation of Tra2β strengthens protein interaction, which shows the importance of the RS domain. The RNA binding activity of Tra2β has been shown to be reduced by its hyperphosphorylation (Venables et al., 2000).

1.11 Aims of project

The main aim in this project was to analyse the role of RNA binding proteins T-STAR, Sam68, hnRNP G, hnRNP G-T and Tra2β and to identify their interacting protein partners. Investigation of these proteins will help us to understand their roles and importance within the cells, and specifically in different kinds of diseases such as cancer:

1.11.1 First aim:

The first aim was to create stable cell lines expressing epitope tagged T-STAR, Sam68, hnRNP G, hnRNP G-T and Tra2β proteins. This was achieved by cloning the open reading frame (ORFs) encoding these RNA binding proteins (T-STAR, Sam68, hnRNP G, hnRNP G-T and Tra2β) into the pcDNA5 vector. Then transfection of these open reading frames to HEK293 knock in cell lines was carried out to enable integration into the genome. The next step was selection of stable cell lines by using blastocidin and hygromicin antibiotics. After setting up the stable cell lines, tetracycline antibiotic was used to increase the protein expression of genes of interest (*T-STAR*, *Sam68*, *hnRNP G*, *hnRNP G-T* and *TRA2B)*. As the cell lysate was used for immunoprecipitation (IP) of tagged proteins there was a possibility that some proteins which do not interact with each other normally in vivo could have the opportunity to interact within the cell lysate (Mili and Steitz, 2004). Specifically there was a possibility that by over-expression of the FLAG-tagged proteins within the cells false interactions could be detected due to the high level of protein expression. In other words, this approach could lead to interaction of some proteins with other proteins which were not original protein partners. To solve this problem after detection of the interaction between proteins, the validity of interactions were also tested at the endogenous level of protein expression.

1.11.2 Second aim:

The second aim of the project was to use these stable cell lines to investigate the cell biology and interaction partners of these proteins. This included making stable cell lines expressing the wild type and P439R Sam68 proteins to compare the localisation of these proteins and their protein interacting networks. After identification of the interacting proteins, the next step was to investigate the possible roles of new interacting protein candidates within the cells such as their roles in alternative splicing regulation of selected target genes.

1.11.3 Third aim:

The final aim of the project was to use these stable cell lines to investigate RNA processing pathways. In particular, some morphological changes occurred after induction of cells expressing hnRNP G-T protein. Therefore, one of the aims of project was focused on the roles of hnRNP G-T protein in: A) Morphological changes B) Regulation of cell proliferation

Chapter 2

2   Identification of protein partners of T-STAR, wild type and P439R Sam68

2.1   Introduction

The STAR family proteins have known important roles, including RNA processing, signal transduction (Vernet and Artzt, 1997; Lukong and Richard, 2003), splicing, localization and translation of mRNAs (Li et al., 2000; Matter et al., 2002; Wu et al., 2002; Mootz et al., 2004; Schumacher et al., 2005; Edenfeld et al., 2006; Paronetto et al., 2007). However we need to find out more information about their interaction networks and pathways to fully understand these proteins' roles.

In this Chapter, the main focus will be on the T-STAR and Sam68 proteins and the proteomic approach carried out to identify their protein partners.

Sam68 protein has a mostly nuclear localization. Initially Sam68 was identified in the cytoplasm during mitosis (Rajan et al., 2008a). There are 24 amino acids in the C-terminus of Sam68 called an NLS (nuclear localization signal) which is essential for the localization of this protein (Ishidate et al., 1997). By making mutation of Sam68 (P439R Sam68, changing proline to arginine), the resulting protein will be localized in the cytoplasm (Reddy, 2000). Therefore, one of the aims of project was to see whether or not changing the localization of a mostly nuclear Sam68 to a cytoplasmic one would change the Sam68 interacting protein partners' network. Figure 12 is a schematic design showing the difference between the wild type and P439R Sam68 proteins.

Figure 12 : Difference between wild type and P439R mutant Sam68. The only difference between the wild type and P439R mutant Sam68 proteins was a point mutation in the codon encoding amino-acid 439 (changing proline to arginine). This amino acid substitution is highlighted by pink. CCA encodes proline whereas CGA encodes arginine. After making this mutation of Sam68 (P439R, changing proline to arginine), the resulting protein localized in the cytoplasm of HeLa cells rather than the nucleus (Reddy, 2000).

To achieve this aim of project the open reading frames of the T-STAR, wild type and P439R mutant Sam68 proteins were cloned into the pcDNA5 vector. Secondly, the

stable cell lines expressing these FLAG tagged RNA binding proteins were made. The next step was purification of the FLAG tagged T-STAR, wild type and P439R mutant Sam68 proteins plus their partners using immunoprecipitation. Finally, immunoprecipitated proteins were identified by mass spectrometry. RHA (RNA helicase A) was detected as the new common interacting protein partner of T-STAR, wild type and P439R Sam68. Furthermore, CCAR1 (Cell division cycle and apoptosis regulator protein 1) was detected as the new candidate interacting protein partner of the wild type and P439R Sam68 proteins. Interestingly it was found out that both wild type (nuclear) and P439R (cytoplasmic) Sam68 have the same interacting protein partners.

Also described in this Chapter are some investigations into the identified endogenous interacting protein partners of T-STAR, wild type and P439R mutant Sam68 to confirm their interactions.

2.2 Methods and Materials

2.2.1 Standard molecular biology

2.2.1(a) PCR

T-STAR, wild type and P439R *Sam68* were PCR amplified with the combination of reagents described in table 1 and 2 respectively. Phusion DNA polymerase, dNTPs and primers were from Finnzymes, Promega and Integrated DNA Technology (IDT) companies respectively. Table 3 shows the primers used in this project.

Reagent	Quantity for a 50µl reaction
dH$_2$O	31.5µl
5X GC Buffer (contains KCl, Tris-HCl and MgCl2)	10µl
10µM dNTPs (Deoxyribonucleotide triphosphates)	1µl
10µM Forward primer	2.5µl
10µM Reverse primer	2.5µl
Template DNA (~100ng)	2µl
Phusion DNA Polymerase	0.5µl

Table 1: PCR recipe for amplification of *T-STAR*. Here the reagents used in a typical 50µl reaction are shown.

Reagent	Quantity for a 50µl reaction
dH₂O	16.5µl
Template DNA (~100ng)	2µl
ImmoMIX	25µl
10µM Forward primer	3µl
10µM Reverse primer	3µl
Phusion DNA Polymerase	0.5µl

Table 2: PCR recipe for making wild type and P439R *Sam68*. Here the reagents used in a typical 50µl reaction are shown.

Primer name	Sequence (5'-3')	Restriction Enzyme
T-STARBGLIIF	AAAAAAAAAAGATCTATGGAGGAGAAGTACCTGCC	*Bgl*II
T-STARXHOIB	AAAAAAAAACTCGAGTCAGTATCTGCCATATGGCTGGT	*Xho*I
MSam68 F-BamHI	AAAAAAAAAGGATCC ATGCAGCGCCGGGACGATCCT	*BamH*I
Wt MSam68 R-SalI	AAAAAAAAAGGATCCTTAATAACGTCCATATGGATGCTCTCTGTATG	*Sal*I
Mutant MSam68R-SalI	AAAAAAAAAGTCGACTTAATAACGTCCATATCGATGCTCTCTGTATG	*Sal*I
pCDNArev	ACAATGCGATGCAATTTCCTC	

Table 3: Primers used for PCR amplification of *T-STAR,* wild type and P439R *Sam68*.

A DNA Engine DYAD thermal cycler (MJ Research) was used for PCR amplifications. The amplification protocol used for amplifying *T-STAR*, wild type and P439R *Sam68* are shown in table 4 and table 5 respectively.

1	Heat Activation	98 °C	30 seconds		
2	Denaturation	98 °C	10 seconds		
3	Annealing	60 °C	20 seconds		
4	Extension	72 °C	1 minute	Cycling to step 2	30
5	Final Extension	72 °C	4minutes		
6	Cooling	15°C	∞		

Table 4: PCR Programme for *T-STAR*.

1	Heat Activation	95 °C	10 minutes		
2	Denaturation	95°C	30 seconds		
3	Annealing	57.2°C	30 seconds		
4	Extension	72 °C	1 minute	Cycling to step 2	25
5	Final Extension	72 °C	10 minutes		
6	Cooling	10°C	∞		

Table 5: PCR Programme for PCR amplification of wild type and P439R *Sam68*.

2.2.1(b) Restriction digests

PCR products and the pcDNA5 vector were incubated for three hours with appropriate restriction enzymes separately for the first and second digestions. The restriction enzymes *BamH*I and *Xho*I were used for digestion of plasmid DNA (pcDNA5). The combinations of enzymes (*Bgl*II, *Xho*I) and (*BamH*I, *Sal*I) were used for digestion of *T-STAR* and *Sam68* (wild type and 439R) PCR products respectively. BSA (Bovine Serum

Albumin) was used for enzymes to work efficiently. The digestion mixes used for PCR products and plasmids are shown in table 6:

PCR product			Plasmid (pcDNA5 vector)		
	First digestion	Second digestion		First digestion	Second digestion
dH$_2$O	9.5 µl	3µl	dH$_2$O	32 µl	8 µl
Purified DNA	20 µl	25µl	Purified DNA	5 µl	20 µl
BSA (10X)	4 µl	4µl	BSA (10X)	5 µl	4µl
Buffer	4 µl	4µl	Buffer	5 µl	4µl
Enzyme	2.5 µl	4µl	Enzyme	3 µl	4µl
Total	40 µl	40 µl	Total	50µl	40 µl

Table 6: Recipes of restriction digest mixes used for PCR products (left) and plasmid digestions (right). PCR products and pcDNA5 vector were incubated with appropriate restriction enzyme separately for between 2-3 hours.

2.2.1(c) Ligations and Re-cleavage

After digestion, the purified digested vector and the DNA insert were ligated to each other in a 10-20 µl reaction volume containing DNA ligase (table 7). T4 DNA ligase and its buffer were purchased from New England biolabs. The ligation mixtures were incubated overnight at 16°C.

Purified digested vector	1-15µl
Purified digested insert	1-15µl
T4 DNA ligase buffer	2 µl
T4 DNA ligase enzyme	2 µl
Total	20 µl

Table 7: Recipe of ligation mixtures.

Heat inactivation of incubation ligation mixtures was carried out at 65°C for 10 minutes. To confirm the ligation mixtures contained the desired clones, the PCRs with reverse primer of the vector and forward primer of the DNA insert were carried out. To enrich for the desired clones, re-cleaving of ligated vectors using an appropriate enzyme in a 20 µl reaction volume was carried out. The re-cleavage was selected to cut the ligated vectors which do not contain the insert in their multi cloning sites. Figure 13 shows a schematic design of cloning of *T-STAR*, wild type and P439R *Sam68* into pcDNA5.
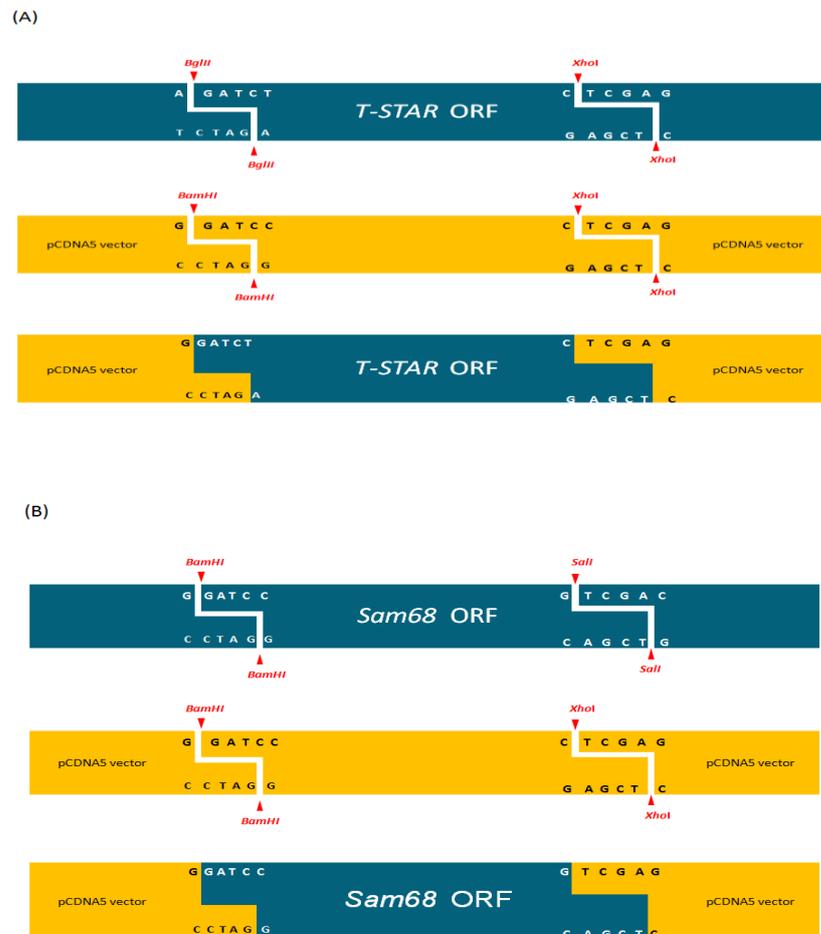
Figure 13 : Schematic picture of digestion and ligation strategy to clone inserts into the pcDNA5 vector.

(A) Schematic picture of the digestion and ligation strategy to clone the *T-STAR* ORF into the pcDNA5 vector. *T-STAR* PCR product (insert) was digested by the restriction enzymes *Bgl*II and *Xho*I. The pcDNA5 vector was digested by the restriction enzymes *BamH*I and *Xho*I. After the ligation procedure, re-cleavage with *BamH*I enzyme was carried out to remove the ligated vectors which did not contain an insert. Ligation to *Bgl*II destroys the *BamH*I site in the pcDNA5 vector. B) Schematic picture of digestion and ligation strategy to clone wild type and P439R *Sam68* into the pcDNA5 vector. The *Sam68* PCR product (insert) was digested by the restriction enzymes *BamH*I and *Sal*I (B) The pcDNA5 vector was digested by the restriction enzymes *BamH*I and *Xho*I (C) After the ligation procedure, re-cleavage with *Not*I enzyme was carried out to remove re-ligated pcDNA5 vectors which did not contain an insert.

2.2.1(d) Purification of DNA after PCR or restriction digests

The QIAquick PCR purification kit (Qiagen) was used to purify PCR products or restriction digested DNA from enzymes, buffers and dNTPs. Purification was done based on the manufacturer's instructions.

2.2.1(e) Dialysis

Residual salt from ligation/re-cleavage mixtures were removed using dialysis discs (Millipore). Samples were placed on filters. These filters were floated in plates which contain dH$_2$O. The dialysis was carried out at room temperature for 1 hour.

2.2.1(f) Transformation into E. coli

Transformation of the plasmids was carried out by electroporation into an *E.coli* strain called DH5α. 2µl of dialysed, ligation mixtures were mixed with 20 µl of the electro-competent cells and transferred into an electroporation cuvette. The competent cells, dialysed re-cleaved ligation mixtures and electroporation cuvette were kept completely cooled. The electroporation voltage was at 1.5kV. The electroporated cells were incubated in 1 ml LB broth for 1 hour at 37°C enable recovery. The LB agar plates which contain 50µg/ml ampicillin were used for plating out competent cells. Plates were incubated at 37°C overnight. (The recipes for LB broth and LB Agar mixtures are shown in table 8).

LB broth mixture(500mL)		LB agar mixture	
dH$_2$O	300ml	dH$_2$O	300ml
Tryptone	5gr	Tryptone	5gr
Yeast Extract	2.5gr	Yeast Extract	2.5gr
Sodium Chloride	5gr	Sodium Chloride	5gr
dH$_2$O up to	500ml	dH$_2$O up to	500ml
		Agar	7.5gr

Table 8: LB broth (left) and LB agar (right) recipe.

2.2.1(g) Purification of plasmid DNA

Single colonies were selected from the plates and inoculated into tubes which contained 5 ml LB broth plus 50µg/ml ampicillin. The tubes were incubated in a shaker for an overnight step at 37°C. The next step was centrifugation of cultures at 8000rpm for 3 minutes. A QIAprep Miniprep Kit (Qiagen) was used to extract plasmid DNAs from pellets. The purification of plasmids was then carried out according to the manufacturer's instructions.

2.2.1(h) Agarose gel electrophoresis

Preparation of 1% agarose gels was carried out by melting 0.80g agar in 80ml TAE buffer (1X Tris-acetate-EDTA) and 8µl of 1000X ethidium bromide. DNA samples were mixed with the appropriate amount of 10X loading dye, and electrophoresed at 100

volts in 1X TAE buffer (50X TAE buffer: 242g Tris base, 57.1ml acetic acid, 100ml of 500mM EDTA pH 8) for 1 hour. Syngene (Gene Genius Bioimaging System) was used to visualize DNA with UV light. 1Kb plus DNA ladder (Invitrogen) was used to estimate sizes.

2.2.1(i) Sequencing

Sequencing was carried out by the Geneservice Company using plasmids concentrations at 100ng/µl. The primers pCDNArev (ACAATGCGATGCAATTTCCTC), BGH (TAGAAGGCACAGTCGAGG) and CMV (CGCAAATGGGCGGTAGGCGTG3) were used for sequencing.

2.2.2 Cell culture

2.2.2(a) Establishment of stable cell lines

Flp-In HEK293 cells were given to our group by Dr Nicholas Watkins and Dr Andrew Knox (Institute for Cell and Molecular Biosciences, Newcastle University). In the genome of these cells there is an FRT site and a blasticidin resistance gene which has been stably integrated. In addition, the Tet repressor is expressed in these cells. Expression of the genes of interest cloned into the FRT site is controlled by the tetracycline promoter. Therefore, a high level of gene expression will be achieved by adding tetracycline to the cell culture medium. To ensure the presence of the FRT site in the Flp-In HEK293 cells, blasticidin was added every third feed. After 24 hours of seeding cells in 10% FBS DMEM (Dulbecco's MEM) with glutamax-1 medium (high glucose (4.5g/l) with L-Glutamine), when they were 60% confluent, the constructs containing the open reading frames of the T-STAR, wild type and P439R Sam68 proteins cloned in the pcDNA5 vector were co-transfected with a plasmid pOG44 (encoding Flp recombinase) at a ratio of 9:1. The empty pcDNA5 vector was co-

transfected with the plasmid pOG44 as the negative control. Selection of transfected cells was carried out 48 hours later using 1µg/ml hygromycin B (Sigma-Aldrich). The stable transfected cells had a hygromycin resistance gene, therefore, they survived after adding the hygromycin to the media. The dead cells were removed from the plate. The live cells were then washed with 1ml of 1xPBS, followed by adding 2 ml of fresh medium and 2 µl of hygromycin B. After a few weeks of selection by hygromycin B, colonies were re-plated. When the cells were confluent enough, expression of the genes of interest were induced by adding 1µg/ml tetracycline (Sigma-Aldrich) to the cell culture medium.

2.2.2(b) HEK293 and HeLa cells

HEK293 and HeLa cells (provided in Professor Elliott's lab, IGM, Newcastle University) were seeded (at a density of 1×10^5 /ml) in 10% FBS DMEM (Dulbecco's MEM) with glutamax-1 medium and were incubated at 37°C in 5% CO_2.

2.2.2(c) Harvesting cells

After removing growth medium (Invitrogen) cells were washed with 1X PBS (Invitrogen). Trypsin-EDTA (Invitrogen) was added to detach the cells from the plate. Next, the growth medium was added, followed by mixing and centrifugation of the cells at 8000rpm for 2 minutes. Cell pellets were washed with 1X PBS, centrifuged at 8000rpm for 2 minutes followed by removal of the supernatant in order to make the cell pellets ready for Western blotting or co-immunoprecipitation. For SDS-PAGE, after harvesting the cells, 2X SDS sample loading dye (4% SDS, 100mM Tris-HCl pH6.8, 20% glycerol, 200mM DTT, Bromophenol Blue) was added to the cell pellets before sonication.

2.2.3 Protein based methods

2.2.3(a) SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Before loading, the sonicated samples were incubated at 96°C for 5 minutes and centrifuged at 8000rpm for 2 minutes. For Western blotting, 4% stacking gels and 10% running gels were made for loading the protein samples. Electrophoresis was carried out at 150V for approximately 50 minutes. For Simply Blue staining (Invitrogen), 4% stacking and 12% resolving gels (Invitrogen) were used and the protein samples were electrophoresed at 15V for more than 2 hours. De-staining of Simply Blue stained gels was carried out by using dH$_2$O. A Page Ruler prestained protein ladder (10kDa-170kDa), Page Ruler plus prestained protein ladder (10kDa-250kDa) (Fermentas) and prestained protein marker broad range (7kDa-175kDa) (New England Biolabs) were used to estimate the protein sizes.

2.2.3(b) Western blot (Protein immunoblot analysis)

Transferring of gels to the nitrocellulose (NC) membranes (Amersham Biosciences) was carried out for 30 minutes at 15V using transfer buffer (0.04% SDS, 58mM Tris, 39mM Glycine, 20% MeOH). Nitrocellulose membranes were then washed with 1xTBS-T buffer (6.5g/l Tris-HCl, 8.75g/l NaCl, 2ml/l Tween) and incubated for 1 hour in the blocking solution (milk powder in 1xTBS-T). Incubations with the primary antibody (table 9) were carried out for 1 hour. Before adding the secondary antibody, the nitrocellulose membrane was washed three times with 1xTBS-T. After 1 hour incubation with the secondary antibody (table 10), the nitrocellulose membrane was washed three times with 1xTBS-T (each time for 5 minutes). All steps were carried out at room temperature. Lastly, the nitrocellulose membranes were incubated with 1ml of both ECLI (0.44% Coumaric acid, 1% Luminol, and 100mM Tris pH 8.5) and ECLII

(0.02% hydrogen peroxide in100mM Tris pH 8.5) solutions for 1 minute and then exposed to film (Kodak).

Primary Antibody	Source	Monoclonal /Polyclonal	Species	Dilution for Western Blot	Dilution for immunofluorescence
Anti β Actin	Sigma	Polyclonal	Rabbit	1:1000	-
Monoclonal anti Flag	Sigma	Monoclonal	Mouse	1:1000	1:200
Rabbit Anti DHX9 (RHA)	ProteinTech Group	Polyclonal	Rabbit	1:2000	1:200
Rabbit anti CCAR1	Sigma	Polyclonal	Rabbit	1:500	1:200
Anti hnRNP L(4D11)	Sigma	Monoclonal	Mouse	-	1:200

Table 9: Primary antibodies used for Western blot and immunofluorescence.

Secondary Antibody	Source	Monoclonal /Polyclonal	Species	Dilution for Western Blot	Dilution for immunofluorescence
Anti-mouse-HRP	Amersham	Polyclonal	Sheep	1:1000	-
Anti-rabbit-HRP	Jackson lab	Polyclonal	Goat	1:1000	-
Anti-sheep-HRP	DAKO A/S	Polyclonal	Rabbit	1:1000	-
Anti-rabbit-488(Green)	Molecular Probes	Polyclonal	Donkey	-	1:400
Anti-mouse-594(Red)	Molecular Probes	Polyclonal	Donkey	-	1:400
Anti-mouse-HRP	Amersham	Polyclonal	Sheep	1:1000	-

Table 10: Secondary antibodies used for Western blot and immunofluorescence.

2.2.3(c) Immunoprecipitation using α-FLAG affinity resin

The transfected Flp-In HEK293 cells were harvested 72 hours after induction with tetracycline (as explained above). 1.5 ml of IP +Benzonase (100mM KCl, 50mM Tris pH 7.4, 5mM NaF, 5mM MgCl2, 0.1% NP-40, 2μl Benzonase , 1X Roche Complete Protease Inhibitor Cocktail) was added to the cell pellets followed by a 20 minute incubation on

ice. The FLAG resin (Sigma) was washed 3 times with TBS pH 7.4 and the samples were centrifuged for 10 minutes at 4°C. 30 µl of the supernatant was kept as Input and the rest of the supernatant was incubated with the FLAG resin for an overnight step at 4°C. The samples were centrifuged at 1000rpm for 30 seconds and 30 µl of the supernatant was kept and labelled SupIP. Next, the FLAG resin was washed six to seven times with IP wash buffer (100mM KCL, 50mM Tris pH 7.4, 0.1% NP-40). To check the efficiency of co-immunoprecipitation, 2X SDS sample loading dye was added to samples for loading on SDS-PAGE and checked by Western blotting.

2.2.3(d) Mass spectrometry

Electrophoresis of protein-FLAG immunoprecipitation (IP) samples on SDS-PAGE gels (NuPAGE 4-12% Bis-Tris Gel, Invitrogen) was followed by staining gels with Simply Blue for 1 hour and de-staining by using dH$_2$O. The protein bands were cut and sent for analysis by Dr Gary Black and Dr Andrew Porter (Department of Biomedical Sciences, School of Life Sciences at Northumbria University).

2.2.3(e) Immunofluorescence

When HeLa cells were 50-70% confluent, coverslips were washed with 1XPBS, followed by fixation of the cells with 4% (w/v) PFA (Paraformaldehyde) for 20 minutes at room temperature. The cells were permeabilised by incubation in 1% Triton X-100 for 15 minutes, and blocked in 10% horse serum/0.1% Triton X-100 for 1 hour. Incubation of the cells in primary and fluorescent conjugated secondary antibodies (tables 9 and 10) diluted in blocking solution was carried out for 1 hour for each of the antibodies. After incubation with the antibodies, cells were washed with 1XPBS each time for 5 minutes. The coverslips were mounted using VectaShield Mounting Medium with 4', 6-

diamidino-2-phenylindole (DAPI) (Vector Laboratories). The fluorescent images were obtained on an Axiovision microscope.

2.3   Results

2.3.1   Cloning of T-STAR into pcDNA5 (generation of T-STAR construct)

Primers T-STARBGLIIF and T-STARXHOIB were used to PCR amplify the open reading frame of T-STAR, using *T-STAR* cloned into the GFP3 as a template. The PCR amplification resulted in a ~ 1Kb product (figure 14). The pcDNA5 vector and PCR product (insert) were digested by the restriction enzymes (*BamH*I, *Xho*I) and (*Bgl*II, *Xho*I) respectively and then ligated together. After the ligation procedure, a re-cleavage with *BamH*I enzyme was carried out to remove the ligated vectors which did not contain an insert. The ligation mixes were transformed into *E.coli DH5α*, which were plated out on media suplimented with ampicillin. 20 colonies were screened by PCR in which insert forward primer (T-STARBGLIIF) and vector reverse primer (BGH reverse primer) were used. Colonies were selected on the LB medium plus 50μg/ml ampicillin. After overnight incubation at 37°C on a shaker, plasmid purification was carried out. To confirm the correct presence of inserts in the pcDNA5 vector, 100ng/μl of constructs were sequenced. The results confirmed the presence of the insert containing the open reading frame of T-STAR within the constructs. The sequencing results are shown in the Appendix A.
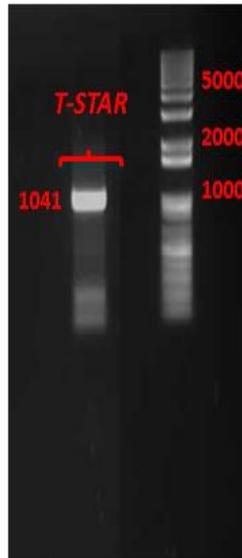
Figure 14: Amplification of the open reading frame (ORF) of _T-STAR_ for cloning. The PCR amplification resulted in a 1041bp product being identified on an ethidium bromide stained agarose gel. A 1kb plus DNA ladder was run in parallel on the gel.

2.3.2 Cloning of wild type and P439R mutant Sam68 into pcDNA5

The only difference between wild type and P439R mutant Sam68 was a point mutation in the codon encoding P439 (changing proline to arginine). To clone the ORFs for these proteins into the pcDNA5, MSam68 F-BamHI was used as the forward primer for both wild type and P439R mutant _Sam68_. As the reverse primers, Wt MSam68 R-SalI and Mutant MSam68R-SalI were used for wild type and P439R _Sam68_ respectively. These primers were used to PCR amplify the open reading frames of wild type and P439R mutant Sam68 using Sam68 cloned into a pACT2 AD construct as a template. The PCR amplification resulted in 1.3Kb products (figure 15).

(A) (B)
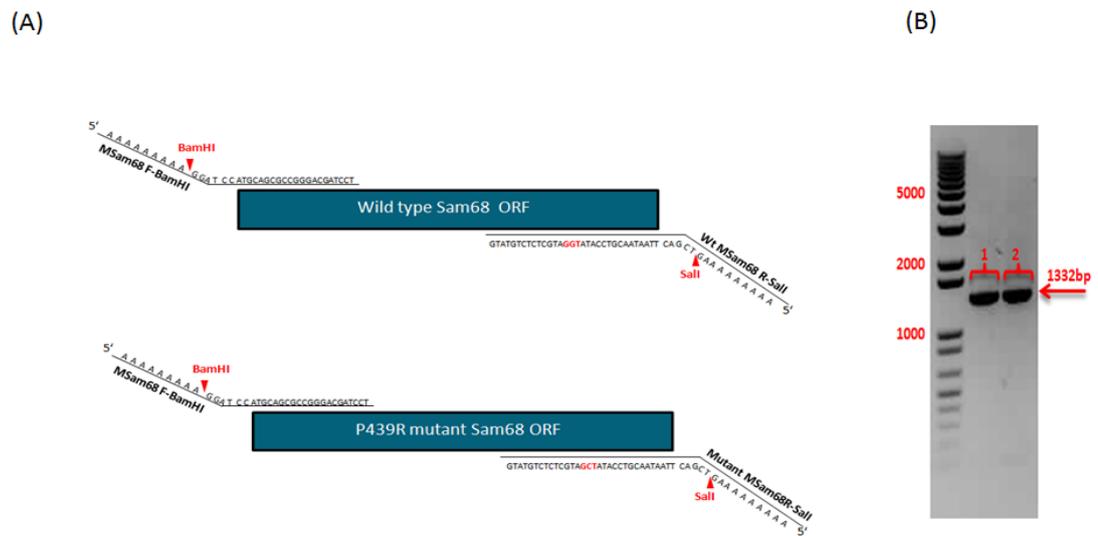
Figure 15: PCR amplification of the wild type and P439R mutant *Sam68* for cloning into pcDNA5. (A) Schematic picture of the PCR amplification of the open reading frame of wild type and mutant Sam68. MSam68 F-BamHI was used as the forward primer for both wild type and P439R mutant *Sam68*. As the reverse primers, Wt MSam68 R-SalI and Mutant MSam68R-SalI were used for wild type and P439R mutant *Sam68* respectively. (B) PCR amplifications resulted in 1332bp products in ethidium bromide stained agaros gel. 1) Wild type *Sam68* and 2) P439R mutant *Sam68*. As it has been shown the size of *Sam68* PCR product is 1332bp. 1kb plus DNA ladder also has been shown in this picture.

The pcDNA5 vector and *Sam68* PCR product (insert) were digested by the restriction enzymes (*BamH*I, *Xho*I) and (*BamH*I, *Sal*I) respectively. *Not*I was used for the re-cleavage process to remove the ligated vectors which did not contain an insert. Sequencing results confirmed the presence of inserts containing the open reading frames of wild type and P439R *Sam68* within the constructs (see Appendix A).

2.3.3 Establishment of stable cell lines expressing T-STAR, wild type and P439R mutant Sam68 FLAG-tagged proteins

Stable cell lines expressing epitope tagged proteins were made by transfecting the T-STAR, wild type and P439R mutant Sam68 open reading frames cloned in the pcDNA5 into the Flp-In HEK293 knock in cell line. In the genome of these knock in cells there is an FRT site and a blasticidin resistance gene which has been stably integrated. In addition, the tet repressor is expressed in these cells. Expression of the gene of interest cloned into the FRT site was controlled by the tetracycline promoter. Therefore, a high level of gene expression would be achieved by adding tetracycline to the cell culture medium. To ensure the presence of the FRT site in the Flp-In HEK293 cells, blasticidin was added every third feed.

After 24 hours of seeding cells in 10% FBS DMEM with glutamax-1 medium , when they were 60% confluent, the open reading frame of T-STAR, wild type and P439R mutant Sam68 proteins cloned in the pcDNA5 vector were co-transfected with a plasmid pOG44 (encoding Flp recombinase). Also, empty pcDNA5 vector was co-transfected with a plasmid pOG44 as the negative control. Selection of transfected cells was carried out 48 hours later using hygromycin B. The stable transfected cells had a hygromycin resistance gene, therefore, they would survive after adding hygromycin to the media. In the case of cell death, untransfected cells were removed from the plate. Live cells then were washed with PBS and this step was followed by adding fresh medium and hygromycin B. After few weeks of selection by hygromycin B, colonies were re-plated.

When the cells were confluent enough, the genes of interest were induced by adding tetracycline to the cell culture medium. Using Western blotting, the presence of FLAG tagged proteins (T-STAR, wild type and P439R mutant Sam68) expression was tested after induction in comparison with negative control (FLAG tagged pcDNA5) (figure 16).

Any FLAG tagged protein at 0, but substantial induction at 24, 48 and 72 hours was detected.

The molecular weight of the endogenous T-STAR and Sam68 proteins are 55kDa and 68 kDa respectively. Because there are 16 and six amino acids in 2x FLAG (D Y K D D D D K D Y K D D D D K) and His tags (H H H H H H) respectively, the weight of recombinant proteins were higher than the weight of the endogenous proteins.

Blots were probed with monoclonal anti-FLAG and anti-mouse conjugated to HRP as the primary and secondary antibodies respectively. Then, blots were re-probed with α-β actin and α-rabbit IgG conjugated to HRP as the primary and secondary antibodies respectively to test that the sample had been loaded equally. As it has been shown in figure 16 equal actin loading was present in each lane.
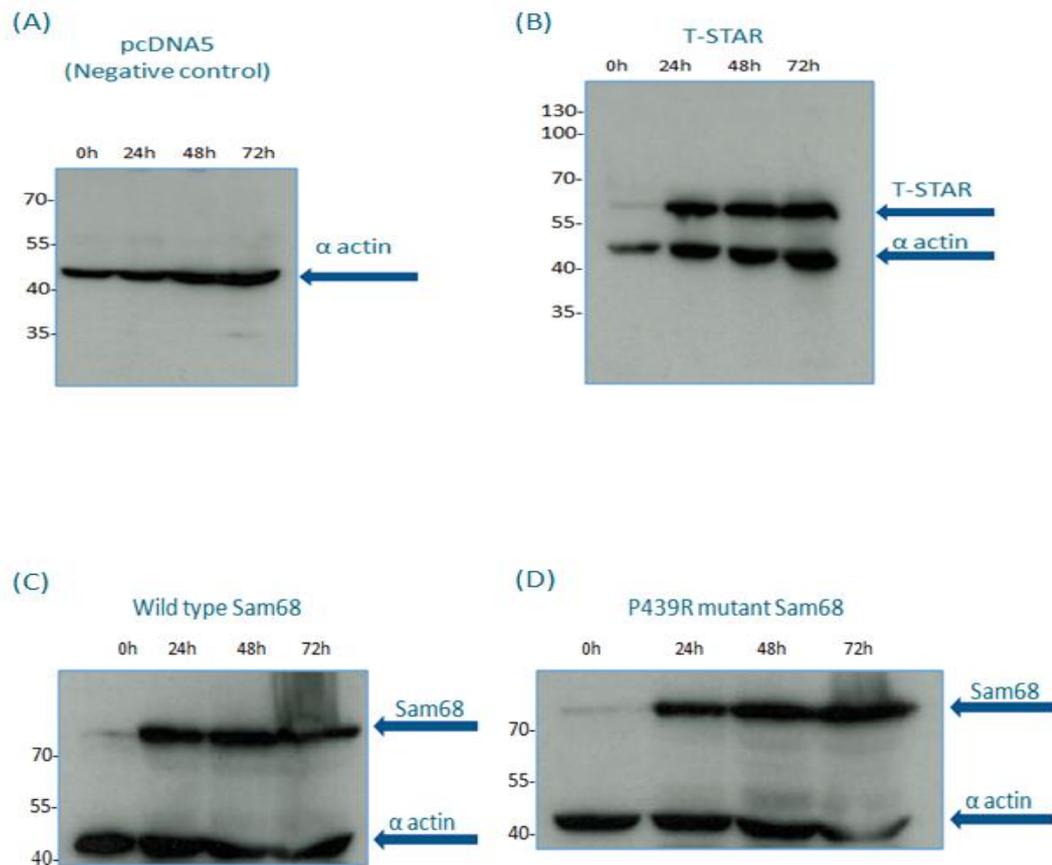
Figure 16: Induced expression of FLAG tagged proteins (T-STAR, wild type and P439R mutant Sam68) in the stable cell lines as monitored by Western blotting. The presence of FLAG tagged proteins (T-STAR, wild type and mutant Sam68) was confirmed after 24, 48 and 72 hours induction in comparison with the negative control (pcDNA5). The weight of endogenous T-STAR and Sam68 proteins are 55kDa and 68 kDa respectively. Because there are FLAG tags and His tags in these recombinant proteins, their weights were higher than the endogenous proteins. T-STAR, wild type and P439R Sam68 were around 60kDa and 80kDa respectively. A) Negative control (pcDNA5), B) T-STAR, C) wild type Sam68, D) P439R Sam68. Monoclonal anti-FLAG and anti-mouse conjugated to HRP were used as the primary and secondary antibodies respectively to detect the presence of FLAG tagged proteins. Blots were re-probed with α-β actin and α-rabbit IgG conjugated to HRP as the primary and secondary antibodies respectively to confirm that the samples had been loaded equally. Equal levels of actin loading control were present in each lane.

2.3.4 Localisation of FLAG tagged Sam68 proteins in the stable cell lines

After detecting the expression and correct size of FLAG-tagged proteins using Western blot, immunofluorescence experiments were carried out to see if these proteins correctly localise in the cell. Sam68 protein has a mostly nuclear localization, even though initially it was identified in the cytoplasm during mitosis (Rajan et al., 2008a). It has been shown that by mutating NLS in Sam68 (P439R, changing proline to arginine), the resulting Sam68 protein is localized in the cytoplasm of HeLa cells (Reddy, 2000). The plan was to test if the P439R mutant Sam68 protein would have a similar cytoplasmic localisation in the Flp-In HEK293 cells after induction in a stable cell line expressing FLAG tagged P439R mutant Sam68. Consistent with this, FLAG-tagged P439R mutant Sam68 was only localized in the cytoplasm and not present within the nucleus (figure 17).
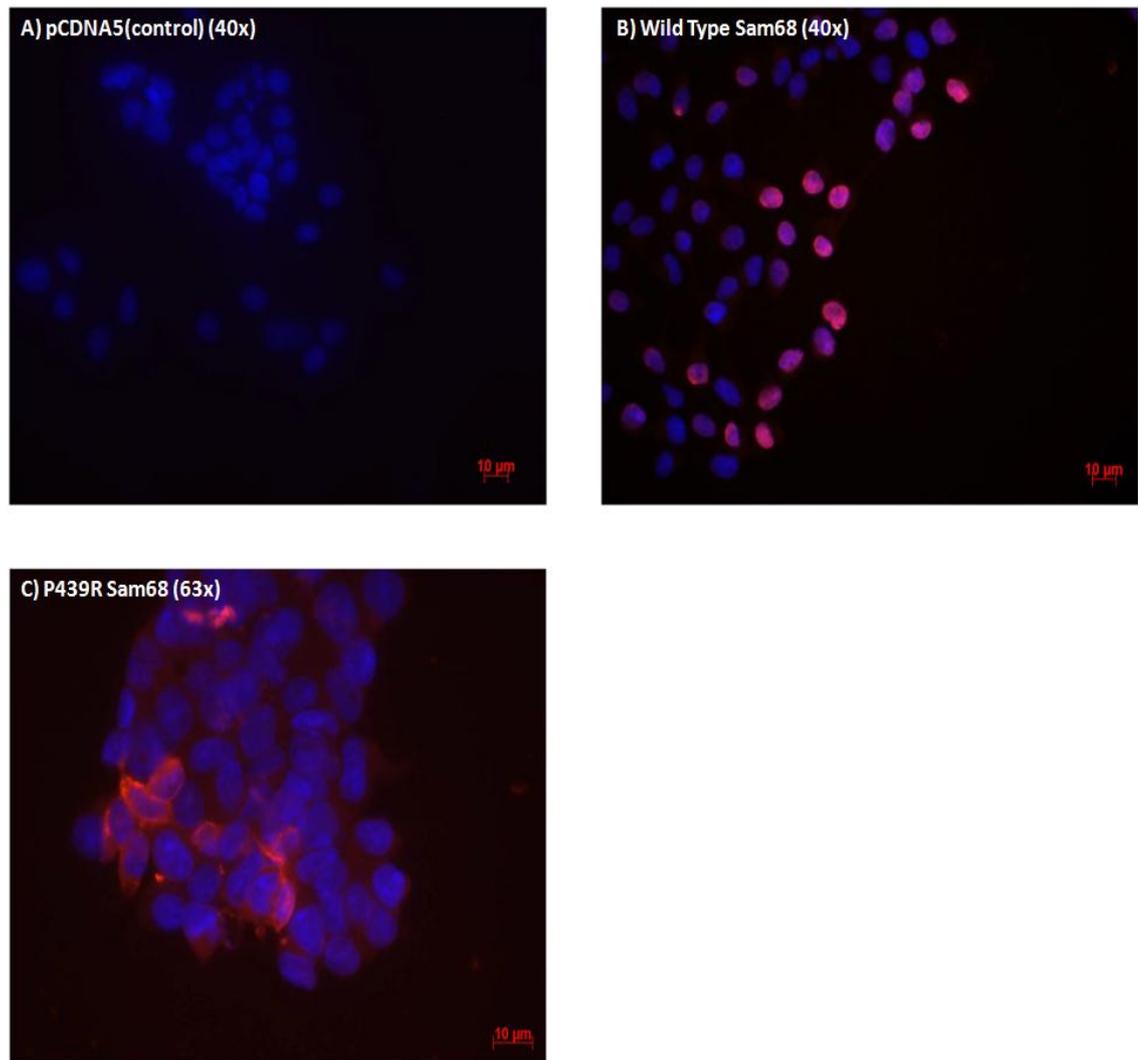
Figure 17: Localization of FLAG tagged wild type and P439R mutant Sam68 proteins after induction in stable Flp-In HEK293 cell lines. A) Negative control (pcDNA5), B) Localization of Wild type Sam68 after induction by tetracycline and C) Localization of P439R mutant Sam68 after induction by tetracycline. Wild type Sam68 is completely nuclear while P439R mutant Sam68 is only localized within the cytoplasm. Monoclonal anti FLAG and α-mouse IgG (594) were used as the primary and secondary antibodies respectively. The pcDNA5 is stably transfected with a control empty vector. These cells did not contain any FLAG epitope containing induced gene and were used as the negative control samples. DAPI was used to stain the nucleus and the red colour is showing the localization of induced proteins. The scale bars are equal to 10μm. Monoclonal anti FLAG and α-mouse IgG (594) were used as the primary and secondary antibodies respectively.

2.3.5 Purification of FLAG tagged proteins using immunoprecipitation

After I confirmed the localization of wild type and P439R Sam68 proteins within the cell, the next step in the project was to investigate more about protein partners of T-STAR, wild type and mutant Sam68. Therefore, 72 hours after tetracycline induction of the stable Flp-In HEK293 cells, the cells were harvested, followed by the immunoprecipitation of the FLAG-tagged proteins (T-STAR, wild and P439R mutant Sam68) using α-FLAG affinity resin. The immunoprecipitated (IP) samples were first analysed by Western blotting to confirm the presence of FLAG-tagged proteins within the input and IP samples. Monoclonal anti FLAG antibody was used as the primary antibody to probe the Western blots and analyse whether the immunoprecipitation was successful. The IP lanes of FLAG tagged proteins in comparison with the control sample (pcDNA5) showed that the immunoprecipitations were successful in retrieving the tagged proteins (figure 18).
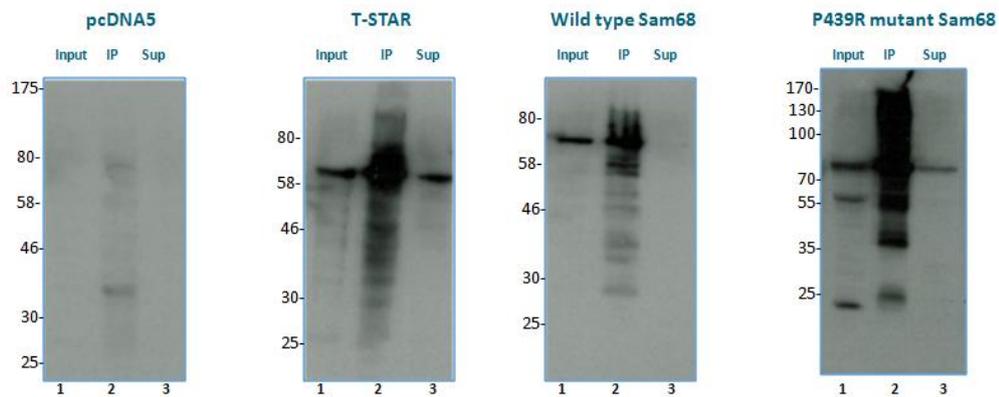
Figure 18: Successful Immunoprecipitation (IP) of the FLAG tagged proteins from stable cell lines. Western blotting carried out to determine IP efficiency. Monoclonal anti FLAG antibody and HRP-conjugated anti-mouse were used as the primary and secondary antibodies. Lanes 1, 2 and 3 show the Input, IP and Sup samples respectively. The IP lanes of FLAG tagged proteins in comparison with the control sample (pcDNA5) showed that the immunoprecipitations were successful. The Size markers are in kDa. Input: soluble supernatant before immunoprecipitation, IP: immunoprecipitated samples Sup: soluble supernatant after immunoprecipitation. All samples were performed in triplicate.

2.3.6 Identification of candidate interacting protein partners of T-STAR using mass spectrometry

After the process of Western blotting showed the efficient recovery of the FLAG-tagged proteins, the next stage was to try and identify any co-immunoprecipitated proteins. In order to identify the candidate interacting protein partners for T-STAR, protein-FLAG immunoprecipitation (IP) samples were loaded on SDS-PAGE gels and electrophoresed, followed by staining of gels with Simply Blue stain for 1 hour. Gels were de-stained by using dH$_2$O overnight. The protein bands were theb cut and sent for mass spectrometry. The gel in figure 19 shows more proteins in the T-STAR lane compared with the control sample lane (empty pcDNA5). T-STAR specific protein bands were cut out of the gel and sent for identification by mass spectrometry. Based on the

sequencing results (table 11), identities of three candidate T-STAR interacting proteins were predicted.

In addition to proving the presence of T-STAR, two other proteins, hnRNP G and RNA helicase A (RHA) were detected as the potential T-STAR partners (figure 19). The expected presence of T-STAR and hnRNP G were consistent with the immunoprecipitation having worked, since they were known T-STAR interacting protein partners (Ehrmann and Elliott, 2010). In addition, RHA was detected as a novel candidate interacting partner for T-STAR. Figure 20 shows the sequences of these proteins. Matched peptides identified by mass spectrometry are shown in red. Extra information including Swiss-Prot entry name, Accession number, Gene name, Recommended Name and Nominal mass (Mr) of these proteins are shown in table 11.
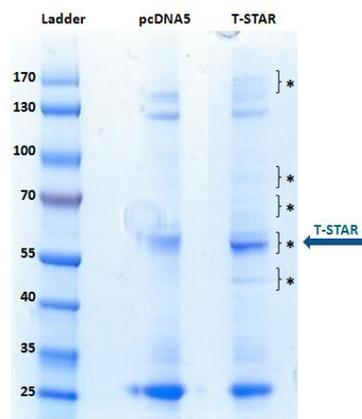
Figure 19: Simply Blue staining for detection of candidate interacting protein partners of T-STAR. After loading IP samples on 4-12% gradient gels and gel elctrophoresis, Simply Blue safe stain was used for staining the gels. Stained gels were washed with dH$_2$O. The protein bands which are highlighted by asterisks were cut and sent for sequencing by mass spectrometry. There are more protein bands in the T-STAR lane in comparison with the control sample (pcDNA5) suggesting the existence of T-STAR interacting proteins. Based on mass spectrometry, T-STAR, hnRNP G and RHA were detected. The presence of T-STAR and hnRNP G showed that the immunoprecipitation had worked. In addition, RHA was detected as a novel candidate interacting partner for T-STAR.

T-STAR

Swiss-Prot entry name: KHDR3_HUMAN   Accession number: O75525

MEEKYLPELMAEKDSLDPSFTHALRLVNREIEKFQKGEGKEEEKYIDVVINKNMKLGQKVLIPVKQ
FPKFNFVGKLLGPRGNSLKRLQEETLTKMSILGKGSMRDKAKEEELRKSGEAKYFHLNDDLHVLIEV
FAPPAEAYARMGHALEEIKKFLIPDYYDEIRQAQLQELTYLNGGSENADVPVVRGKSTLRTRGVTT
PAITRGRGGVTARPVAVGVPRGTPTPRGVLSTRGPVSRGRGLLTPRARGVPPTGYRPPPPPPTQET
YGEYDYDDGYGTAYDEQSYDSYDNSYSTPAQSAADYYDYGHGLSEDAYDSYGQEEWTNSRHKAP
SARTAKGVYRDQPYGRY

HnRNPG (RBMX):

Swiss-Prot entry name: HNRPG_HUMAN   Accession number: P38159

MVEADRPGKLFIGGLNTETNEKALEAVFGKYGRIVEVLLMKDRETNKSRGFAFVTFESPADAKDAR
DMNGKSLDGKAIKVEQATKPSFESGRRGPPPPPRSRGPPRGLRGGRGGSGGTRGPPSRGGHMD
DGGYSMNFNMSSSRGPLPVKRGPPPRSGGPPPKRSAPSGPVRSSSGMGGRAPVSRGRDSYGGP
PRREPLPSRRDVYLSPRDDGYSTKDSYSSRDYPSSRDTRDYAPPPRDYTYRDYGHSSSRDDYPSRGY
SDRDGYGRDRDYSDHPSGGSYRDSYESYGNSRSAPPTRGPPPSYGGSSRYDDYSSSRDGYGGSRD
SYSSSRSDLYSSGRDRVGRQERGLPPSMERGYLLHVIPTAVQAADSQEVVAVEEADLIEGEAEADT
RNKQNFGPKSQFKETKSGNYSIITTQGLLKGKIVLLFLNSLLSSPP

RNA helicase A:

Swiss-Prot entry name: DHX9_HUMAN   Accession number: Q08211

MGDVKNFLYAWCGKRKMTPTYEIRAVGNKNRQKFMCEVQVEGYNYTGMGNSTNKKDAQSNA
ARDFVNYLVRINEIKSEEVPAFGVASPPPLTDTPDTTANAEGDLPTHHGGPLLHILALKAENNSEVG
ASGYGVPGPTWDRGANLKDYYSRKEEQEVQATLESEEVDLNAGLHGNWTLENAKARLIQYFQKE
KIQGEYKYTQVGPDHNRSFIAEMTIYIKQLGRRIFAREHGSNKKLAAQSCALSLVRQLYHLGVVEAY
TGLTKKKEGETVEPYKVNLSQDLEHQLQNIIQELNLEILPPPEDPSVPVALNIGKLAQFEPSQRQNQ
VGVVPWSPPQSNWNPWTSSNIDEGPLAFATPEQISMDLKNELMYQLEQDHDLQAILQERELLPV
KKFESEILEAISQNSVVIIRGATGCGKTTQVPQFILDDFIQNDRAAECNIVVTQPRRISAVSVAERVAF
ERGEEPGKSCGYSVRFESVLPRPHASIMFCTVGVLLRKLEAGIRGISHVIVDEIHERDINTSFLLVVLR
DVVQAYPEVRIVFMSATIDTSMFCEYFFNCPSLKLWRTYPVQEYFLEDCIQMTHFVPPPKDKKKKD
KDDDGGEDDDANCNLICGDEYGPETRLSMSQLNEKETPFELIEALLKYIETLNVPGAVLVFLPGWN
LIYTMQKHLEMNPHFGSHRYQILPLHSQIPREEQRKVFDPVPVGVTKVILSTNIAETSITINDVVYVD
SCKQKVKLFTAHNNMTNYSTVWASKTNLEQRKGRAGRSTAGFCFHLCSRARFERLETHMTPEMF
RTPLHEIALSIKLLRLGGIGQFLAKAIEPPPLDAVIEAEHTLRELDALDANDELTPLGRILAKLPIEPRFG
KMMIMGCIFYVGDAICTIAAATCFPEPFVNEGKQLGYIHRNFAGNRFSDHVALLSVFQAWDDAM
GGEEAEIRFCEHKRLNMATLRMTWEAKVQLKEILINSGFPEDCLLTQVFTNTGPDNNLDVVISLLA
FGVYPNVCYHKEKRKILTTEGRNALIHKSSVNCPFSSQDMNYPSPFFVFGEKIRTRAISAKGMTLVP
PLQLLLFASKKVQSDGQIVLVDDWIKLQISHEAAACITGLRAAMEALVVEVTKQPAIISQLDPVNER
MLNMIRQISEPSAAGINLMIGSTRYGDGPRPPKMARYDNGSGYRRGGSSYSGGGYGGGYSSGGY
GSGGYGGSATPSGRICAGVGGGYRGVSRGGFRGNSGGDYRGPSGGYRGSGGFQRGGGRGAYG
TGYLDIEEEVAAIKLGYVSSVCRQ

Figure 20: Mass spectrometry results for T-STAR interacting proteins. Showing the sequences of identified proteins. Matched peptide sequences are shown in red. The percentage of the

protein that was covered by the peptide sequences detected by Mass Spec. were 13%, 5% and 2% for T-STAR, hnRNP G and RNA helicase A (RHA) respectively.

	Swiss-Prot entry name	Accession number	Gene name	Recommended Name	Nominal mass(Mr)
1	KHDR3_HUMAN	O75525	KHDRBS3/ T-STAR	KH domain-containing, RNA-binding, signal transduction-associated protein 3	38833
2	HNRPG_HUMAN	P38159	RBMX	Heterogeneous nuclear ribonucleoprotein G	47419
3	DHX9_HUMAN	Q08211	DHX9	ATP-dependent RNA helicase A	143405

Table 11: Candidate protein partners of T-STAR identified by FLAG immunoprecipitation. Based on mass spectrometry results, the three proteins T-STAR, hnRNP G and RHA were detected by immunoprecipitation and mass spectrometry. Information about these proteins such as Swiss-Prot entry name, Accession number, Gene name, Recommended Name and Nominal mass (Mr) are from two websites: **http://www.genecards.org/** and **http://www.uniprot.org/**.

2.3.7   Identification of the interacting partners of wild type and P439R mutant Sam68 using mass spectrometry

A similar analysis was carried out for Sam68. After loading IP samples on 4-12% gradient gels and gel electrophoresis, Simply Blue safe stain was used for staining the gels. The stained gels were washed with dH$_2$O. As shown in figure 21 there were more protein bands in the wild type and P439R mutant Sam68 lanes in comparison with the control sample (pcDNA5). Candidate Sam68 interacting proteins bands were cut from the gel and sent for identification. Based on mass spectrometry results, in addition to proving the presence of Sam68, a new candidate protein called CCAR1 (cell cycle and

apoptosis regulatory protein 1) was detected for wild type Sam68. On the other hand, for P439R mutant Sam68, in addition to detecting the presence of Sam68, two new candidate interacting protein partners called RHA (RNA helicase A) and CCAR1 were detected. Matched peptides identified by mass spectrometry are shown in red in figures 22 and 23. Extra information including Swiss-Prot entry name, Accession number, Gene name, Recommended name and Nominal mass (Mr) of these proteins are shown in tables 12 and 13.
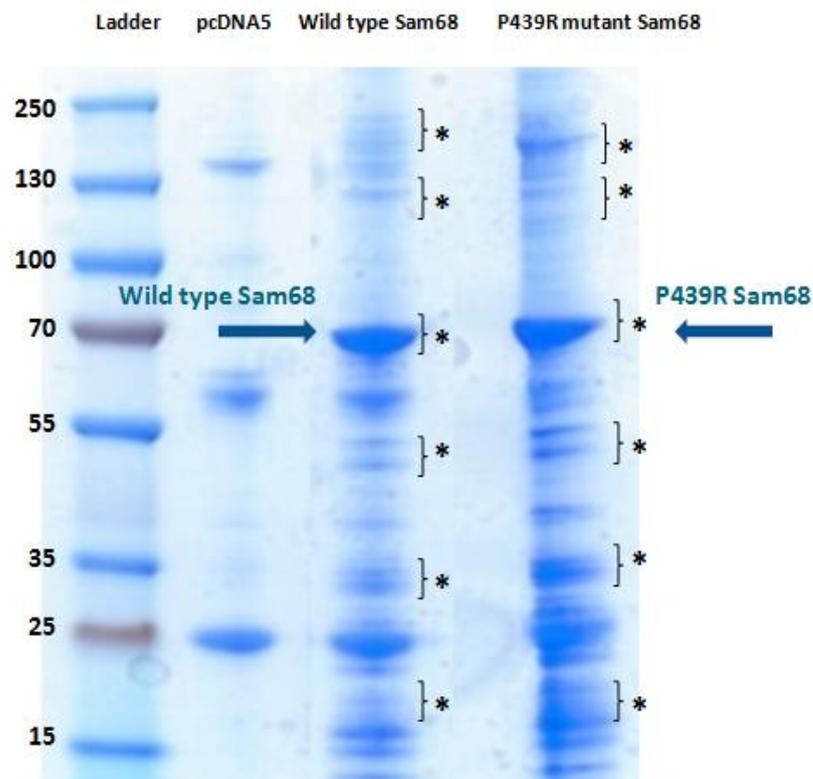
Figure 21: Simply Blue staining for detection of wild type and P439R mutant Sam68 interacting protein partners. After loading the IP samples on 4-12% gradient gels and gel electrophoresis, Simply Blue safe stain was used for staining the gels. Stained gels were washed with dH$_2$O. Protein bands which are highlighted by asterisks were cut out of the gel and sent for sequencing. There were more protein bands in the wild type and P439R mutant Sam68 lanes in comparison with the control sample (pcDNA5).

Wild type Sam68

Swiss-Prot entry name: DHX9_HUMAN   Accession number: Q08211

MQRRDDPASRLTRSSGRSCSKDPSGAHPSVRLTPSRPSPLPHRPRGGGGGPRGGARASPATQPPPLL
PPSTPGPDATVVGSAPTPLLPPSATAAVKMEPENKYLPELMAEK**DSLDPSFTHAMQLLSVEIEK**IQKG
ESK**KDDEENYLDLFSHK**NMKLKERVLIPVKQYPKFNFVRKILGPQGNTIKRLQEETGAKISVLGKGSMR
DKAKEEELRKGGDPKYAHLNMDLHVFIEVFGPPCEAYALMAHAMEEVKKFLVPDMMDDICQEQFLE
LSYLNGVPEPSRGRGVSVRGRGAAPPPPPVPRGRGVGPPRGALVRGTPVRGSITRGSTVTRGVPPPP
TVRGAPTPRARTAGIQRIPLPPTPAPETYEDYGYDDTYAEQSYEGYEGYYSQSQGESEYYDYGHGELQ
DSYEAYGQDDWNGTRPSLKAPPARPVKGAYREHPYGRY

CCAR1 (cell cycle and apoptosis regulatory protein 1)

Swiss-Prot entry name: KHDR1_HUMAN   Accession number: Q07666

MAQFGGQKNPPWATQFTATAVSQPAALGVQQPSLLGASPTIYTQQTALAAAGLTTQTPANYQLTQ
TAALQQQAAAAAAALQQQYSQPQQALYSVQQQLQQPQQTLLTQPAVALPTSLSLSTPQPTAQITVS
YPTPRSSQQQTQPQKQRVFTGVVTKLHDTFGFVDEDVFFQLSAVKGKTPQVGDR**VLVEATYNPNM
PFK**WNAQRI**QTLPNQNQSQTQPLLK**TPPAVLQPIAPQTTFGVQTQPQPQSLLQAQISAASITPLLQT
QPQPLLQQPQQKAGLLQPPVRIVSQPQPARRLDPPSRFSGRNDRGDQVPNRKDDRSRERERERRRS
RERSPQRKRSRERSPRRERERSPRRVRRVVPRYTVQFSKFSLDCPSCDMMELRRRYQNLYIPSDFFDA
QFTWVDAFPLSRPFQLGNYCNFYVMHREVESLEK**NMAILDPPDADHLYSAK**VMLMASPSMEDLYH
KSCALAEDPQELRDGFQHPARLVKFLVGMKGKDEAMAIGGHWSPSLDGPDPEKDPSVLIKTAIRCCK
ALTGIDLSVCTQWYRFAEIRYHRPEETHKGRTVPAHVETVVLFFPDVWHCLPTRSEWETLSRGYKQQL
VEKLQGERKEADGEQDEEEKDDGEAKEISTPTHWSKLDPKTMKVNDLRKELEGRALSSKGLKSQLIAR
LTKQLKVEEQKEEQKELEKSEKEEDEDDDRKSEDDKEEEERKRQEEIERQRRERRYILPDEPAIIVHPNW
AAKSGKFDCSIMSLSVLLDYRLEDNKEHSFEVSLFAELFNEMLQRDFGVRIYK**SLLSLPEKEDK**KEKDKK
SKKDERKDKKEERDDETDEPKPKRRKSGDDKDKKEDRDERKKEDKRKGDSKDDDETEEDNNQDEYD
PMEAEEAEDEEDDRDEEEMTKRDDKRDINRYCKERPSKDKEKEKTQMITINRDLLMAFVYFDQSHCG
YLLEKDLEEILYTLGLHLSRAQVKKLLNKVVLRESCFYRKLTDTSKDEENHEESESLQEDMLGNRLLLPTP
TVKQESKDVEENVGLIVYNGAMVDVGSLLQKLEKSEKVRAEVEQKLQLLEEKTDEDEKTILNLENSNKS
LSGELREVKK**DLSQLQENLK**ISENMSLQFENQMNKTIR**NLSTVMDEIHTVLK**KDNVKNEDKDQKSKE
NGASV

Figure 22: Mass spectrometry results for wild type Sam68 interacting proteins. This picture shows the sequences of identified proteins. Matched peptide sequences identified by mass spectrometry are shown in red. Sam68 and CCAR1 were detected as the interacting protein partners. The peptide coverages recovered by Mass Spec. were 7% for both wild type Sam68 and CCAR1 proteins.

	Swiss-Prot entry name	Accession number	Gene name	Recommended Name	Nominal mass(Mr)
1	KHDR1_HUMAN	Q07666	KHDRBS1/Sam68	KH domain-containing, RNA-binding, signal transduction-associated protein 1	48456
2	CCAR1_HUMAN	Q8IX12	CCAR1	Cell division cycle and apoptosis regulator protein 1	133338

Table 12: Candidate protein partners of wild type Sam68 identified by FLAG immunoprecipitation. Based on mass spectrometry results, the presence of Sam68 in the immunoprecipitate shows the IP has worked. In addition to proving the presence of Sam68, a new candidate interacting protein called CCAR1 (cell cycle and apoptosis regulatory protein 1) was detected for wild type Sam68. Information about these proteins such as Swiss-Prot entry name, Accession number, Gene name, Recommended Name and Nominal mass (Mr) are from two websites: **http://www.genecards.org/** and **http://www.uniprot.org/**

P439R mutant Sam68

Swiss-Prot entryname: KHDR1_HUMAN   Accession number: Q07666

MQRRDDPAARLTRSSGRSCSKDPSGAHPSVRLTPSRPSPLPHRSRGGGGGPRGGARASPATQPPPLL
PPSNPGPDATVVGSAPTPLLPPSATAAAKMEPENKYLPELMAEK**DSLDPSFTHAMQLLSVEIEK**IQKG
ESK**KDDEENYLDLFSHK**NMKLKERVLIPVKQYPK**FNFVGK**ILGPQGNTIKRLQEETGAKISVLGKGSM
RDKAKEEELRKGGDPKYAHLNMDLHVFIEVFGPPCEAYALMAHAMEEVKKFLVPDMMDDICQEQF
LELSYLNGVPEPSRGRGVSVRGRGAAPPPPPVPRGRGVGPPRGALVRGTPVRGSITRGATVTRGVPP
PPTVRGAPTPRARTAGIQRIPLPPTPAPETYEDYGYDDSYAEQSYEGYEGYYSQSQGESEYYDYGHGEL
QDSYEAYGQDDWNGTRPSLKAPPARPVKGAYREHPYGRY

CCAR1 (cell cycle and apoptosis regulatory protein 1)

Swiss-Prot entryname: CCAR1_HUMAN   Accession number: Q8IX12

MAQFGGQKNPPWATQFTATAVSQPAALGVQQPSLLGASPTIYTQQTALAAAGLTTQTPANYQLTQ
TAALQQQAAAAAAALQQLQQPQQTLLTQPAVALPTSLSLSTPQPTAQITVSYPTPR**SSQQQTQPQQ
RVFTGVVTK**LHDTFGFVDEDVFFQLSAVK**GKTPQVGD**RVLVEATYNPNMPFKWNAQR**IQTLPNQN
QSQTQPLLK**TPPAVLQPIAPQTTFGVQTQPQPQSLLQAQISAASITPLLQTQPQP LLQQPQQK**AGLLQ
PPVRIVSQPQPAR**RLDPPSRFSGRNDRGDQVPNRKDDRSRERERERRRSRERSPQRKRSRERSPRRE
RERSPRRVRRVVPRYTVQFSK**FSLDCPSCDMM**ELRRRYQNLYIPSDFFDAQFTWVDAFPLSRPFQLG
NYCNFYVMHREVESLEK**NMAILDPPDADHLYSAKVMLMASPSMEDLYHKSCALAEDPQELRDGFQ
HPARL**VKFLVGMKGKDEAMAIGGHWSPSLDGPDPEKDPSVLIK**TAIRCCK**ALTGIDLSVCTQWYRF
AEIR**YHRPEETHKGRT**VPAHVETVVLFFPDVWHCLPTR**SEWETLSR**GYKQQLVEKLQGERK**EADGEQ
DEEEKDDGEAKE**IST**PTHWSK**LDPKTMKVNDLRKELESRALSSKGLKSQLIARLTKQLKVEEQKEEQKE
LEKSEKEEDEDDDRKSEDDKEEEERKRQEEIERQRRERRYILPDEPAIIVHPNWAAKSGKFDCSIMSLSV
LLDYRLEDNKEHSFEVSLFAELFNEMLQRDFGVRIYKSLLSLPEKEDKKEKDKKSKKDERKDKK**EERDDE
TDEPKPK**RRKSGDDKDKKEDRDERKKEDKRKDDSKDDDETEEDNNQDEYDPMEAEEAEDEEDDR
DEEEMTKRDDKR**DINRYCKERPSK**DKEKEKTQM**ITINR**DLLMAFVYFDQSHCGYLLEKDLEEILYTLGL
HLSRAQVKKLLNKVVLRESCFYRK**LTDTSKDEENHEESESLQEDMLGNR**LLLPTPTVK**QES**

RNA helicase A

Swiss-Prot entryname: DHX9_HUMAN   Accession number: Q08211

MGDVKNFLYAWCGKRKMTPSYEIRAVGNKNRQKFMCEVQVEGYNYTGMGNSTNKKDAQSNAAR
DFVNYLVRINEIK**SEEVPAFGVASPP**PLTDTPDTTANAEGDLPTTMGGPLPPHLALK**AENNSEVGASG
YGVPGPTWDRGANLK**DYYSR**KEEQEVQATLESEEVDLNAGLHGNWTLENAKARL**NQYFQKEKIQGE
YKYTQVGPDHNR**SFIAEMTIYIK**QLGRR**IFAREHGSNK**KLAAQSCALSLVRQLYHLGVVEAYSGLTKKK
EGETVEPYK**VNLSQDLEHQLQNIIQELNLEILPPVSIKLFSSHLHPSVPVALNIGK**LAQFEPSQR**QNQVG
VVPWSPPQSNWNPWTSSNIDEGPLAFVSAILFLRSESCVLLLRLYLISMSSVTFLMQILQERELLPVKKF
ESEILEAISQNSVVIIRGATGCGKTTQVPQFILDDFIQNDR**AAECNIVVTQPRR**ISAVSVAERVAFERGE
EPGK**SCGYSVRFESILPRPHASIMFCTVGVLLRKLEAGIR**GISHVIVDEIHER**DINTDFLLVVLR**DVVQAY
PEVR**IVLMSATIDTSMFCEYFFNCPIIEVYGRTYPVQEYFLEDCIQMTHFVPPPK**DKKKKDKDDDGGED
DDANCNLIC**GDEYGPETR**LSMSQLNEKETPFELIEALLKYIETLNVP**GAVLVFLPGWNLIYTMQK**HLEM**

Figure 23 : Mass spectrometry results for P439R mutant Sam68 interacting proteins. This picture shows the sequences of identified proteins as interacting with the P439R mutant Sam68 protein. Matched peptide sequences are shown in red. Sam68, RHA and CCAR1 were detected as the interacting protein partners. The peptide coverages for identified proteins by Mass Spec. were 9%, 26% and 29% for P439R Sam68, CCAR1 and RHA respectively.

	Swiss-Prot entry name	Accession number	Gene name	Recommended Name	Nominal mass(Mr)
1	KHDR1_HUMAN	Q07666	KHDRBS1/Sam68	KH domain-containing, RNA-binding, signal transduction-associated protein 1	48456
2	CCAR1_HUMAN	Q8IX12	CCAR1	Cell division cycle and apoptosis regulator protein 1	133338
3	DHX9_HUMAN	Q08211	DHX9/RHA	ATP-dependent RNA helicase A	115243

Table 13: Protein partners of P439R mutant Sam68 identified by FLAG immunoprecipitation. Based on mass spectrometry results, the presence of Sam68 in the precipitated proteins shows the IP has worked. In addition to proving the presence of Sam68, two new candidate proteins called CCAR1 (cell cycle and apoptosis regulatory protein 1) and RHA (RNA helicase A) were detected for P439R Sam68. Information about these proteins such as Swiss-Prot entry name, Accession number, Gene name, Recommended Name and Nominal mass (Mr) are from two websites: http://www.genecards.org/ **and** http://www.uniprot.org/

2.3.8 Confirmation of the presence of RNA Helicase A (RHA) in T-STAR, wild type and P439R mutant Sam68 IP samples

After sending cut bands from the SDS-PAGE gel for mass spectrometry, I next carried out experiments to test the presence of RHA in each of the IP samples. Therefore, Western blots of all samples were probed with rabbit anti DHX9 (RHA) antibody. The presence of this protein (RHA) in all IP samples including T-STAR, wild type Sam68 and even P439R mutant Sam68 was detected (figure 24) but not in the IPs from stable cell lines made with empty pcDNA5. The immunoprecipitated RHA migrated at exactly the predicted molecular weight for RHA (140kDa). Rabbit anti DHX9 (RHA) and anti-rabbit conjugated to HRP were used as primary and secondary antibodies respectively.

Although based on mass spectrometry sequencing results RHA was detected as a new candidate interacting protein partner of T-STAR and the P439R mutant Sam68, using Western blotting and the RHA antibody, the presence of this protein in wild type Sam68 IP samples was detected as well.
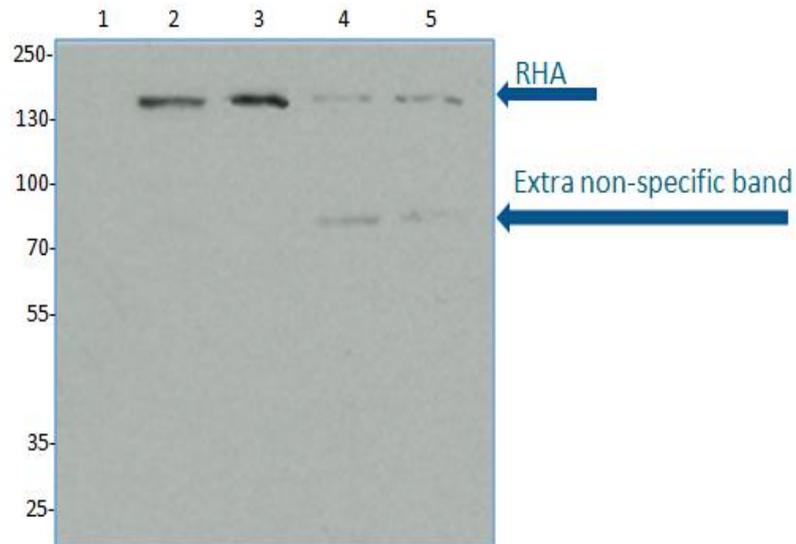
Figure 24: Western blot which confirms the presence of RHA in T-STAR, wild type and P439Rmutant Sam68 cell lysate IP samples. This experiment shows the presence of RHA protein in T-STAR, P439R mutant and even wild type Sam68 cell lysate IP samples. Although based on mass spectrometry sequencing results RHA was detected as a new candidate interacting protein partner of only T-STAR and P439R mutant Sam68, but interestingly, by using Western blot and RHA antibody, presence of this protein in wild type Sam68 IP samples was detected as well. Rabbit Anti-DHX9 (RHA) and anti-rabbit IgG (HRP) were used as primary and secondary antibodies respectively. (1): negative control (pcDNA5), (2): positive control (HEK293 cell lysate), (3): T-STAR IP sample, (4): wild type Sam68 IP sample, (5): P439R mutant Sam68 IP sample. The size marker is in kDa.

2.3.9   Confirmation of the presence of the CCAR1 protein in wild type and P439R mutant Sam68 IP samples

After cutting bands from the SDS PAGE gel and sending them for mass spectrometry, the next step was confirmation of the presence of CCAR1 protein in the IP samples. Therefore, to test the presence of CCAR1 protein in the wild type and P439R mutant Sam68 IP samples, these samples were probed with anti CCAR1 antibody after Western blotting. The presence of this protein in all IP samples including wild type and P439R mutant was detected (figure 25). Rabbit anti CCAR1 and anti-rabbit-HRP were used as primary and secondary antibodies respectively. No CCAR1 was detected in the negative control (pcDNA5, empty vector). Immunoprecipitated CCAR1 migrated at exactly the predicted molecular weight of 140kDa.
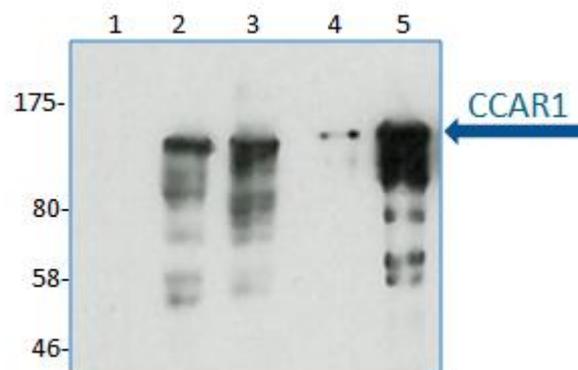
Figure 25: Confirming the presence of CCAR1 protein in wild type and mutant Sam68 cell lysate IP samples. To test the presence of CCAR1 in wild type and P439R mutant Sam68 IP samples, these samples were probed with anti CCAR1 antibody. This picture shows the presence of CCAR1 in both wild type and P439R mutant Sam68 cell lysate IP samples. Rabbit anti CCAR1 and anti-rabbit IgG (HRP) were used as primary and secondary antibodies respectively. (1): negative control (pcDNA5), (2 and 3): positive controls (HEK293 cell lysate and HeLa cell lysate), (4): wild type Sam68 IP sample, (5): P439R mutant Sam68 IP sample. The size marker is in kDa.

2.4   Summary and discussion

The aim of this Chapter was to establish stable cell lines expressing FLAG tagged proteins (T-STAR, wild type and P439R mutant Sam68), and to use them to investigate their interacting proteins partners. To achieve this, the open reading frames of these RNA binding proteins were cloned in to the pcDNA5 vector. This vector encodes a FLAG epitope. These constructs were then transfected into the Flp-In HEK293 cells for integration into the genome. After making stable cell lines and inducing gene expression by using tetracycline antibiotic it was possible to express the protein of interest.

I firstly used these stable cell lines for cell biological investigation into the localization of wild type and P439R mutant Sam68 after induction in the Flp-In HEK293 cells. Wild type Sam68 is nuclear and P439R mutant Sam68 protein is completely cytoplasmic. These results were similar to the previous study carried out by Reddy et al in HeLa cells (Reddy, 2000). The point mutation (P439R, changing proline to arginine) in the tyrosine-rich C-terminal domain of Sam68 changes the localization of Sam68 protein, making Sam68 a completely cytoplasmic protein.

Based on these findings, the next step was to see whether or not changing the localization of nuclear Sam68 to the cytoplasm would change the Sam68 interacting protein partner network in Flp-In HEK293 cells. Therefore, the next steps were to purify FLAG tagged proteins (T-STAR, wild type and P439R mutant Sam68) and their partners using immunoprecipitation and to identify the immunoprecipitated proteins using mass spectrometry.

Based on mass spectrometry sequencing results, in addition to proving the presence of Sam68, CCAR1 protein was detected in the wild type Sam68 IP lane. Also, two potential proteins, CCAR1 and RHA were detected in the P439R mutant Sam68 IP lane. As it has been previously shown that Sam68 and RHA can interact with each other (Reddy et al.,

2000) this confirmed the IP approach was identifying *bona fide* targets. But this was the first time that CCAR1 was detected as a potential interacting partner for Sam68.

Based on mass spectrometry results, RHA was detected only as a partner of the P439R mutant Sam68 IP. Interestingly, by using Western blot and DHX9 (RHA) antibody, RHA protein was found to be a partner of wild type Sam68 as well. To prove the presence of CCAR1 in wild type and mutant Sam68 IP samples, these samples were probed with anti CCAR1 antibody and the presence of this protein in all IP samples including wild type and P439R mutant Sam68 was detected. Hence, in terms of interacting protein partners of both wild type and P439R Sam68, there was no difference detected between wild type and mutant versions of Sam68.

In addition to proving the presence of T-STAR in the immunoprecipitates two other proteins (hnRNP G and RNA helicase A (RHA)) were detected as T-STAR partners by mass spectrometry. The expected presence of T-STAR showed that the immunoprecipitation had worked. In addition, the interaction between hnRNP G and T-STAR has already been described. For example, their interactions to regulate CD44 minigene alternative splicing have been reported (Venables *et al.*, 1999; Stoss *et al.*, 2001). Therefore, it was encouraging to see hnRNP G protein in the IP sample as one of the T-STAR interacting protein partners. But it was the first time that RHA was detected as an interacting protein partner for T-STAR. The presence of RHA in the T-STAR IP samples was confirmed by using an anti DHX9 (RHA) antibody and Western blot.

RHA is one member of the DEXH helicase superfamily (Lee and Hurwitz, 1993; Zhang et al., 1995). DNA or RNA helicases are important in resolving double stranded of nucleic acids (Zhang and Grosse, 2004). Using an RNA unwinding activity, the members of the DExD/H family are involved in different roles including RNA secondary structure modulation, translation and RNA transportation (Rocak and Linder, 2004; Fuller-Pace, 2006). While the helicase parts of these proteins are highly conserved, the N- and C-

terminal domains of the DExD/H family members are different. These diversities seem
to be important in their interactions (Fuller-Pace, 2006).

RNA helicase A has been detected in human cells (Lee and Hurwitz, 1992), and an
orthologue of this protein has been detected in *Drosophila melanogaster* called MLE
(Kuroda et al., 1991). In Drosophila MLE protein has some roles in dosage
compensation of the X chromosome (Kuroda et al., 1991). It seems that the role of
MLE is to increase the transcriptional activity of X chromosome in male Drosophila
(Kuroda et al., 1991). RHA also is called nuclear DNA helicase II (NDH II) (Friedemann et
al., 2005). RHA also interacts with SMN which is important for small nuclear
ribonucleoprotein assembly (snRNPs) (Pellizzoni et al., 2001; Terns and Terns, 2001).
RHA is a component of the prespliceosomes (Hartmuth et al., 2002) and its localization
is similar to hnRNPs (Zhang et al., 1999b). This protein is localized mostly in the
nucleoplasm in human cells and the nucleolus in mouse cells (Zhang et al., 1999a).

RHA and Sam68 play a role in RNA transportation (Reddy et al., 2000). There are two
routes for nuclear-cytoplasmic transportation of retroviral mRNA. Complex
retroviruses such as HIV (Human Immunodeficiency Virus) have a protein called Rev
containing a sequence called a NES (Nuclear Export Sequence). The NES can bind to a
RRE (Rev Response Element) in RNA and to the nuclear export receptor called CRM1. In
this way, Rev protein acts as an adaptor between nuclear export receptor and viral
mRNA (Pollard and Malim, 1998). The second route of retroviral RNA transport is
mediated by the CTE (Constitutive Transport Element) of type D retroviruses (Bray *et
al.*, 1994; Zolotukhin *et al.*, 1994; Tabernero *et al.*, 1996; Ernst *et al.*, 1997; Pasquinelli
et al., 1997; Saavedra *et al.*, 1997). RHA is able to bind to the CTE of retroviral RNAs
(Tang et al., 1997). The CTE is involved in promotion of nuclear export of viral RNA and
it seems that RHA has a role in this transportation. There is a C-terminal region in the
structure of RHA which is involved in translocation across the nuclear membrane (Tang
et al., 1999). In addition, there is another protein called Tap (an RNA binding protein)
which can bind to the CTE (Yoon *et al.*, 1997; Gruter *et al.*, 1998). RHA and Tap interact

with each other (Tang and Wong-Staal, 2000) . Therefore, it seems that these two proteins are involved in nuclear transportation of CTE-containing RNAs (both viral and cellular RNAs) (Reddy et al., 2000).

On the other hand, it has been reported that Sam68 can independently of Rev protein mediate transportation of RNA (Reddy et al., 1999). RHA binds to Sam68 and Tap, and all of these proteins are involved in nuclear transportation of RNAs which contain either an RRE or CTE. For example while there is a Rev protein pathway for HIV as an efficient way for nuclear export of RNA, this retrovirus and D retroviruses can use another pathway which include Sam68,Tap and RHA. It seems that this pathway is used by cellular mRNA as well (Reddy et al., 2000).

In addition to the roles of RHA in RNA transportation and processing (Tang et al., 1997), RHA protein is involved in other activities as well including translation of mRNA (Hartman et al., 2006), transcription (Kuroda et al., 1991; Nakajima et al., 1997), DNA repair (Schlegel et al., 2003) and tumorigenesis (Abdelhaleem, 2004). Generally, members of the DExD/H family have roles in transcription regulation through recruiting transcription factors and stabilization of the transcriptional initiation complex (Fuller-Pace, 2006). For example, RHA is involved in activation of transcription by interacting with CBP (CREB-binding protein) and RNA polymerase II. RHA acts as a bridge between these two components (Nakajima et al., 1997). RHA also interacts with topoisomerase IIα and double-stranded DNA, and it can also activate transcription by unwinding the structure of chromatin in transcription activated site (Zhou *et al.*, 2003; Fuller-Pace, 2006).

Interestingly, it has been shown that RHA acts as a bridge between Pol II holoenzyme and BRCA1 (breast cancer-specific tumour suppressor) (Miki *et al.*, 1994; Rowell *et al.*, 1994; Anderson *et al.*, 1998). The transcription of some genes involved in apoptosis, cell cycle and DNA repair is regulated by BRCA1 (Starita and Parvin, 2003; Yoshida and

Miki, 2004; Rosen *et al.*, 2006). It hence seems that RHA has both specific and general
roles in transcription (Fuller-Pace, 2006).

Also there are some known roles for RHA protein in cancer as well. For example, RHA is
over-expressed in lung cancer cells (Wei et al., 2004). RNA helicase A by increasing the
transcription level of MDR1 (multidrug-resistant) gene in cancer cells seems to help
cancer cells to be resistant to chemotherapeutic treatments (Zhong and Safa, 2004).

There are two RNA binding domains (dsRBDs) and one RGG box in the N and C termini
of RHA respectively for binding to RNA (Zhang and Grosse, 1997). In RHA, the dsRBDs
regardless of their roles in binding to RNA are involved in protein-protein interactions
as well (Reichman et al., 2003). On the other hand, the RGG box in RHA is involved in
physical contacts with different proteins such as RNA polymerase II, small nuclear
ribonucleoproteins, and spliceosomes (Pellizzoni et al., 2001; Terns and Terns, 2001). It
seems that in addition to the roles of RHA protein in transcription regulation, RHA is
involved in the spliceosome assembly (Pellizzoni et al., 2001).

I also detected another protein in wild type and P439R mutant Sam68 IP samples
called CCAR1 (cell cycle and apoptosis regulatory protein is also known as CARP-1
(Jiang et al., 2010)). CCAR1 is involved in gene expression regulation (Kim et al., 2008).
One group of transcription activators are called nuclear receptors (NRs) including
steroid and thyroid receptors (Tsai and O'Malley, 1994). Also, some co-activators have
the ability to change chromatin structure and activate RNA polymerase II, and can be
recruited by nuclear receptors (Kim et al., 2008). p160 co-activators after binding to
nuclear receptors function as scaffolds for gathering co-activator complexes on
promoters (Stallcup et al., 2003). p160 co-activators also have the role of recruiting
other co-activators such as CoCoA (Coiled-coil coactivator), histone methyltransferases
and histone acetyltransferases such as CARM1 (Coactivator-Associated Arginine
Methyltransferase 1) and CBP/p300 respectively (Kim et al., 2003; Stallcup et al., 2003).
Generally, a set of co-activator complexes make a group of proteins called the

Mediator. The proteins in the Mediator complex interact with nuclear receptors and other transcription activators (Fondell *et al.*, 1996; Ito *et al.*, 1999; Naar *et al.*, 1999; Kang *et al.*, 2002). CCAR1 not only is involved in coordination with p160 co-activators, but also has role in recruiting the Mediator complex. CCAR1 has direct and indirect interactions (through CoCoA) with nuclear receptors (Kim et al., 2008) (figure 26).

Since Sam68 and P300/CBP also interact with each other (Hong *et al.*, 2002), there is a possibility to consider Sam68 as a mediator between CCAR1 and P300/CBP. It seems that presence and interaction of Sam68 with these proteins might be able to play a role to recruit other co-activators. This idea needs to be tested by more experiments. Figure 26 shows the role of CCAR1 in transcription. The possible role of Sam68 also is shown in this figure: that Sam68 also acts to recruit CCAR1 and P300.
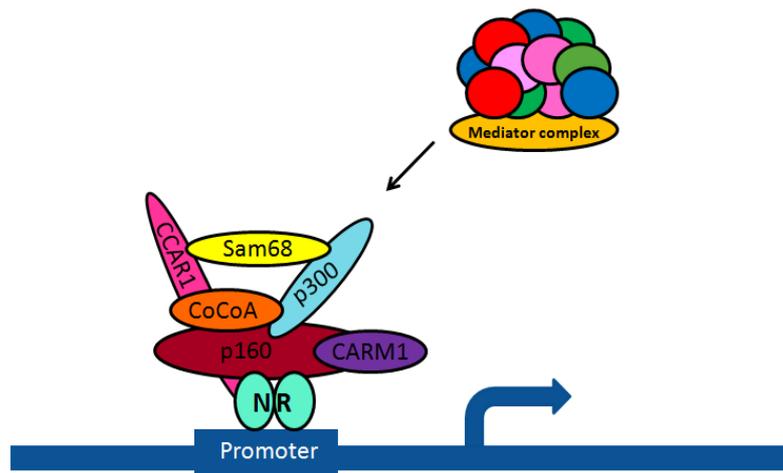
Figure 26 : Role of CCAR1 in transcription. CCAR1 coordinates with both p160 co-activators and Mediator complex. CCAR1 recruits the Mediator complex. CCAR1 has also a direct and indirect interaction (through CoCoA) with nuclear receptors (Kim et al., 2008). NR: nuclear receptors. Possible role of Sam68 in having interaction with CCAR1 and P300/CBP plus recruiting other co-activators of transcription is shown in the picture.

CCAR1 acts as a P53 co-activator (Kim et al., 2008). CCAR1 has roles in apoptosis, cell proliferation regulation (Ou et al., 2009) and breast cancer (Kim et al., 2008). CCAR1

functions as an intracellular transducer in these pathways (Ou et al., 2009). The Wnt/β-catenin signalling pathway is involved in different kinds of cancers especially colorectal cancers. Some target genes of Wnt/β-catenin signalling pathway have roles in cell proliferation regulation. CCAR1 seems to help β-catenin in increasing transcription level of these genes. Interestingly, CCAR1 seems to have a role in controlling the growth of cancer cells for example in colon cancer cells (Ou et al., 2009).

To summarise, in this project RHA was detected as interacting with T-STAR, wild type and P439R mutant Sam68. In addition, CCAR1 interacts with both wild type and P439R mutant Sam68. Interestingly although wild type and P439R mutant Sam68 have different localization, both have same interacting partners (CCAR1 and RHA). To follow up these results I needed to check whether or not the interactions of these proteins might happen only after induction and over-expression of the gene of interest, or if these interactions happen at endogenous levels of expression as well. To achieve this aim I carried out more experiments. In the next Chapter more investigations regarding T-STAR, wild type and P439R mutant Sam68 interactions with endogenous CCAR1 and RHA will be described.

Chapter 3

3 Investigating the functional importance of T-SAR, wild type and P439R Sam68

3.1 Introduction

In the previous Chapter, I described some new interacting partners for T-STAR, wild type and P439R mutant Sam68. These partner proteins were RHA (RNA helicase A) which interacted with T-STAR and P439R mutant Sam68 proteins, and CCAR1 (Cell division cycle and apoptosis regulator protein 1) which interacted with both wild type and mutant Sam68 immunoprecipitated proteins. Both of these proteins, as mentioned in the previous Chapter, have important roles within the cell.

Since the interactions between the proteins were only observed after over-expression of T-STAR or Sam68, it was necessary to check for interactions between these proteins expressed at endogenous levels. The first aim of this Chapter was to test the interactions of endogenous RHA and CCAR1 with T-STAR and Sam68. My second aim was to get more information regarding the localization of these proteins within cells, which would enable us to achieve more information about the roles of these protein interactions. The third aim was to identify the interacting domains by making short forms of these proteins and seeing if these also interacted.

Addressing these aims would enable us to learn more about the interaction networks of T-STAR and Sam68. The final aim of this Chapter was investigation of possible links

between Sam68 and RHA in terms of their roles in the transcription and alternative splicing regulation of some target genes.

There are some studies showing the role of Sam68 and RHA in transcription regulation. One possibility that Sam68 is involved in transcription is through binding the Cap Binding Protein (CBP). Sam68 and CBP interact and co-localise with each other (Hong *et al.*, 2002; Babic *et al.*, 2004). CBP and P300 act as transcriptional adaptors (Blobel, 2000; Goodman and Smolik, 2000; Chan and La Thangue, 2001). CBP has a histone acetyltransferase activity (Bannister and Kouzarides, 1996). It seems that Sam68 competes with other cofactors to bind to CBP and after binding, this RNA binding protein acts such a bridge to link CBP to TF (transcription repressor factor). In this way Sam68 represses the transcription of CBP targets (Bielli et al., 2011) (figure 27).
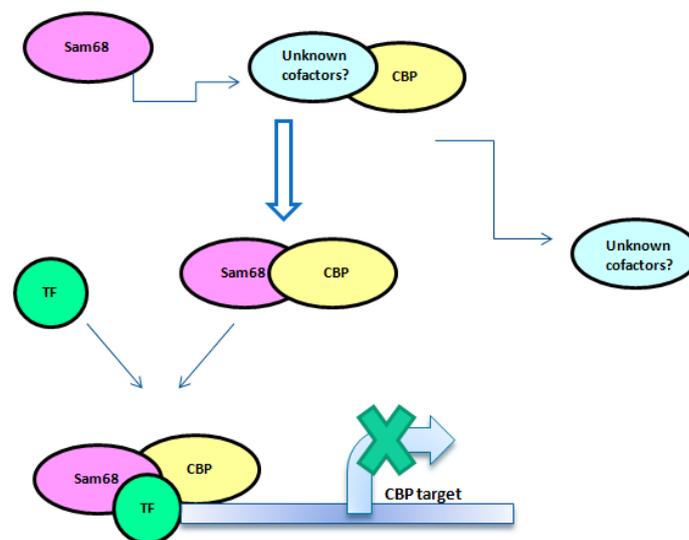
Figure 27: Role of Sam68 in repression of CBP targets. Sam68 might compete with other cofactors to bind to CBP. Sam68 acts like a bridge to link CBP and transcription repressor factor (TF). In this case, Sam68 has repression role in transcription of CBP targets. The figure is modified from (Bielli et al., 2011).

RHA and CBP also interact with each other. In addition, RNA helicase A belongs to the holoenzyme complex which contains RNA polymerase II as well (Nakajima et al., 1997). In another experiment using glycerol density gradient analysis RHA and CBP were detected as interacting partners but RNA polymerase II did not co sediment with RHA (Kitagawa et al., 2002). Since both of these proteins, Sam68 and RHA are involved in transcription I also wanted to know whether RHA and Sam68 interact with each other to regulate transcription or not.

RHA seems to be involved in transcription regulation not only by changing the chromatin structure but also by acting like a bridge to mediate interaction between RNA polymerase II and CBP (Nakajima et al., 1997). RNA helicase A has similar role to mediate interactions between Breast cancer-specific tumor suppressor protein 1 (BRCA1) (a transcription factor) and RNA polymerase II (Anderson et al., 1998). Transcription of some genes which have roles in apoptosis, cell cycle and chromatin remodelling is regulated by BRCA1 (Mullan et al., 2006).

One of the genes that RHA has a role in its transcription regulation is *cyclin D1*. It has been shown that in breast cancer cells there is a positive correlation between the expression level of *RHA* , *cyclin D1* and *EGFR* (epidermal growth factor receptor) (Huo et al., 2010). After binding of epidermal growth factor (EGF) to EGFR, this receptor translocates to the nucleus. ATRS (AT-rich sequence) within the promoter of target genes are bound by EGFR. This binding can lead to transcription activation of the target gene. Huo et al showed the presence of EGF leads to interaction of RHA and EGFR and these two proteins make a complex. RNA helicase A then binds to the ATRS and acts as a mediator (Huo et al., 2010) (figure 28).
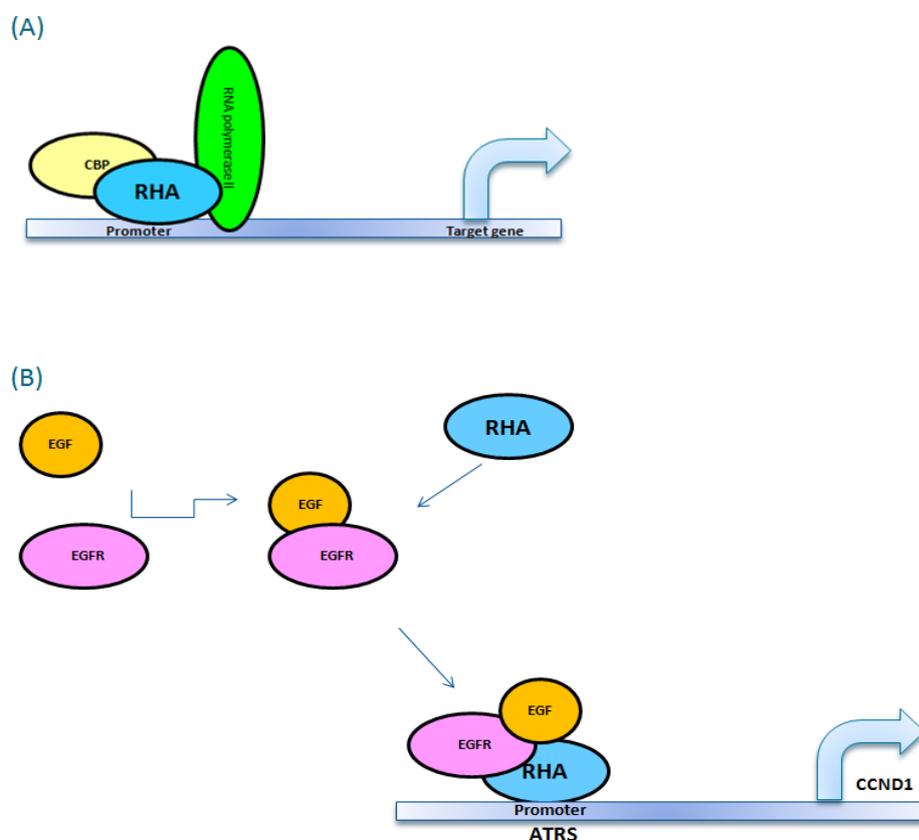
Figure 28: Role of RHA in transcription regulation. A) RHA seems to be involved in transcription regulation by acting as a bridge to mediate an interaction between RNA polymerase II and CBP (Nakajima et al., 1997). B) In breast cancer after binding of EGF to EGFR, this receptor translocates to the nucleus. Then RNA helicase A binds to the ATRS within the promoter and acts as a mediator. This binding can lead to transcription activation of the target gene. (Huo et al., 2010).

I planned to check the role of these two proteins (Sam68 and RHA) in transcription regulation of the same target gene (*cyclin D1/CCND1*). *Cyclin D1* has an important role in the regulation of the cell cycle and is involved in different kind of cancers such as breast cancer (Xu et al., 2013). On the other hand, Sam68 seems to be over expressed in breast and prostate cancers (Elliott and Rajan; Busa *et al.*, 2007; Rajan *et al.*, 2008b;

Elliott and Rajan, 2010). Also, in breast cancer cell lines *RHA* and *cyclin D1* expression level have positive correlations with each other (Huo et al., 2010).

To extend my studies on the RHA and Sam68 interaction, I also looked at their role in alternative splicing regulation of a target gene called *NRXN3*. By using these different approaches there would be an opportunity to investigate Sam68 interaction networks and their roles within the cell.

Chapter 3: Further investigation into the interactions of T-STAR, wild type Sam68 and P439R Sam68

3.2   Methods and Materials

3.2.1   *Standard molecular biology*

Some of standard molecular biology techniques including restriction digestion, purification of DNA and plasmid DNA, ligations, re-cleavage, dialysis, agarose gel electrophoresis and sequencing are described in the methods section in the previous Chapter (Chapter2, method section).

3.2.1(a) PCR amplification of short forms of Sam68 and RHA

The PCR amplification protocols to make short forms of Sam68 comprising the N-KH, N-KH-RG, KH-RG and RG-tyrosine rich domains were identical to the *Sam68* PCR amplification described previously (more information is available in Chapter 2, method section, tables 2 and 5).

RHA was amplified by PCR using a similar combination of reagents and protocol to that used for *T-STAR* amplification (previously described in Chapter 2, method section, tables 1 and table 4). The primers used for these amplifications are shown in table 14.

Primer name	Sequence (5'-3')	Restriction Enzyme site
WSam68 F-KH-RG-BamHI	AAA AAA AAA GGATCC TTCAATTTTGTGGGGAAGATTCTTGGACCA	*BamH*I
WSam68 F-RG-Tyrosine BamHI	AAA AAA AAA GGATCC CCTGAACCCTCTCGTGGTCGTG	*BamH*I
Wt MSam68 R-SalI	AAAAAAAAAGGATCCTTAATAACGTCCATATGGATGCTCTCTGTATG	*Sal*I
WSam68 R-N-KH-SalI	AAAAAAAAAGTCGACTTATTCCTCCTTGGCTTTGTCTCTCATTG	*Sal*I
WSam68 R-N-KH-RG-SalI	AAAAAAAAAGTCGACTTATGTCCGAGCTCTTGGTGTTGG AG	*Sal*I
WRHA-F-BamHI	AAA AAA AAA GGATCC ATGGGTGACGTTAAAAATTTTCTGTATGCCTG	*BamH*I
WRHA-R-SalI	AAA AAA AAAGTCGACTTAATAGCCGCCACC TCCTCTTCC	*Sal*I
MSam68 F-BamHI	AAAAAAAAAGGATCC ATGCAGCGCCGGGACGATCCT	*BamH*I
BGHrev	TAGAAGGCACAGTCGAGG	---------

Table 14: Primers used to make RHA and short forms of Sam68.

3.2.1(b) Heat shock transformation

10μl of the plasmid DNA was added to 100μl of *E. coli* cells, DH5α (provided in Professor Elliott's lab) and mixed. The sample was left on ice for 30 minutes and then incubated firstly at 42°C for 1 minute and secondly on ice for 2 minutes. 1 ml of the LB medium was then added, and the sample was incubated at 37°C for 1 hour. After centrifugation, the excess LB medium was removed and the cell pellets were mixed with the remaining medium. These resuspended cells were then plated out on the LB agar plates which contained 50μg/ml ampicillin. The plates were incubated at 37°C overnight (LB broth and LB Agar mixtures are shown in Chapter2, method section, table 8).

3.2.2 Cell culture (HEK 293 cells)

HEK293 cells (provided in Professor Elliott's lab, IGM, Newcastle University) were seeded in 10% FBS DMEM (Dulbecco's MEM) with glutamax-1 medium. They were incubated at 37°C in 5% CO_2. The Generation of stable cell lines and harvesting cells are described previously (Chapter2, method section).

3.2.3 Protein based methods

SDS polyacrylamide gel electrophoresis (SDS-PAGE), Western blot (Protein immunoblot analysis), immunoprecipitation using α-FLAG affinity resin, and immunofluorescence have been already described (Chapter2, method section). Table 15 shows the antibodies used in Western blot, immunofluorescence and immunoprecipitation.

Chapter 3: Further investigation into the interactions of T-STAR, wild type Sam68 and P439R Sam68

Primary Antibody	Source	Monoclonal /Polyclonal	Species	Dilution for Western Blot	Dilution for immunofluorescence
Anti T-STAR	Made in Elliott's Lab	Polyclonal	Rabbit	1:1000	-
Control IgG	Santa Cruz	Polyclonal	Rabbit	1:1000	-
Anti Sam68 (C-20)	Santa Cruz	Polyclonal	Rabbit	1:1000	-
Rabbit Anti DHX9 (RHA)	ProteinTech Group	Polyclonal	Rabbit	1:2000	1:200
Rabbit anti CCAR1	Sigma	Polyclonal	Rabbit	1:500	1:200
Anti hnRNP L(4D11)	Sigma	Monoclonal	Mouse	-	1:200

Table 15: Primary antibodies. Antibodies used for Western blot and immunofluorescence.

Secondary Antibody	Source	Monoclonal /Polyclonal	Species	Dilution for Western Blot	Dilution for immunofluorescence
Anti-mouse-HRP	Amersham	Polyclonal	Sheep	1:1000	-
Anti-rabbit-HRP	Jackson lab	Polyclonal	Goat	1:1000	-
Anti-rabbit-488(Green)	Molecular Probes	Polyclonal	Donkey	-	1:400
Anti-mouse-594(Red)	Molecular Probes	Polyclonal	Donkey	-	1:400

Table 16: Secondary antibodies used for Western blot and immunofluorescence.

3.2.3(a) Immunoprecipitation of endogenous T-STAR and Sam68 proteins using protein A Dynabeads

In order to detect interacting protein partners of T-STAR and Sam68, HEK293 cells were harvested and the IP lysis buffer (50mM Tris pH 7.4, 5mM NaF, 5mM MgCl2, 100mM KCl, 1X Roche Complete Protease Inhibitor Cocktail and 0.1% NP-40) used to resuspend the cell pellets, followed by a 20 minutes incubation on ice. Then, the samples were centrifuged for 15 minutes at 4°C at maximum speed. 30 μl of the

supernatant was kept and labelled as Input. 1200 µl of the soluble cell lysate was added to a 1.5 ml centrifugation tubes containing an appropriate amount of rabbit polyclonal Sam68/rabbit T-STAR (table 15). There were two control groups as well: (i) 1200 µl of the soluble cell lysate was added to a 1.5 ml centrifugation tubes which had appropriate amount of normal rabbit IgG; (ii) in the no antibody control group, cell lysate was incubated with no antibody. All the samples were incubated at 4°C on a rotator overnight.

The following day, a magnetic rack and 250 µl of 0.1M sodium phosphate (pH 8.1) were used for washing 50 µl of protein A Dynabeads (Invitrogen). The cell lysates were added to the beads followed by one hour incubation on a rotator at room temperature. Using a magnetic rack, 30 µl of the supernatants were kept and labelled SupIP. 1X PBS was then used for washing the beads. In order to check the immunoprecipitation of proteins, 2X SDS sample loading dye was added to samples for loading on SDS-PAGE gels, followed by Western blotting.

3.2.4   RNA-based methods

3.2.4(a) RNA extraction

Trizol® (Invitrogen) was used for extracting RNA according manufacturer's instructions. After adding 100µl of Trizol® to the harvested cells, samples were mixed for 10 minutes at the maximum speed on a shaker. Samples were then incubated for 5 minutes at room temperature. This step was followed by the addition of 20µl of chloroform to samples, shaking for a few seconds and incubation for 3 minutes at room temperature. After centrifugation of samples at 4°C for 15 minutes at maximum speed, the clear layers were transferred to new 1.5 ml centrifuge tubes. Nucleic acids

were precipitated by adding 50µl of isopropanol to the tube and incubation of sample for 10 minutes at room temperature, followed by another centrifugation (at 4°C) for 10 minutes.   After removing the supernatant, 75% ethanol-25% nuclease free dH$_2$O (provided in Professor Elliott's lab) was used for washing the RNA pellet. After centrifugation of samples (at 4°C) for 5 minutes, the supernatant was removed and the tube was left with an open top on a hot block (at 65°C) for 5 minutes. Then 30-50 µl of nuclease free dH$_2$O was added to the tube, and the sample was incubated on the hot block (at 55°C) for 5 minutes and then vortexed. A nanoDrop Spectrophotometer (NanoDrop Technologies) was used for quantifying the RNA concentration.

3.2.4(b) RT-PCR (Reverse transcriptase PCR)

A SuperScript® VILO™ cDNA Synthesis Kit (Invitrogen) was used to make cDNA from RNA samples according to the manufacturer's instructions. Table 17 shows the reaction mixture used.

Reagent	Quantity for a 10µl reaction
(Diethylpyrocarbonate) DEPC-treated water (nuclease free dH$_2$O)	Up to 10µl
5X VILO™ Reaction Mix	2µl
10X SuperScriptR Enzyme Mix	1µl
RNA (500ng/µl)	Xµl

Table 17: RT-PCR recipe. Based on SuperScript® VILO™ cDNA Synthesis Kit recipe.

RT-PCR reactions were incubated firstly at 25°C for 10 minutes and this step was followed by two other steps, 42°C for 120 minutes and 85°C for 5 minutes respectively. A DNA Engine DYAD thermal cycler (MJ Research) was used for RT-PCR amplifications.

3.2.4(c) Real time PCR (estimation of Cyclin D1 /CCND1 transcript level) after induction of gene of interest

After RNA extraction, making cDNA from RNA (500ng/µl) and diluting cDNA in a ratio of 1:10, real time PCR was carried out to estimate the level of *cyclin D1*/*CCND1* mRNA in different samples. The amplicon was 81bp in length. The forward and reverse primers had a 10µM concentration in 10µl PCR reactions. SYBR® Green PCR Master Mix (5ml) (part no: 4309155) was used as the master mix. Reagents and primers used in the real time PCR reaction are shown in table 18 and table 19. The company which supplied primers was Integrated DNA Technology (IDT).

Reagent	Quantity for a 10µl reaction
Nuclease free dH$_2$O	2µl
SYBR® Green PCR Master Mix(5ml)	5µl
10µM Forward primer	1µl
10µM Reverse primer	1µl
cDNA (1:10 diluted)	1µl

Table 18: Real time PCR recipe to check level of cyclin D1/CCND1 mRNA in different samples.
The reagents and PCR programmes used in a typical 10µl reaction are shown.

Primer name	Sequence (5'-3')
Cyclin D1 F	TCTACACCGACAACTCCATCC
Cyclin D1 R	TTCCACTTGAGCTTGTTCACC
α Actin (F) SYBR	CATCGAGCACGGCATCGTCA
α Actin (R) SYBR	TAGCACAGCCTGGATAGCAAC
GAPDH (F) SYBR	CATACCAGGAAATGAGCTTGACAA
GAPDH (R) SYBR	AACAGCGACACCCATCCTC
βTubulin (F) SYBR	CTTCGGCCAGATCTTCAGAC
βTubulin (R) SYBR	AGAGAGTGGGTCAGCTGG AA

Table 19: Primers used in real time PCR.

All reactions were repeated three times using an ABI 7900 HT Fast real-time PCR system (Applied Biosystems). The amplification protocol is shown in table 20. The ΔCt method was used to estimate the relative expression level of *cyclin D1*/*CCND1* mRNA. After calculating the average of Ct value of samples, the mean Ct value of endogenous expression of *α Actin, βTubulin* and *GAPDH* (control genes) were used for normalization of the *cyclin D1*/*CCND1* Ct value in different samples. Data were analysed using Excel software (Student's *t*-test).

1	Heat activation	50 °C	2 minutes		
2	Initial denaturation	95°C	10 minutes		
3	Denaturation	95°C	15 seconds		
4	Annealing/elongation	60 °C	1 minute	Cycling to step 3	45

Table 20: PCR Programme for real time PCR.

3.2.4(d) Splicing assay

After RNA extraction, synthesis of cDNA from RNA (500ng/μl) (RT-PCR) and dilution of cDNA (ratio of 1:5), a 25 μl polymerase chain reaction was carried out to measure splicing patterns. The amplification protocol, PCR programmes and the primers used in these reactions are shown in tables 21, 22 and 23 respectively. A DNA Engine DYAD thermal cycler (MJ Research) was used for PCR amplifications. GoTaq® DNA Polymerase and dNTPs were from the Promega company. The company which supplied primers was Integrated DNA Technology (IDT).

Reagent	Quantity for a 25μl reaction
Nuclease free dH$_2$O	17.35μl
5X Green Go Taq Reaction Buffer	5μl
10μM Forward primer	0.5μl
10μM Reverse primer	0.5μl
cDNA (1:5 diluted)	1μl
GoTaq® DNA Polymerase	0.15μl
10mM dNTPs (Deoxyribonucleotide triphosphates)	0.5μl

Table 21: PCR recipe to check alternative splicing patterns of a target gene (*NRXN3*). The reagents and PCR programmes used in a typical 25μl reaction are shown.

Chapter 3: Further investigation into the interactions of T-STAR, wild type Sam68 and P439R Sam68

Step	Function	Temperature	Time	Number of cycles	
1	Heat Activation	94 °C	2 minutes		
2	Denaturation	94°C	1 minute		
3	Annealing	60°C	1 minute		
4	Extension	72 °C	1 minute	Cycling to step 2	30
5	Final Extension	72 °C	10 minutes		
6	Cooling	4°C	∞		

Table 22: PCR Programme used to analyse alternative splicing patterns followed by the target gene (*NRXN3*).

Primer name	Sequence (5'-3')
NRXN3 F	TTGGCACAGTTGACATCTCC
NRXN3 R	GTCCTTTGTCCTTTCCACCA

Table 23: Primers used in the PCR reaction to check the alternative splicing patterns followed by the target genes (*NRXN3*).

Both agarose gel electrophoresis explained previously (Chapter2, method section) and the QIAXCEL multi-capillary electrophoresis system were used for running PCR products. To estimate PCR product sizes and their relative concentrations, the Bio calculator software was used.

3.3   Results

3.3.1   Do RHA and CCAR1 proteins co-localise with Sam68 nuclear bodies (SNBs) within the cell?

After identification of RHA and CCAR1 proteins as candidate interacting proteins for Sam68 and T-STAR, the next aim was to check the localisation of these proteins within the cell. Specifically I was interested to see whether or not these proteins co-localized with Sam68. Sam68 seems to be exclusively nuclear in most somatic cells. Sam68 protein has also been observed in the cytoplasm of secondary spermatocytes. In addition, Sam68 has some interactions with signalling molecules which are known to be cytoplasmic (Lukong and Richard, 2003; Paronetto *et al.*, 2006; Paronetto *et al.*, 2009). Some signalling components and splicing regulators such as Sam68 and hnRNP L (Rajan et al., 2009) are present in Sam68 Nuclear Bodies (SNBs). In cancer cells, Sam68 protein has a general nucleoplasmic distribution and is also concentrated in nuclear SNB structures (Chen et al., 1999b). The literature suggests that RHA and CCAR1 (Fujita *et al.*, 2005; Kotani *et al.*, 2010; Lu *et al.*, 2012) are mostly nuclear.

HeLa cells and appropriate antibodies (tables 15 and 16) were used to check the localisation of RHA, CCAR1 and Sam68. In our lab we had just rabbit polyclonal antibodies against Sam68, CCAR1 and RHA proteins. Therefore, there was no possibility to test the co-localisation of these proteins at the same time in the same cells. Rajan et al in 2009 had shown that Sam68 and hnRNP L are both present within the SNBs (Sam68 Nuclear Bodies) and the nucleoplasm, and co-localise with each other (Rajan et al., 2009). Therefore, a monoclonal mouse hnRNP L antibody was used as a marker instead of a Sam68 antibody, to gauge where Sam68 protein would be.

HeLa cells were grown on coverslips. When cells were confluent coverslips were washed with PBS and then fixed in 4% (w/v) PFA (paraformaldehyde). Cells were then

incubated in 1%Triton X-100 to be permeabilised. Cells were incubated in blocking solution (10% Horse serum/Triton X-100). This step was followed by incubation of cells in primary and fluorescent conjugated secondary antibodies respectively, diluted in blocking solution. PBS was used for washing the cells. Cells were mounted using VectaShield Mounting Medium with 4',6-diamidino-2-phenylindole (DAPI) and pictures were taken by a fluorescence Axioplan2 microscope. None of RHA or CCAR1 were detected in SNBs as labelled by hnRNP L (figure 29 and 30). However, endogenous RHA (mostly), CCAR1 and Sam68 were each localised within the nucleoplasm.
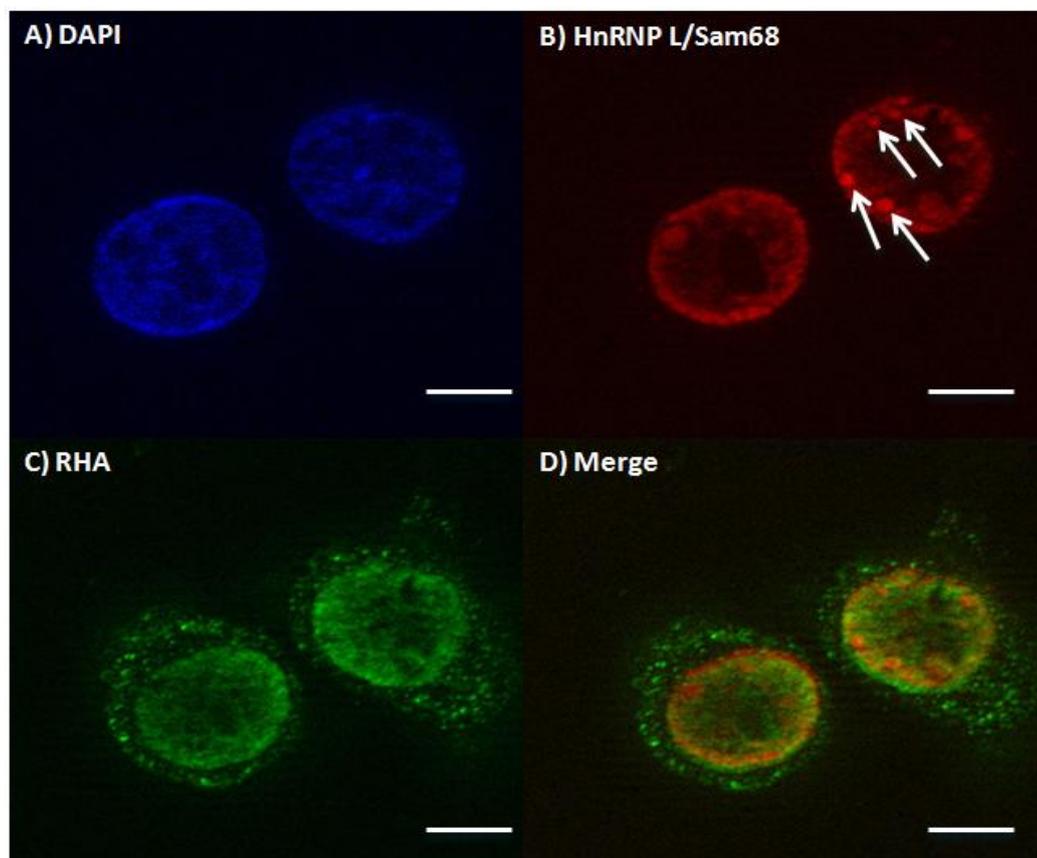
Figure 29 : Localization of endogenous RHA and hnRNP L/Sam68 in HeLa cells. Rajan et al in 2009 showed that Sam68 and hnRNP L are present within the SNBs and the nucleoplasm. They showed these protein co-localise with each other (Rajan et al., 2009). Therefore, monoclonal mouse hnRNP L antibody was used instead of Sam68 antibody to enable co-localization with SNBs. All these proteins including RHA (mostly), hnRNP L /Sam68 are distributed within the nucleoplasm. SNBs (Sam68 Nuclear Bodies) are shown with white arrows. A) DAPI staining of nuclear DNA B) HnRNP L localization. Mouse anti hnRNP L (4D11) and anti-mouse-594 (Red) antibodies were used as primary and secondary antibodies respectively. C) RHA localisation. Rabbit anti DHX9 (RHA) and anti-rabbit-488 (Green) were used as primary and secondary antibodies respectively. D) Merged picture of localization of both hmRNPL /Sam68 and RHA proteins. These results show that although RHA and hnRNP L/Sam68 do not co-localize in SNBs, they co-localise within the nucleoplasm. The scale bars are equal to 20µm.
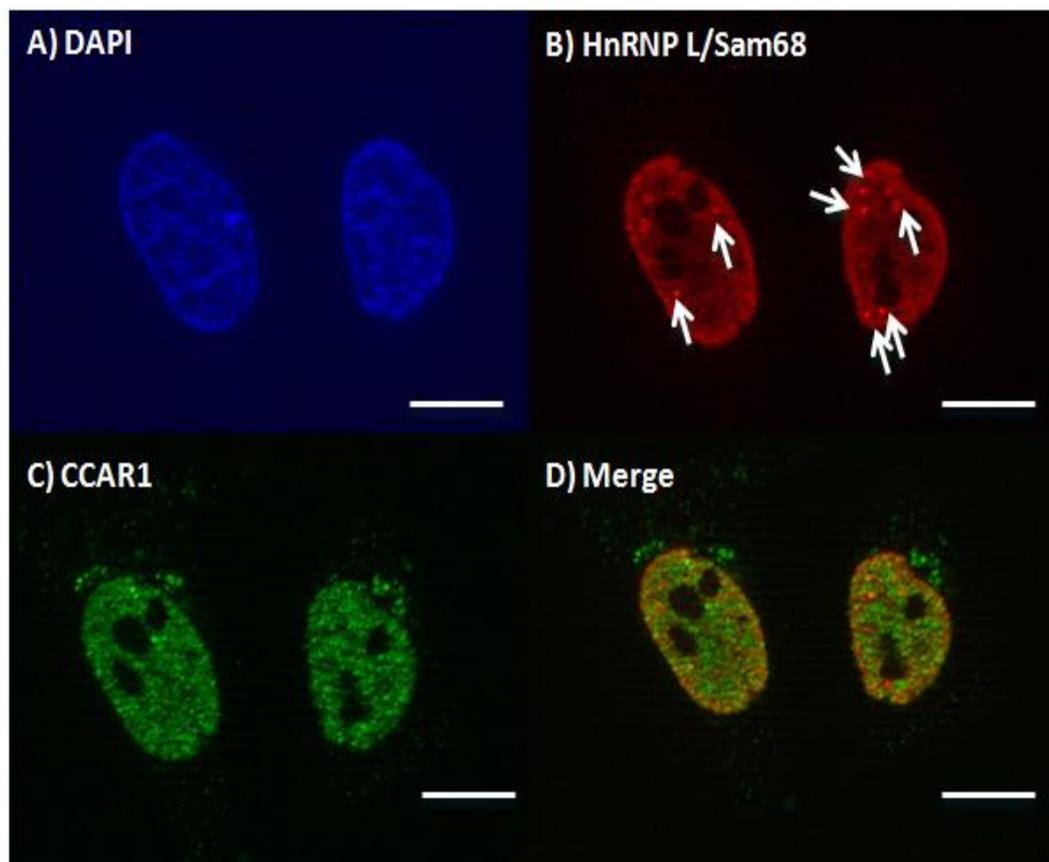
Figure 30: Localization of endogenous CCAR1 and hnRNP L/Sam68 in HeLa cells. Monoclonal mouse hnRNP L antibody was used as a surrogate of Sam68 antibody to enable co-localization. CCAR1 and hnRNP L/Sam68 were distributed within the nucleoplasm and are nuclear. SNBs (Sam68 Nuclear Bodies) are shown with white arrows, labelled by hnRNP L stainintg. A) DAPI staining B) HnRNP L localization. Mouse anti hnRNP L (4D11) and anti-mouse-594(Red) antibodies were used as primary and secondary antibodies respectively. C) CCAR1 localization. Rabbit anti CCAR1 and anti-rabbit-488 (Green) were used as primary and secondary antibodies respectively. D) Merged picture of localization of both hmRNPL and CCAR1 proteins. These results show that although CCAR1 and hnRNP L do not co-localize in SNBs, they co-localize in the nucleoplasm. The scale bars are equal to 20µm.

3.3.2   Does endogenous Sam68 interact with both RHA and CCAR1 proteins?

Following confirmation of Sam68 protein interactions with RHA and CCAR1 after induction with tetracycline, I tested whether or not the interactions between these proteins happen in an endogenous situation as well. One third of a soluble HEK293 cell lysate was incubated overnight at 4°C with appropriate amounts of either Sam68, rabbit IgG (first control) or no antibody (second control). The next day, these soluble cell lysates were then incubated with protein A dynabeads for one hour at room temperature. This step was followed by washing the beads with 1X PBS. To check the efficiency of co-immunoprecipitation, 2X SDS sample loading dye was added to the samples for loading on SDS-PAGE, followed by Western blotting. Anti Sam68 (C-20) and anti-rabbit-HRP as primary and secondary antibodies (tables 15 and 16) were used respectively to probe the Western blots to confirm precipitation of Sam68 protein (figure 31-A).

The immunoprecipitated (IP) samples were then probed with either rabbit anti-DHX9 (RHA) or rabbit anti CCAR1 as primary and anti-rabbit IgG (HRP) as the secondary antibodies (tables 15 and 16) respectively. As shown in figure 31-B, RHA was only detected in the Sam68 IP lane and not in any of the control IP lanes. This experiment shows that endogenous RHA interacts with Sam68 at both endogenous and induced expression levels of Sam68.

On the other hand, in contrast to RHA, endogenous CCAR1 was not detected in the Sam68 immunoprecipitated (IP) lane. This means that the interaction between Sam68 and CCAR1 only happens after over-expression of Sam68 in cells, and there is not any detectable interaction between these two proteins when Sam68 is at endogenous level of expression (figure 31-C). This most likely means that the interaction between Sam68 and CCAR1 only occurs when Sam68 has a high level of expression and not in the normal cellular situation.

(A) Probing with anti Sam68 antibody
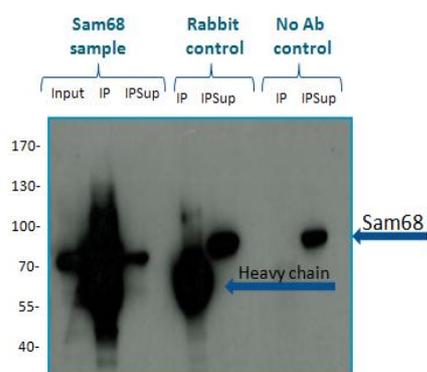
(B) Probing with anti DHX9 (RHA) antibody
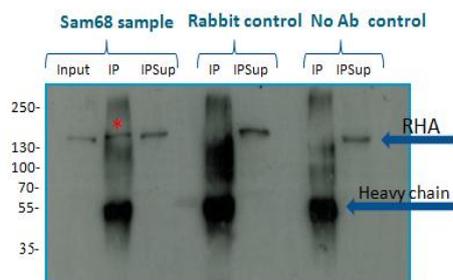
(C) Probing with anti CCAR1 antibody
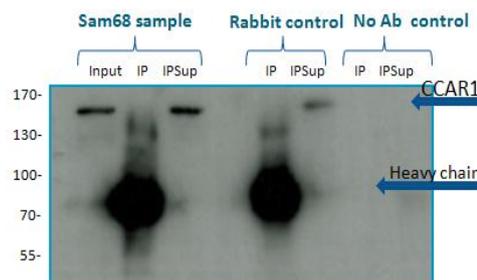
Figure 31: Testing for co-immunoprecipitation of RHA and CCAR1 with Sam68. One third of the soluble cell lysate was incubated overnight at 4°C with appropriate amounts of Sam68 antibody, rabbit IgG or no antibody. The next day, soluble cell lysates were incubated with protein A dynabeads for one hour at room temperature. 1X PBS was then used for washing the beads. The next steps were adding 2X SDS loading dye to samples, loading samples on SDS-PAGE, Western blotting and probing blot using appropriate antibodies. A) Proving the presence of Sam68 in the IP lane but not the control lanes to show the efficiency of immunoprecipitation. Anti Sam68 (C-20) and anti-rabbit-HRP as primary and secondary antibodies were used respectively. B) Proving the presence of RHA in the Sam68 IP samples. Rabbit anti-DHX9 (RHA) and anti-rabbit IgG (HRP) were used as primary and secondary antibodies to check the presence of RHA. The presence of RHA in the immunoprecipitated (IP) lane is shown by a red asterisk in the Sam68 IP sample. C) Testing presence of CCAR1 in the Sam68 IP sample using rabbit anti CCAR1 as primary and

anti-rabbit IgG (HRP) as secondary antibodies. CCAR1 was not detected in the Sam68 IP lane. The size markers are in kDa.

3.3.3 *Mapping interacting domains of Sam68 with RHA via preparation of stable cell lines expressing short forms of Sam68*

After proving the interaction of RHA and Sam68 at both endogenous and induced levels of expression of Sam68, the next step was to test which domain/domains of Sam68 might interact with RHA. To achieve this aim four short forms of Sam68 protein were made. The short forms had different domain content within their structures. These four truncated proteins were called N-KH (short form 1), N-KH-RG (short form 2), KH-RG (short form 3) and RG-tyrosine (short form 4) respectively.

Short form 1 (N-KH) contained the N-terminus, Qua1 and KH domains of Sam68 (figure 32-A). Short form2 in addition to these domains contained the Qua2 and RG-rich (arginine and glycine) domain (figure 32-B). Short form 3 (KH-RG) contained the KH, Qua2 and RG domains (figure 32-C). In the structure of short form 4 (RG-tyrosine), only the RG-rich domain and tyrosine-rich C-terminus were present (figure 32-D).

(A) Sam68 short form 1 (N-KH)

```
DNA: ATGCAGCGCCGGGACGATCCTGCCTCGCGCCTCACCCGGTCCTCGGGCCGC
+1:  M  Q  R  R  D  D  P  A  S  R  L  T  R  S  S  G  R
DNA: AGCTGCTCCAAGGACCCGTCAGGTGCCCACCCCTCGGTGCGTCTGACCCCG
+1:  S  C  S  K  D  P  S  G  A  H  P  S  V  R  L  T  P
DNA: TCTCGGCCGTCGCCCGCTTCCTCACCGGCCCCGGGGAGGGGGAGGTGGGCCC
+1:  S  R  P  S  P  L  P  H  R  P  R  G  G  G  G  P
DNA: AGAGGAGGCGCTCGGGCCTCGCCCGCCACCCAGCCGCCGCCGCTGCTGCCT
+1:  R  G  G  A  R  S  P  A  T  Q  P  P  P  L  L  P
DNA: CCCTCCACCCCTGGTCCCGACGCGACGGTGGTGGGTTCCGCGCCGACCCCG
+1:  P  S  T  P  G  P  D  A  T  V  V  G  S  A  P  T  P
DNA: CTGCTGCCCCCGTCAGCCACAGCCGCGGTCAAGATGGAGCCGGAGAATAAG
+1:  L  L  P  P  S  A  T  A  A  V  K  M  E  P  E  N  K
DNA: TACCTGCCTGAACTCATGGCCGAGAAGGACTCGCTCGACCCGTCCTTCACT
+1:  Y  L  P  E  L  M  A  E  K  D  S  L  D  P  S  F  T
DNA: CACGCCATGCAGCTGCTGTCCGTAGAAATTGAGAAGATTCAGAAGGGAGAG
+1:  H  A  M  Q  L  L  S  V  E  I  E  K  I  Q  K  G  E
DNA: TCAAAAAAAGATGACGAGGAGAATTATTTGGATTTATTTCTCATAAGAAC
+1:  S  K  K  D  D  E  E  N  Y  L  D  L  F  S  H  K  N
DNA: ATGAAGCTGAAAGAACGCGTGCTGATACCTGTCAAGCAGTATCCAAAGTTC
+1:  M  K  L  K  E  R  V  L  I  P  V  K  Q  Y  P  K  F
DNA: AATTTTGTGGGGAAGATTCTTGGACCACAAGGAAATACAATCAAAAGACTC
+1:  N  F  V  G  K  I  L  G  P  Q  G  N  T  I  K  R  L
DNA: CAGGAAGAGACTGGTGCAAAGATCTCTGTCTTGGGGAAGGGTTCAATGAGA
+1:  Q  E  E  T  G  A  K  I  S  V  L  G  K  G  S  M  R
DNA: GACAAAGCCAAGGAGGAA
+1:  D  K  A  K  E  E
```

(B) Sam68 short form 2 (N-KH-RG)

```
DNA: ATGCAGCGCCGGGACGATCCTGCCTCGCGCCTCACCCGGTCCTCGGGCCGC
+1:  M  Q  R  R  D  D  P  A  S  R  L  T  R  S  S  G  R
DNA: AGCTGCTCCAAGGACCCGTCAGGTGCCCACCCCTCGGTGCGTCTGACCCCG
+1:  S  C  S  K  D  P  S  G  A  H  P  S  V  R  L  T  P
DNA: TCTCGGCCGTCGCCGCTTCCTCACCGGCCCCGGGGAGGGGGAGGTGGGCCC
+1:  S  R  P  S  P  L  P  H  R  P  R  G  G  G  G  G  P
DNA: AGAGGAGGCGCTCGGGCCTCGCCCGCCACCCAGCCGCCGCCGCTGCTGCCT
+1:  R  G  G  A  R  A  S  P  A  T  Q  P  P  P  L  L  P
DNA: CCCTCCACCCCTGGTCCCGACGCGACGGTGGTGGGTTCCGCGCCGACCCCG
+1:  P  S  T  P  G  P  D  A  T  V  V  G  S  A  P  T  P
DNA: CTGCTGCCCCCGTCAGCCACAGCCGCGGTCAAGATGGAGCCGGAGAATAAG
+1:  L  L  P  P  S  A  T  A  A  V  K  M  E  P  E  N  K
DNA: TACCTGCCTGAACTCATGGCCGAGAAGGACTCGCTCGACCCGTCCTTCACT
+1:  Y  L  P  E  L  M  A  E  K  D  S  L  D  P  S  F  T
DNA: CACGCCATGCAGCTGCTGTCCGTAGAAATTGAGAAGATTCAGAAGGGAGAG
+1:  H  A  M  Q  L  L  S  V  E  I  E  K  I  Q  K  G  E
DNA: TCAAAAAAAGATGACGAGGAGAATTATTTGGATTTATTTTCTCATAAGAAC
+1:  S  K  K  D  D  E  E  N  Y  L  D  L  F  S  H  K  N
DNA: ATGAAGCTGAAAGAACGCGTGCTGATACCTGTCAAGCAGTATCCAAAGTTC
+1:  M  K  L  K  E  R  V  L  I  P  V  K  Q  Y  P  K  F
DNA: AATTTTGTGGGGAAGATTCTTGGACCACAAGGAAATACAATCAAAAGACTC
+1:  N  F  V  G  K  I  L  G  P  Q  G  N  T  I  K  R  L
DNA: CAGGAAGAGACTGGTGCAAAGATCTCTGTCTTGGGGAAGGGTTCAATGAGA
+1:  Q  E  E  T  G  A  K  I  S  V  L  G  K  G  S  M  R
DNA: GACAAAGCCAAGGAGGAAGAGTTGCGCAAGGGTGGAGACCCCAAATATGCC
+1:  D  K  A  K  E  E  E  L  R  K  G  G  D  P  K  Y  A
DNA: CATTTAAATATGGATCTGCATGTCTTCATTGAAGTCTTTGGACCCCCGTGT
+1:  H  L  N  M  D  L  H  V  F  I  E  V  F  G  P  P  C
DNA: GAAGCTTATGCTCTTATGGCCCATGCTATGGAAGAAGTCAAGAAGTTCCTA
+1:  E  A  Y  A  L  M  A  H  A  M  E  E  V  K  K  F  L
DNA: GTACCAGATATGATGGATGATATCTGTCAGGAGCAGTTTCTAGAATTGTCC
+1:  V  P  D  M  M  D  D  I  C  Q  E  Q  F  L  E  L  S
DNA: TACTTGAACGGAGTACCTGAACCCTCTCGTGGTCGTGGGGTATCTGTGAGA
+1:  Y  L  N  G  V  P  E  P  S  R  G  R  G  V  S  V  R
DNA: GGACGAGGAGCTGCCCCTCCTCCTCCACCTGTTCCCAGAGGACGTGGTGTT
+1:  G  R  G  A  A  P  P  P  P  P  V  P  R  G  R  G  V
DNA: GGACCACCTAGAGGAGCTTTGGTTCGTGGAACCCCAGTGAGAGGCTCCATC
+1:  G  P  P  R  G  A  L  V  R  G  T  P  V  R  G  S  I
DNA: ACCAGAGGTGCCACTGTGACTCGAGGAGTGCCACCCCCACCTACTGTGAGG
+1:  T  R  G  A  T  V  T  R  G  V  P  P  P  T  V  R
DNA: GGTGCTCCAACACCAAGAGCTCGGACA
+1:  G  A  P  T  P  R  A  R  T
```

Figure 32 : Four different short forms of Sam68. Four different short forms of Sam68 were made to be cloned into the pcDNA5 vector. Different domains are shown in different colours. Dark green: extended N-terminal domain of Sam68; green: proline rich sequences (SH3 binding sits); red: N-terminus; bold and underlined: Qua1 domain; bold, blue and underlined: KH Domain; blue font color with under line: Qua2 domain; yellow: RG-rich domain; purple: tyrosine-rich C-terminal domain; gray: Tyrosine residues (candidate SH2 binding sites). The picture is modified from (Ehrmann and Elliott, 2010).

3.3.3(a) Short forms of Sam68 cloned in pcDNA5

Forward and reverse primers (table 14) were used to PCR amplify the open reading frames encoding different Sam68 short forms, using Sam68 cloned into a pACT2 AD construct as a template. The cloning strategy of these short forms of Sam68 was similar to the cloning strategy of wild type Sam68 (explained previously, Chapter 2, result section 2-3-2). Briefly, PCR amplification resulted in different sizes of PCR products. Short form 1 (N-KH): 630bp; short form 2 (N-KH-RG): 1047bp; short form 3 (KH-RG): 549bp; and short form 4 (RG-tyrosine rich): 501bp (figure 33).

Figure 33: PCR amplification of short forms of Sam68 for cloning into the pcDNA5 vector. Agarose gel electrophoresis of PCR reaction products showing: (1) Short form 1 (N-KH). MSam68 F-BamHI and WSam68 R-N-KH-SalI were used as the forward and reverse primers respectively. The PCR product size was 630bp. (2) short form 2 (N-KH-RG). MSam68 F-BamHI and WSam68 R-N-KH-RG-SalI were used as the forward and reverse primers respectively. The PCR product size was 1047bp. (3) Short form 3 (KH-RG). WSam68 F-KH-RG-BamHI and R-N-KH-RG-SalI were used as the forward and reverse primers respectively. The PCR product size was 540bp. 4) Short form 4 (KH-tyrosine). WSam68 F-RG-Tyrosine BamHI and Wt MSam68 R-SalI were used as the forward and reverse primers respectively. The PCR product size was 501bp. The 1kb plus DNA ladder was run in parallel and is also shown in this picture.

Chapter 3: Further investigation into the interactions of T-STAR, wild type Sam68 and P439R Sam68

Digestion of the short forms of Sam68 (insert) and the pcDNA5 vector was carried out by restriction enzymes (*BamH*I, *Sal*I) and (*BamH*I, *Xho*I) respectively. This step was followed by ligation of the digested PCR products into the digested sites of pcDNA5. *Not*I was used for the re-cleavage process to remove the ligated vectors which did not contain an insert. 10 μl of plasmid DNA was then used for heat shock transformation into *E. coli*. LB agar plates which contained 50μg/ml ampicillin were used for plating out the cells. Plates were incubated at 37°C overnight to select colonies. Insert forward primers including: MSam68 F-BamHI (for short form 1 and 2), WSam68 F-KH-RG-BamHI (for short form 3), WSam68 F-RG-Tyrosine BamHI (for short form 4) and vector reverse primer, BGHrev were used to screen the colonies by PCR. Colonies which contained the insert were grown in the LB medium plus 50μg/ml ampicillin. After overnight incubation at 37°C on a shaker, plasmid purification was carried out.

To confirm the correct presence of inserts in the pcDNA5 vector, 100ng/μl of constructs were sequenced. The results confirmed the presence of inserts containing the open reading frames of short forms 1 (N-KH), 2 (N-KH-RG), 3 (KH-RG) and 4 (RG-tyrosine) in the constructs. The sequencing results are shown in the Appendix A.

3.3.3(b) Establishment of stable cell lines encoding short forms of FLAG-tagged Sam68

Dr Andrew Knox and Dr Nicholas Watkins (Institute for Cell and Molecular Biosciences, Newcastle University) gave us the pcDNA5 vector and Flp-In HEK293 cell. In Flp-In HEK293 cells there was tet repressor expression. Tetracycline promoter in these cells which could control gene cloned downstream of this promoter could control the expression of gene of interest. Therefore, addition of tetracycline antibiotic to the medium could lead to high level of expression of short forms of Sam68. When cells were 60%-70% confluent, the open reading frame of the short forms of Sam68 (short

form1: N-KH; short form2: N-KH-RG; short form3: KH-RG; and short form4: RG-tyrosine rich) cloned in pcDNA5 were co-transfected with pOG44 plasmid. Flp recombinase was encoded by pOG44, for integration of each gene of interest into the FRT site of the genome of these cells. Hygromycin B was used for selection of the transfected cells. Transfected cells were resistant to the hygromycin antibiotic. Therefore, after few weeks of washing dead cells from the medium and addition of antibiotic, stably transfected HEK 293 cells were growing. Single colonies were transfered to new plates. After establishment of the cells in a new plate they were ready to be induced by tetracycline to express the FLAG tagged short forms of Sam68.

3.3.3(c) Testing gene expression of short forms of Sam68 after induction

Before mapping the interacting domains of Sam68 with RHA using immunoprecipitation, the expression of these short forms of Sam68 was tested after 72 hours induction with tetracycline. When stable cell lines expressing different short forms of Sam68 were confluent enough, expression of the genes of interest were induced by adding tetracycline to the cell culture medium. After 3 days cells were harvested and sonicated. Samples were loaded on a 13% running gel, electrophoresed and analysed by Western blotting. Predicted sizes of these short forms proteins were 30.38kDa (short form 1: N-KH); 37.60kDa (short form2: N-KH-RG); 19.22kDa (short form3: KH-RG); and 18.02kDa (short form4: RG-tyrosine).

Wild type Sam68 was also loaded as the positive control. Figure 34 shows the resulting Western, in which bands for wild type, short forms1 and 2 could be detected. Since the stable cell line expressing Sam68 short forms 3 and 4 were resistant to antibiotic, the absence of Sam68 short form 3 and 4 proteins on the gel could be because of instability of these proteins in these cells. This instability could be because of either their small sizes being unable to fold properly or being not functional proteins.
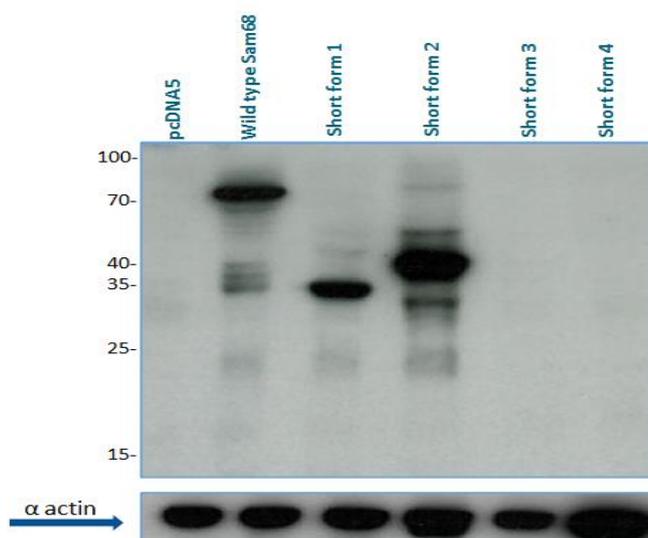
Figure 34: Induced expression of FLAG tagged short forms of Sam68 in their respective stable cell lines as monitored by Western blotting. The presence of FLAG tagged protein (wild type, short forms1 and 2) expressions were confirmed after 72 hours induction in comparison with negative control (pcDNA5). Because there are FLAG tags and His tags in these recombinant proteins, their weights were higher than expected. The predicted weight of wild type Sam68 and short forms 1 and 2 are 68kDa, 30.38kDa and 37.60kDa respectively. It seems that short forms 3 and 4 were not stable and could be degraded quickly after their production. Monoclonal anti FLAG and anti-mouse-HRP were used as the primary and secondary antibodies respectively to detect the presence of FLAG tagged proteins. Blots were re-probed with α-β actin and α-rabbit IgG (HRP) as the primary and secondary antibodies respectively to confirm that the samples had been loaded equally. Equal levels of actin loading control were present in each lane. Size markers are in kDa.

3.3.3(d) Purifying protein partners of short forms of FLAG tagged Sam68 using immunoprecipitation

One of the aims in this project was to detect which part/parts of Sam68 might be involved in mediating the interaction with RHA. Since just two short forms of Sam68 were stable, it was possible to continue this analysis only with short forms 1 and 2. While short form 1 had RG and tyrosine domain deletions, in the structure of short forms 2 only the tyrosine domain was deleted. The plan was to immunoprecipitate Sam68 short forms 1 and 2 to test whether they interact with RHA or not.

Therefore, stable Flp-In HEK293 cells expressing Sam68 short forms 1 and 2 were induced by tetracycline for three days. Cells were then harvested, followed by the immunoprecipitation of FLAG-tagged proteins. α-FLAG affinity resin was used for immunoprecipitation of these proteins.

After harvesting cells, 1.5ml of the IP lysis buffer was added to samples followed by incubation of the cells on ice. FLAG resin washed with TBS buffer was added to the cell pellets and samples were centrifugated at 4°C. 30μl of the supernatant was kept as the input. After on overnight incubation of the rest of the supernatant with FLAG resin, samples were centrifuged and 30μl of the supernatant was kept as the IP sup. Then the FLAG resin was washed with the IP washing buffer. To check immunoprecipitation efficiency, after adding 2XSDS sample loading dye, samples were loaded on SDS-PAGE and probed with appropriate antibodies.

I tested firstly the presence of FLAG tagged proteins in Immunoprecipitated (IP) samples using Western blotting to analyse whether the immunoprecipitation had been successful. Monoclonal anti FLAG antibody was used as the primary antibody to probe the Western blots. IP lanes of FLAG tagged proteins in comparison with the control

samples (stable cell lines made with pcDNA5) showed that the immunoprecipitations were successful in retrieving the tagged proteins (figure 35).
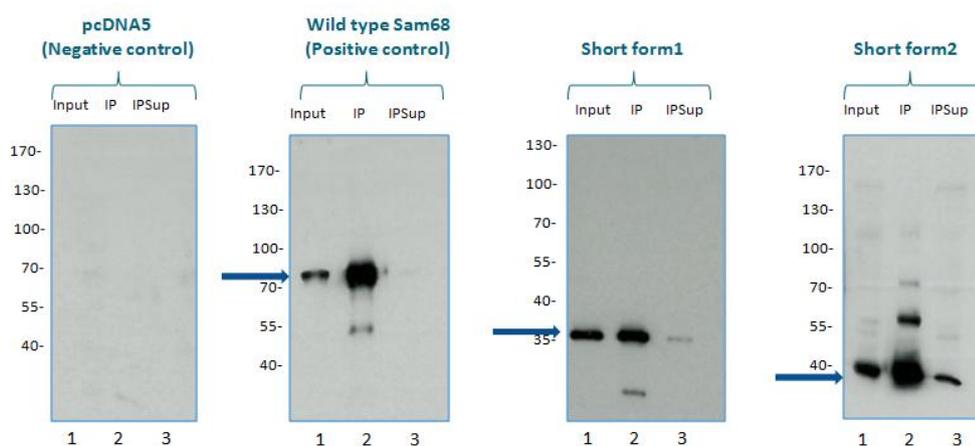
Figure 35: Testing immunoprecipitation (IP) of FLAG tagged proteins from stable cell lines. Western blotting was carried out to determine IP efficiency. Monoclonal anti FLAG antibody and HRP-conjugated anti-mouse were used as the primary and secondary antibodies. In the above picture 1, 2 and 3 show the input (soluble supernatants before immunoprecipitation), IP (immunoprecipitated sample) and IP Sup (soluble supernatants after immunoprecipitation) respectively. IP lanes of the FLAG tagged proteins in comparison with the negative control sample (pcDNA5) showed that co-immunoprecipitations were successful. The size markers are in kDa. The position of the FLAG tagged proteins on the Western blots are indicated by arrows.

3.3.3(e) Which domain/domains of Sam68 are involved in interaction with RHA?

After confirming that the experiment was recovering the Flag tagged proteins, the next step was to probe the blots with anti RHA antibody. The aim was to detect which domain/domains of Sam68 is involved in interaction with RHA protein. In the short form 1, just the tyrosine-rich C-terminal domain was deleted. This means that in the structure of this protein all parts of Sam68 except tyrosine rich domain were present. On the other hand, in short form 2 in addition to the tyrosine domain, the RG (arginine-glycine rich) domain was deleted as well. Western blots were probed with anti-DHX9 (RHA), and anti-rabbit IgG (HRP) were used as secondary antibody. RHA was present in the immunoprecipitated (IP) lane of the positive control (full length Sam68-Flag protein). RHA protein was also present in the Input and IPsup lanes of all other samples. However, there was no RHA band in the immunoprecipitated (IP) lanes of Sam68 short forms 1 and 2, meaning that neither of these two shorter Sam68 protein isoforms interacts with RNA helicase A. Since in the short form 2 all domains except tyrosine rich domain were present, it seems that this tyrosine rich domain has an important role in the interaction between Sam68 and RHA (figure 36).
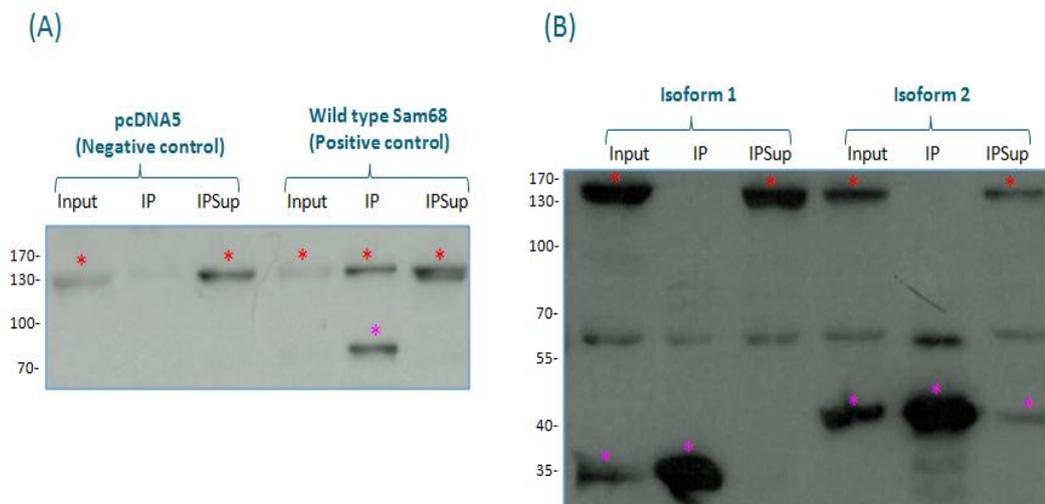
Figure 36: Testing the presence of RHA in immunoprecipitated (IP) samples of Sam68 short forms 1 and 2. (A) Western blot showing the presence of RHA protein in the IP lane of the positive control (wild type Sam68). After confirming the presence of FLAG tagged Sam68 (pink asterisk), Western blots were re-probed with anti RHA antibody. RHA was present in the Input and IPsup lanes of both positive and negative controls and only in the IP lane of the positive control (shown by red asterisks). (B) This picture shows that RHA was not present in the IP lanes of Sam68 short forms1 and 2. After confirming the presence of FLAG tagged short forms 1 and 2 (pink asterisks), Western blots were re-probed with RHA antibody. Although RHA was present in Input and IPsup samples of Sam68 short forms 1 and 2 (shown by red asterisks), RHA protein was not present in any of the IP lanes. The size marker is in kDa. Anti-DHX9 (RHA) and anti-rabbit IgG (HRP) were used as primary and secondary antibodies respectively.

3.3.4   Functional analysis of whether Sam68 and RHA cooperate in the cell

To find out more about the links between Sam68 and RHA, firstly the open reading frame of RHA was cloned in the pcDNA5 vector. Then a stable cell line expressing FLAG tagged RHA protein was made. The presence of induced RHA was confirmed after 72

hours of induction by tetracycline. The aim was to investigate more about the possible roles of RHA and Sam68 in transcription and alternative splicing regulation. Therefore, the transcription level of *CyclinD1* and alternative splicing regulation (exon inclusion) of *NRXN3* after induction of Sam68 and RHA with tetracycline were tested to answer some of the questions regarding any potential cooperation between these two proteins.

3.3.4(a) Cloning the open reading frame of RHA in pcDNA5 vector

Forward and reverse primers (table 14) were used to PCR amplify the open reading frame of RHA from a pcDNA3-HA (GAT) construct (gifted from Professor Nakajima, Institute of Medical Science, Tokyo Medical University). WRHA-F-BamHI and WRHA-R-SalI were used as forward and reverse primers. The cloning strategy of RHA was similar to cloning strategy of wild type Sam68 (explained previously, Chapter 2, result section 2-3-2). The PCR amplification resulted in a 3818bp size of PCR product (figure 37).
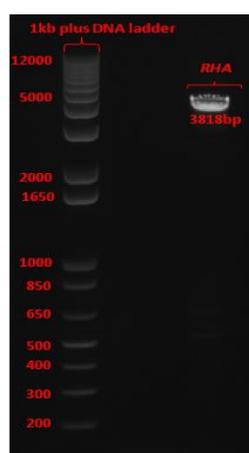
Figure 37: PCR amplification of *RHA* for cloning into pcDNA5 vector. WRHA-F-BamHI and WRHA-R-SalI were used as forward and reverse primers respectively. The PCR product size was 3818bp. The migration of the 1kb plus DNA ladder is also shown in this picture.

Digestion of RHA (insert) and the pcDNA5 vector was carried out by using restriction enzymes (*BamH*I, *Sal*I) and (*BamH*I, *Xho*I) respectively. This step was followed by ligation of the digested PCR products into the complementary digested sites of pcDNA5. *Not*I was used for the re-cleavage process to remove the ligated vectors which did not contain an insert. Transformation of the ligation mixes into *E.coli*, plating out, and screening the colonies by PCR, as well as plasmid purification were carried out similarly to what had been done for the short forms of Sam68.

To confirm the correct presence of RHA in the pcDNA5 vector, 100ng/µl of the constructs were sequenced. The results confirmed the presence of inserts containing the open reading frames of RNA helicase A in the constructs. These sequencing results are shown in the Appendix A.

3.3.4(b) Establishment of stable cell line encoding FLAG-tagged RHA

When Flp-In HEK293 cells were 60%-70% confluent, the open reading frame of RHA cloned in the pcDNA5 was co-transfected with pOG44 plasmid. Hygromycin B was used for selection of the transfected cells. Transfected cells were resistant to the hygromycin antibiotic. Therefore, after few weeks of washing dead cells from the medium and addition of new antibiotic, stably transfected HEK 293 cells were forming colonies. Single colonies were transferred to the new plates. After establishment of cells on a new plate they were ready to be induced by tetracycline to express the FLAG tagged RHA.

3.3.4(c) RHA makes FLAG-tagged proteins of the appropriate size after induction

When stable cell lines expressing RHA were confluent enough, tetracycline was added to the medium to induce expression of FLAG tagged RHA. After 3 days induction, cells were harvested and sonicated. The samples were loaded on an 8% running gel. The

predicted size of this tagged RHA protein was 140kDa plus the mass of the tag. A protein of exactly this size was visible on the Western (figure 38).
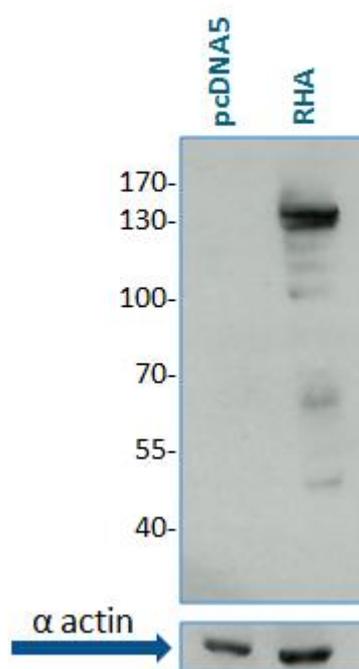
Figure 38: Induced expression of FLAG tagged RHA in the stable cell lines as monitored by Western blotting. The presence of FLAG tagged RHA expression was confirmed after 72 hours induction in comparison with the negative control (Flp-In HEK293 cells stably transfected with pcDNA5, empty vector). Monoclonal anti FLAG and anti-mouse-HRP were used as the primary and secondary antibodies respectively to detect the presence of FLAG tagged proteins. Blots were re-probed with α-β actin and α-rabbit IgG (HRP) as the primary and secondary antibodies respectively, to confirm that the sample had been loaded equally. Equal levels of actin loading control were present in each lane. The size markers are in kDa.

3.3.4(d) Experiment to test the role of Sam68 and RHA in regulation of cyclin D1/CCND1 transcript level

To find out more about the roles of Sam68 and RHA in transcription regulation and also to investigate more about any possible links between these two proteins, their

roles in transcription regulation of a target gene (*cyclin D1/CCND1)* was tested. *Cyclin D1* has an important role in regulation of the cell cycle, and is involved in several kinds of cancers including breast cancer (Xu et al., 2013). In prostate and breast tumours Sam68 is also over expressed (Elliott and Rajan; Busa *et al.*, 2007; Rajan *et al.*, 2008b; Elliott and Rajan, 2010), and it seems that in breast cancer cells there is positive correlation between the expression level of RHA and cyclin D1 (Huo et al., 2010). Therefore, it was interesting to test whether the level of *cyclin D1/CCND1* mRNA changed after over expression of Sam68 and RHA in Flp-In HEK293 cells.

After 3 days of induction with tetracycline, induced cells were harvested. This step was followed by RNA extraction and preparation of cDNA from RNA (500ng/µl). cDNA samples of stable cell lines made with pcDNA5 (empty vector), wild type Sam68, P439R mutant Sam68 and RHA were then diluted 1:10 in dH_2O. Real time PCR was then carried out in triplicate to estimate the level of *cyclin D1/CCND1* mRNA in each sample.

The ΔCt method was used to estimate the relative expression level of *cyclin D1/CCND1* mRNA. After calculating the average of Ct value of samples, the mean Ct value of endogenous expression of *α Actin, βTubulin* and *GAPDH* (control genes) was used for normalization of the *cyclin D1/CCND1* Ct value. Figure 39 shows the level of *cyclinD1* mRNA in wild type Sam68, P439R mutant Sam68 and RHA after induction with tetracycline in comparison with pcDNA5 (empty vector, control group). Data were analysed using Excel software (Student's *t*-test). While there was no significant difference between wild type and control groups ($P > 0.05$), interestingly the level of *cyclin D1/CCND1* mRNA was significantly higher in the P439R mutant Sam68 group in comparison with the control group (pcDNA5, empty vector) ($P < 0.05$). Discussed in Chapter 2, P439R mutant Sam68 is cytoplasmic in localisation. However, the level of *cyclin D1/CCND1* mRNA in a stable cell line expressing RHA after addition of tetracycline was not significantly different from the control group (pcDNA5) ($P > 0.05$).
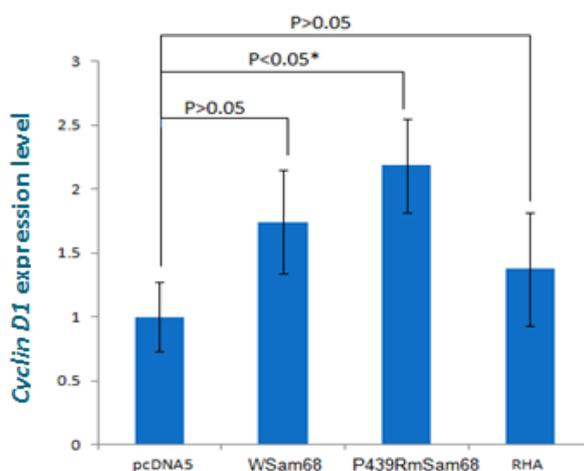
Figure 39: *Cyclin D1/CCND1* **expression level after 3 days induction with tetracycline.** After 3 days of induction with tetracycline antibiotic, the level of *cyclin D1/CCND1* mRNA was estimated in stable cell lines expressing pcDNA5 (empty vector, control group), WSam68 (wild type Sam68), P439RmSam68 (mutant Sam68), and RHA (RNA helicase A) using real time PCR. There was no significant difference between the control group (pcDNA5) and WSam68 (wild type Sam68) (P>0.05). Interestingly, I could detect significant differences between the control group (pcDNA5) and P439RmSam68 (mutant Sam68) (P<0.05). In addition, the level *of cyclin D1/CCND1* expression in the cells with induced expression of RHA was not significantly different from control group (P>0.05). All samples were tested in triplicate. The asterisk shows that the difference between two groups was significant.

3.3.4(e) Do Sam68 and RHA co-regulate the same target genes at the splicing level?

To investigate more about the roles of Sam68 and RHA, role of these proteins was tested in alternative splicing regulation of a target gene called *Neurexin3* (*Nrxn3*). Using alternative promoters and the presence of alternatively spliced exons within the mRNAs of *Neurexin1-3* can lead to production of many different proteins (Missler and Sudhof, 1998; Craig and Kang, 2007; Loya *et al.*, 2010). It has been shown that Sam68 has role in alternative splicing of *Neurexin1* in the cerebellum (Iijima et al., 2011). In

113

addition, T-STAR and Sam68 are involved in regulation of alternative splicing of *Neurexin3* (Ehrmann et al., 2013) . The question was whether Sam68 and RHA had roles in splicing regulation of alternatively spliced segment 4 (AS4) into *Neurexin3* mRNAs in human cells (Flp-In HEK293 cells). Therefore, primers were designed for this cassette exon (figure 40). As shown in this figure there are alternative promoters, two alternatively spliced segments (cassette exons), an alternative five prime splice site and also two alternative finishes. These alternative events mean that from the same gene, many different proteins can be produced. The question was whether or not Sam68 and RHA have roles in AS4 exon splicing.
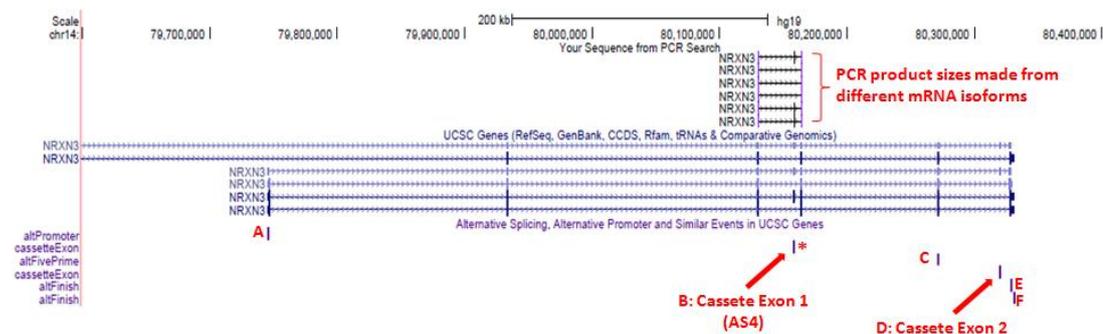
Figure 40: *NRXN3* can make different types of mRNA using different alternative promoters, cassette exons and alternative 3'ends. *NRXN3* has an alternative promoter (A), 2 cassette exons (B and D), an alternative 5' splice site (C) and two alternative 3' ends (E and F). Roles of Sam68 and RHA were tested in inclusion of the AS4 exon (shown by red asterisk). Here a screenshot from UCSC shows the sizes of the expected PCR products with or without this cassette exon.

After 3 days of induction with tetracycline, induced Flp-In HEK293 cells containing pcDNA5 (empty vector, negative control), wild type Sam68, P439R mutant Sam68 and RHA were harvested. This step was followed by RNA extraction, cDNA synthesis from RNA (500ng/µl) and diluting the cDNA (at a ratio of 1:5). Then 25 µl polymerase chain reactions were carried out. GoTaq® enzyme, NRXN3 F and NRXN3 R were used as DNA polymerase, forward and reverse primers respectively. Both the QIAXCEL multi-

capillary electrophoresis system and agarose gel electrophoresis were used to analyse PCR products. While wild type Sam68 was confirmed to have a role in repression of exon inclusion of cassette exon 1 (AS4 exon) (there was a decrease in inclusion of the AS4 exon, so the upper exon included band was weaker after Sam68 induction), the P439R mutant Sam68 and RHA did not play any role in splicing regulation of this target (figure 41). Since P439R mutant Sam68 is cytoplasmic, this protein is not present in the nucleus to be involved in regulation of splicing of this cassette exon. Furthermore, it seems that RHA also does not play any role in splicing regulation of this target gene (*NRXN3*).
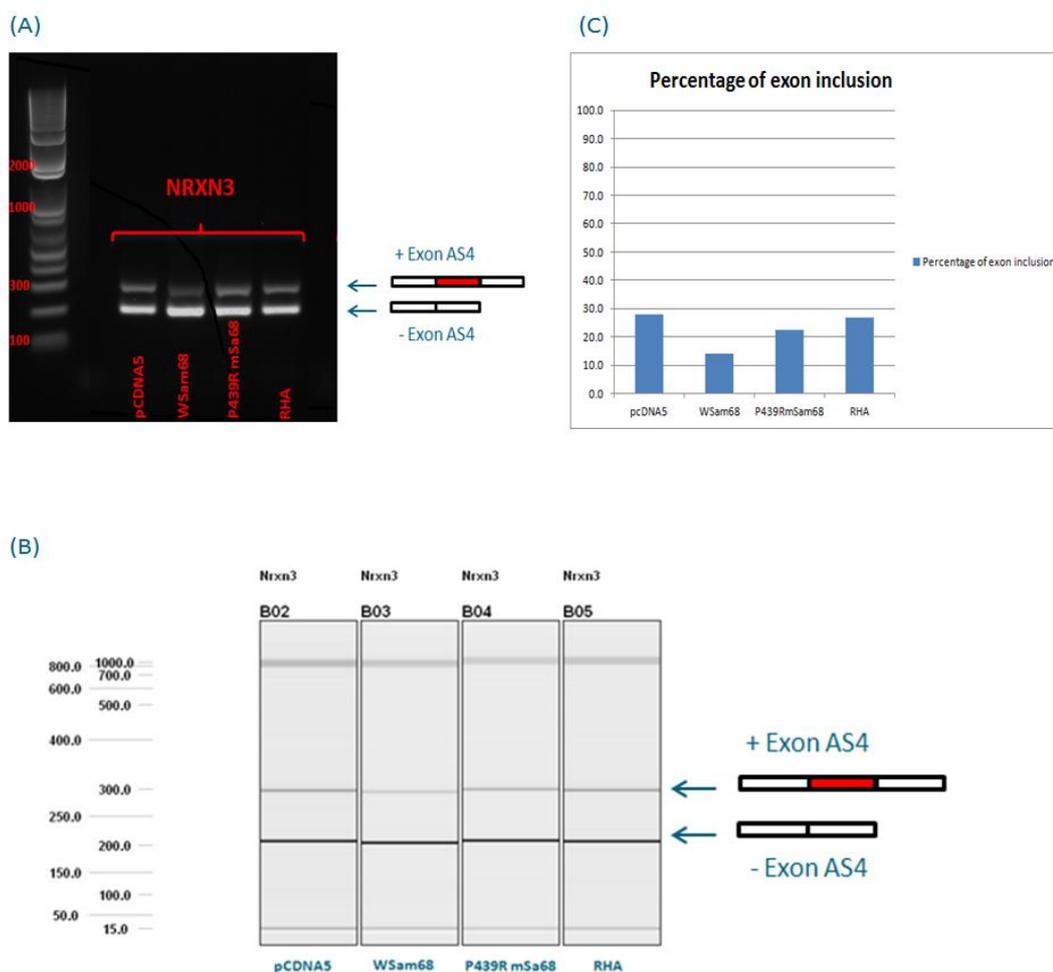
Figure 41 : Testing the role of Sam68 and RHA in splicing regulation. To check the possible role of these two proteins in splicing regulation of the *NRXN3* cassette exon 1 (AS4 exon), HEK293 cells were harvested which over-expressed these proteins and RNA samples were extracted. After making cDNA, RT-PCR was used to check for the presence of this cassette exon in mRNA. A) PCR samples loaded on agarose gel. RT-PCR products made with and without this specific exon would be 294 and 209 bp. The 1kb plus DNA ladder also has been shown in this picture. B) The QIAXCEL multi-capillary electrophoresis system was also used to check the exon inclusion in each sample. C) This graph shows the percentage of exon inclusion in each sample. pcDNA5: 27.8%, WSam68 (wild type Sam68): 14.2%, P439RmSam68 (mutant Sam68): 22.4%, RHA: 26.7%. Together these data show while wild type Sam68 has role in repression of exon inclusion of AS4 exon, p439R Sam68 and RHA do not play any detectable role.

3.3.5 Testing the interaction between endogenous T-STAR and RHA proteins

Following the confirmation of a protein interaction between T-STAR and RHA using FLAG-tagged protein bait after induction with tetracycline, there was still another question to be answered: whether or not the interaction between these two proteins happens in an endogenous situation. Since T-STAR is expressed only in brain, testis and very small amount in kidney (Venables *et al.*, 1999; Venables *et al.*, 2004), HEK293 cell line (derived from Human Embryonic Kidney cells) was chosen to check the interaction between T-STAR and RHA.

Immunoprecipitation of endogenous T-STAR protein using protein A Dynabeads was carried out to test this hypothesis. One third volume of a soluble HEK293 cell lysate was incubated overnight at 4°C with appropriate amounts of either T-STAR, rabbit IgG (first control) or no antibody (second control). The next day, soluble cell lysates were incubated with protein A dynabeads for one hour at room temperature. 1X PBS was then used for washing the beads. To check the efficiency of co-immunoprecipitation, 2X SDS sample loading dye was added to the samples before loading on SDS-PAGE and followed by Western blotting.

Since the heavy chains of both T-STAR and the rabbit IgG used for immunoprecipitation have similar weights, it was not possible to directly check the efficiency of T-STAR immunoprecipitation. Therefore, for testing this efficiency, investigation of the presence of one of already known interacting partner of T-STAR in the IP samples could be used for proving the efficiency of T-STAR immunoprecipitation.

Since T-STAR has a known protein interaction with Sam68 (Venables et al., 1999), Sam68 rabbit polyclonal IgG and anti-rabbit IgG (HRP) were used as primary and secondary antibodies respectively to look for Sam68 protein in the immunoprecipitates. A very small amount of Sam68 was co-immunoprecipitated with

T-STAR (figure 42) which showed that probably the small amount of T-STAR expression in HEK293 cells is a barrier for highly efficient co-immunoprecipitation.

To test for co-immunoprecipitation of RHA, the IP samples were probed with rabbit anti-DHX9 (RHA) and anti-rabbit IgG (HRP) as primary and secondary antibodies respectively. Although RHA was not detected in the T-STAR IP lane, this may be because of negligible expression of T-STAR in kidney cells and not because of a lack of any interaction of T-STAR with RHA (figure 43). Therefore, to check the endogenous interaction between T-STAR and RHA it would probably be better to use other cell lines or tissues which express higher amounts of T-STAR such as testis tissue in future experiments.
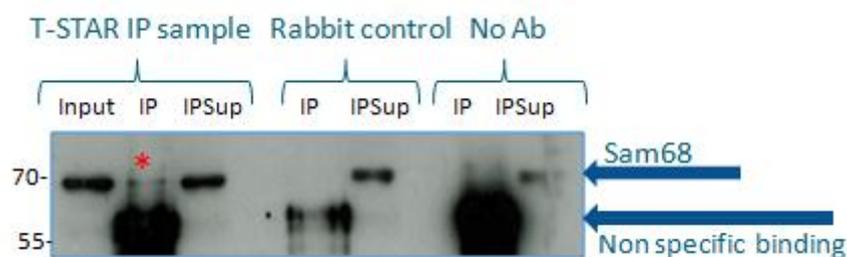
Figure 42: Experiment to test the presence of Sam68 protein in the T-STAR IP sample. One third of a soluble HEK293 cell lysate was incubated overnight at 4°C with appropriate amounts of either T-STAR antibody, rabbit IgG or no antibody. The next day, soluble cell lysates were incubated with protein A dynabeads for one hour at room temperature. 1X PBS was then used for washing the beads. After washing 2X SDS loading dye was added to samples, followed by SDS-PAGE, Western blotting and probing blots with appropriate antibodies. Presence of Sam68 is shown by red asterisk. The size marker is in kDa. Pageruler prestained protein ladder was used as the marker which has range from 10kDa to 250 kDa.
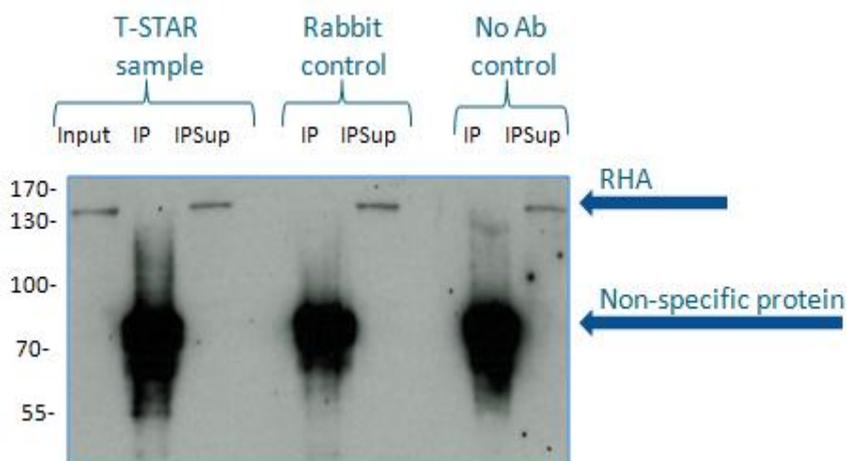
Figure 43: Testing the presence of RHA protein in the T-STAR cell lysate IP samples. IP samples were probed with Rabbit anti-DHX9 (RHA) and anti-rabbit IgG (HRP) as primary and secondary antibodies respectively. The size markers shown are in kDa.

Chapter 3: Further investigation into the interactions of T-STAR, wild type Sam68 and P439R Sam68

3.4   Summary and Discussion:

In this Chapter the aim was to investigate more about T-STAR, Sam68 and their newly identified interacting candidates.

To compare the distribution of Sam68 relative to RHA and CCAR1 the localization of these proteins was monitored within the cell. Specifically, to see whether or not these proteins co-localized with Sam68. Appropriate antibodies were used to check localization of RHA, CCAR1 and Sam68 in HeLa cells. Rajan et al in 2009 had shown that Sam68 and hnRNP L are present within SNB (Sam68 Nuclear Bodies) and the nucleoplasm and co-localized with each other (Rajan et al., 2009). Therefore, monoclonal mouse hnRNP L antibody was used instead of Sam68 antibody. Consistent to the findings of Denegri et al (Denegri *et al.*, 2001) in this study SNBs were also detected in HeLa cells. The results showed RHA or CCAR1 were not present in the SNBs, but endogenous RHA, CCAR1 and Sam68 were localized within the nucleoplasm. Since Sam68 was present both in the SNBs and nucleoplasm it seems that Sam68 interactions with RHA and CCAR1 occur in the nucleoplasm instead of SNBs.

To investigate more about CCAR1 and RHA, endogenous interaction of Sam68 was tested with these proteins. A Sam68 co-IP using protein A Dynabeads was carried out, and the IP samples were probed with appropriate antibodies. These results suggest that CCAR1 interacts with Sam68 only after Sam68 over-expression and not in the endogenous situation. Based on these results while normal level of Sam68 does not stimulate CCAR1 interaction, an increased level of Sam68 expression can lead to interaction with CCAR1 protein. Other studies have shown the role of Sam68 and CCAR1 proteins separately in transcription and cancer. Sam68 and CCAR1 are both over-expressed in breast cancer (Busa *et al.*, 2007; Kim *et al.*, 2008; Rajan *et al.*, 2008b; Elliott and Rajan, 2010). Therefore, there could be a possible link between expression level of Sam68 and CCAR1 which needs to be tested.

Chapter 3: Further investigation into the interactions of T-STAR, wild type Sam68 and P439R Sam68

The interaction between RHA and Sam68 can occur at both endogenous and over expressed levels of Sam68. This means that the interaction between these two proteins is not dependent on expression level of Sam68. Based on these results the next step was testing the possible links between RHA and Sam68 proteins.

To answer which domain/domains of Sam68 is/are interacting with RHA, short forms of Sam68 (short forms 1, 2, 3 and 4) were produced by expression in stable cell lines. The drawback of this approach was out of four truncated proteins only two short forms had detectable expression level. This could be because short forms 3 and 4 were not stable in these cells and could possibly be degraded quickly after their production. This suggests these proteins were degraded because of either their small sizes or being not functional proteins within the cells.

Immunoprecipitation was carried out to identify protein partners of Sam68 short forms 1 and 2. Then after confirmation of the presence of these two short forms in cell lysate IP samples, these samples were probed with anti-DHX9 (RHA) antibody. Although the positive control full length Sam68 clearly interacted with RHA, there was not any band in IP lanes of Sam68 short forms 1 and 2, so it seems that neither of these two isoforms interacts with RNA helicase A. In fact, short form 2 contained all domains of Sam68 except the tyrosine rich domain. This shows that tyrosine rich domain has a functional role in the interaction between Sam68 and RHA. Figure 44 shows the structure difference between full length Sam68 and Short form 2 Sam68 (N-KH-RG).
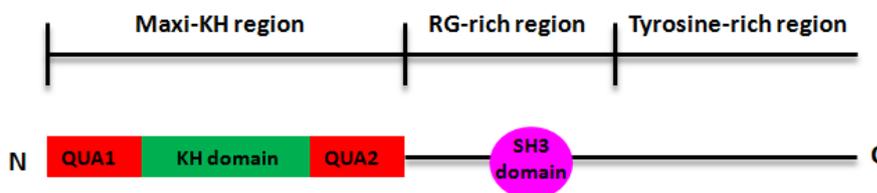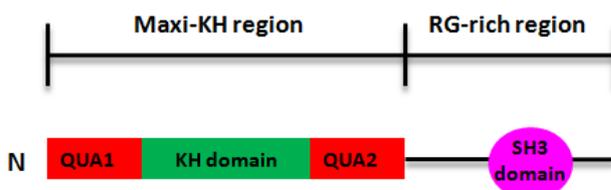
Figure 44: Sam68 full length and short form 2. Deletion of Tyrosine-rich domain of Sam68 interrupts the interaction between Sam68 and RHA. These results show that this region plays an important role in mediating the interaction of these two proteins. A) The picture of full length Sam68. B) The picture of Sam68 short form 2 (N-KH-RG).

According to the important role of tyrosine rich domain to act as a substrate for tyrosine kinase enzymes, this domain can lead to interaction of Sam68 protein with proteins which contain SH2 domains (Najib et al., 2005). Tyrosine-rich sites can be phosphorylated by tyrosine kinases. Phosphorlation of Sam68 is important for its roles in signal transduction within the cell (Vernet and Artzt, 1997). There seem to be two possibilities that may explain why deletion of tyrosine rich domain could disrupt the interaction between Sam68 and RHA. The most obvious explanation is that the tyrosine rich domain is the important one for Sam68 interactions. It is possible that by deletion of the Sam68 tyrosine rich domain, Sam68 interaction with one of the

proteins involved in signal transduction and activation of Sam68-RHA interaction was disrupted and therefore, Sam68 was not able to interact with RHA. The second explanation could be that the deletion of tyrosine-rich domain of Sam68 could change folding of Sam68 protein structure which could negatively affect Sam68 interaction with RHA.

As mentioned in the discussion section of the previous Chapter, RHA, Sam68 and Tap have been implicated in nuclear transport of RNAs which contain either an RRE or CTE (Reddy et al., 2000). Therefore, the plan was to test whether or not these two proteins (Sam68 and RHA) might be involved in other events within the cell such as transcription or alternative splicing regulations. One hypothesis was that there may be links between the roles of Sam68 and RHA in transcription regulation of target genes. Separate studies related to the role of Sam68 and RHA in transcription. In this study the plan was to find a common transcription target to see whether its transcription regulation can be influenced by both Sam68 and RHA proteins. One target gene (known to be regulated by Sam68 and RHA) was *cyclin D1/CCND1.* There was no study to test the role of Sam68 and RHA at the same time in transcription regulation of *cyclin D1*. Therefore, in this study the roles of Sam68 and RHA proteins were tested in transcription regulation of *cyclin D1* after over-expression of Sam68 and RHA in Flp-In HEK293 cells.

According to the previous studies Sam68 has mostly a repressive role in transcription (Hong et al., 2002). Mouse fibroblasts arrest their cell cycle after Sam68 over expression. Levels of *cyclin D1/CCND1* mRNA are reduced after over expression of Sam68, suggesting Sam68 protein might repress expression of *cyclin D1/CCND1* (Hong et al., 2002). Small ubiquitin-like modifier (SUMO) is involved in a post translational modification of Sam68 called sumoylation. It has been shown that sumoylation of Sam68 promotes the role of this protein in inhibition of *cyclin D1/CCND1* gene expression (Babic et al., 2006). Therefore the level of *cyclin D1/CCND1* mRNA was

tetsted by real time PCR after over-expression of RHA, wild type and mutant Sam68 in Flp-In HEK293 cells.

The results in this study showed that while wild type Sam68 does not repress *cyclin D1/CCND1* transcription, P439R mutant Sam68 which is a cytoplasmic protein, increases the level of cyclin D1/*CCND1* mRNA in comparison with the control group (pcDNA5, empty vector). It might be possible that mutant Sam68 (P439R mutant Sam68) helps the accumulation of *cyclin D1 /CCND1* transcripts. The study previously carried out by Taylor et al had shown that Sam68 seems to have a role to inhibit *cyclin D1/CCND1* transcript accumulation (Taylor et al., 2004).

On the other hand, the level of *cyclin D1/CCND1* mRNA in a stable cell line expressing RHA after addition of tetracycline was higher than in the control group but it was not significantly different from the control group (pcDNA5).

The last experiment to investigate the potential roles of Sam68 and RHA proteins was testing the role of Sam68 and RHA in alternative splicing regulation of a target gene *(NRXN3)*. It has been shown that Sam68 is involved in regulation of alternative splicing of the *Neurexin1* gene in the cerebellum (Iijima et al., 2011) and *Neurexin3* (Ehrmann et al., 2013) . Therefore, in this study the roles of Sam68 and RHA were tested in inclusion regulation of alternatively spliced segment 4 (AS4) into *Neurexin3* mRNAs in human cells. After induction of stable cell lines expressing empty vector (negative control, pcDNA5), RHA, wild type and P439R mutant Sam68, the percentage of exon inclusion in each sample was calculated. Similar to our group's previous study (Ehrmann et al., 2013) in this study it was found out that wild type Sam68 has a role in alternative splicing of *NRXN3.* P439R Sam68 does not seem to have a role, likely caused by cytoplasmic location of this protein. In addition, although RHA has a role in splicing regulation of some genes such as its role in splicing of glutamate receptor (Bratt and Ohman, 2003), RNA helicase A protein had no effect on alternative splicing

of *NRXN3* cassette exon 1 (AS4). In conclusion, although RHA and Sam68 have some reported roles in transcription and alternative splicing, more studies are needed to evaluate their common effects on other target genes.

In this study also the interaction of T-STAR and RNA helicase A was tested at their endogenous level of expression. Firstly, HEK293 cells were used to check this interaction. Immunoprecipitation of endogenous T-STAR protein using protein A Dynabeads was carried out and immunoprecipitated (IP) samples were probed with appropriate antibodies, but no RHA was detected in the immunoprecipitated (IP) or the input lanes. This might be because of negligible expression of T-STAR in kidney cells. Although in this study the interaction of RHA with T-STAR was detected after over-expression of this member of STAR family, to investigate more about the interaction between T-STAR and RHA at the endogenous level of T-STAR protein expression, using other cell lines or tissues which have more T-STAR expression might be a better option to try.

To summarise, opposite to Sam68-CCAR1 interaction which happens after over-expression of Sam68, in this study it was found out that RHA and Sam68 interact with each other in both endogenous and over-expressed situation. In addition, the tyrosine rich domain of Sam68 seems to be involved in this interaction. Even just deletion of tyrosine rich domain of Sam68 can inhibit the interaction of this protein with RHA. It should be emphasised that more studies are needed to be carried out to investigate common roles and possible links of these two proteins in signal transduction, transcription and splicing.

In the next Chapter, the investigations regarding hnRNP G, hnRNP G-T and Tra2β protein interaction networks will be described.

Chapter 4

4 Investigation into the cellular roles of hnRNP G (RBMX), hnRNP G-T (RBMXL2) and Tra2β using inducible cell lines

4.1 Introduction

As mentioned previously in Chapter 1, a major aim of this project was to investigate the interacting protein partners of the RNA binding proteins T-STAR, Sam68, hnRNP G, hnRNP G-T and Tra2β. In the two previous Chapters (Chapters 2 and 3), the main focus was on findings regarding T-STAR and Sam68. In the present Chapter main focus will be on similar approaches to investigate protein partners of hnRNP G, hnRNP G-T and Tra2β. The important roles of these RNA binding proteins within cells have been explained in the introduction Chapter. Briefly, hnRNPs not only play key roles in regulation of transcript packaging and translation, but also are involved in alternative splicing (Han et al., 2010). Tra2β also has an important role in alternative splicing regulation (Elliott et al., 2012). Therefore, the primary objective of study was to make stable cell lines expressing FLAG tagged hnRNP G, hnRNP G-T and Tra2β, followed by identification and analysis of interacting groups of these proteins, and analysis of splicing patterns.

To achieve these aims, the open reading frames of hnRNP G, hnRNP G-T and Tra2β were cloned into the pcDNA5 vector and made stable cell line expressing the FLAG tagged proteins, similar to what had been carried out for T-STAR and Sam68. The stable transfected cells had a hygromycin resistance gene, therefore they would survive after adding hygromycin to the media. After few weeks of selection by

hygromycin B, colonies were re-plated. When the cells were confluent enough, expression of the genes of interest were induced by adding tetracycline to the cell culture medium. The next step was purification of hnRNP G, hnRNP G-T and Tra2β proteins plus their partners using immunoprecipitation. Finally, immunoprecipitated proteins were identified by mass spectrometry. The presence of FLAG tagged proteins and some of the candidate protein partners were confirmed using Western blotting.

As will be explained in the result sections, after over expression of hnRNP G-T, there were also some morphological changes identified in the cells. Cells became unattached, circular and floating within the media. One of the aims of the study was testing alternative splicing regulation of some target genes in comparison with control groups to identify some targets responding to hnRNP G-T. It should be mentioned that hnRNP G, hnRNP G-T and Tra2β are the protein names for *RBMX*, *RBMXL2* and *TRA2B* gene products respectively. In this Chapter the appropriate names will be used for the protein or gene.

Chapter 4: Investigation into the cellular roles of hnRNP G (RBMX), hnRNP G-T (RBMXL2) and Tra2β using inducible cell lines

4.2 Methods and Materials

4.2.1 Standard molecular biology

Some of the standard molecular biology techniques including purification of DNA and plasmid DNA, ligations, re-cleavage, dialysis, transformation into *E.coli*, agarose gel electrophoresis and sequencing have already been described (Chapter2, method section).

4.2.1(a) PCR amplification of RBMX, RBMXL2 and TRA2B

The PCR recipes used to amplify *RBMX* and *TRA2B* were identical to those used for the *T-STAR* amplification (table 1). The *RBMXL2* PCR amplification was similar to that used for *Sam68* (table 2). The DNA Engine DYAD thermal cycler (MJ Research) was used for PCR amplifications. The amplification protocols are identical to table 4 (for *RBMX* and *TRA2B*) and table 5 (for *RBMXL2*) (more information is available in Chapter 2, method section). Primers used for amplifying the open reading frames of *RBMX*, *TRA2B* and *RBMXL2* are shown in table 24.

Primer name	Sequence (5'-3')	Restriction Enzyme
Tra2B-FLBamHF	AAAAAAAGGATCC ATGAGCGACAGCGGCGAGCAGA	*BamH*I
Tra2B-FL-SalI	AAAAAAAGTCGACTTAATAGCGACGAGGTGAGTATGATCGA	*Sal*I
RBMX-BamhI-F	AAAAAAAGGATCC ATGGTTGAAGCAGATCGCCCAGGA	*BamH*I
RMBX-SalI-R	AAAAAAAGTCGACCTAGTATCTGCTTCTGCCTCCCC	*Sal*I
RBMXL2-BamhI-F	AAAAAAAGGATCCATGGTTGAAGCGGATCGCCCG	*BamH*I
RMBXL2-SalI-R	AAAAAAAGTCGAC TTA GTATCTGCTCCGGCCTCCTC	*Sal*I
BGHrev	TAGAAGGCACAGTCGAGG	---------

Table 24:Primers used to amplify *TRA2B*, *RBM* and *RBMXL2*

Chapter 4: Investigation into the cellular roles of hnRNP G (RBMX), hnRNP G-T (RBMXL2) and Tra2β using inducible cell lines

4.2.1(b) Restriction digests

The restriction enzymes *BamH*I and *Xho*I were used for digestion of plasmid DNA (pcDNA5). *BamH*I and *Sal*I were also used for digestion of PCR products. BSA (Bovine Serum Albumin) was used for enzymes to work efficiently. PCR products were incubated with each restriction enzyme separately for 3 hours. Digestion mixes are shown in table 25:

	PCR product			Plasmid (pcDNA5 vector)	
	First digestion	Second digestion		First digestion	Second digestion
dH$_2$O	9µl	……	dH$_2$O	32 µl	…..
Purified DNA	20 µl	30µl	Purified DNA	5 µl	38 µl
BSA (10X)	4 µl	3µl	BSA (10X)	5 µl	3µl
Buffer	4 µl	4µl	Buffer	5 µl	5µl
Enzyme	3 µl	3µl	Enzyme	3 µl	4µl
Total	40 µl	40 µl	Total	50µl	50 µl

Table 25: First and second digestion strategy used for PCR Products (left) and Plasmid digestions (right) mixtures.

4.2.2   Cell culture

Generation of stable cell lines and harvesting of cells have been described previously (Chapter 2, method section).

4.2.3 Protein based methods

SDS polyacrylamide gel electrophoresis (SDS-PAGE), Western blot (Protein immunoblot analysis), immunoprecipitation using α-FLAG affinity resin and mass spectrometry have been described previously (Chapter 2, method section). Tables 27 and 28 show the antibodies used for Western blots and immunoprecipitations.

Primary Antibody	Source	Monoclonal /Polyclonal	Species	Dilution for Western Blot
Monoclonal anti Flag	Sigma	Monoclonal	Mouse	1:1000
Anti Actin	Sigma	Polyclonal	Rabbit	1:1000
Anti Sam68 (C-20)	Santa Cruz	Polyclonal	Rabbit	1:1000
Anti hnRNP C1/C2(4F4)	Santa Cruz	Monoclonal	Mouse	1:200
Anti Tra2β antibody (ab31353)	Abcam	Polyclonal	Rabbit	1:2000

Table 26: Primary antibodies. Antibodies used for Western blot.

Secondary Antibody	Source	Monoclonal /Polyclonal	Species	Dilution for Western Blot
Anti-mouse-HRP	Amersham	Polyclonal	Sheep	1:1000
Anti-rabbit-HRP	Jackson lab	Polyclonal	Goat	1:1000

Table 27: Secondary antibodies. Secondary antibodies used for Western blot

Chapter 4: Investigation into the cellular roles of hnRNP G (RBMX), hnRNP G-T (RBMXL2) and Tra2β using inducible cell lines

4.2.4 *RNA based methods*

The RNA extraction procedure has been explained in the previously (Chapter 3, method section).

4.2.4(a) RT-PCR (Reverse transcriptase PCR)

RT-PCR using the SuperScript® VILO™ cDNA Synthesis Kit (Invitrogen) has been mentioned already (Chapter 3, method section). In addition, the Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo scientific) was used to transcribe cDNA from RNA samples. Table 28 shows the reaction set up.

Reagent	Quantity for a 10µl reaction
(Diethylpyrocarbonate) DEPC-treated water (nuclease free dH$_2$O)	Up to 50µl
5X Reaction Mix	10µl
Maxima Enzyme Mix	5µl
RNA (15µg)	Xµl

Table 28: RT-PCR recipe. Based on Maxima First Strand cDNA Synthesis Kit for RT-qPCR recipe.

RT-PCR reactions were incubated firstly at 25°C for 10 minutes and this step was followed by two other steps, 50°C for 30 minutes and 85°C for 5 minutes respectively. The DNA Engine DYAD thermal cycler (MJ Research) was used for RT-PCR amplifications.

4.2.4(b) Splicing assay

Splicing assay steps including PCR recipe and programme have been described in the previous Chapter (Chapter 3, method section). Table 29 shows the primers used in PCR reactions. To estimate PCR product sizes and their relative concentrations, the Bio calculator software was used.

Primer name	Sequence (5'-3')
MINK1F	ACCTCATCTATCGCCACTGC
MINK1R	AGGCTGGAAGGACGCTGTC
ACOT9	CAAAGGGCAGCTTACTCCTG
ACOT9R	CTCCCGCAACTTATCTCGAA
ENAHF	GAGCCTGTAACTTCTAAGGCCTC
ENAHR	GCTGTGATAAGGGTGTGGATTT
MYOFF	AGTGGAATCAGGTCGTCAATCT
MYOFR	TCAGGTAACAAGGTCCAAACG

Table 29: Primers used in PCR reactions to check alternative splicing patterns of target genes.

4.2.5   MTT cell proliferation assay

For this experiment an MTT cell proliferation assay kit (Cambridge bioscience, item number 10009365) was used. 10000 cells/well were seeded in 100µl of 10% FBS DMEM with glutamax-1 medium in six 96-well plates (plates 1-6, one plate for each day). Cells were incubated at 37°C in a 5% CO_2 incubator overnight. Next day, 2 µl of 0.05mg/ml tetracycline was added to each well of all plates and cells were incubated

for 24 hours. After 1 day, 10 μl of MTT reagent was added and mixed gently with the media of all wells in one of the plates. Also, 2 μl of 0.05mg/ml tetracycline was added to each well of the other plates and they were incubated at 37°C. After 4 hours incubation of cells with MTT reagent at 37°C, dark crystals called formazan (cell metabolization products), were present in each well. Media was removed gently and 100μl of crystal dissolving solution was added to each well and mixed. By adding this solution crystals were dissolved and a purple colour solution was produced. This procedure was continued for 6 days and the Microplate reader was used to measure the absorbance of all samples at 570nm. All samples were tested in triplicate. Also, there were blank wells for all samples which did not contain any cells but MTT reagent was added to them as a background control. Data were analysed using Excel software (Student's t-test).

4.2.6 Apoptosis, cell proliferation and DNA damage assay

An apoptosis, DNA damage and cell proliferation kit (BD Biosciences) was used to check the percentage of apoptotic and DNA damaged cells. Seeding cells was carried out at a density of 1×10^5/ml. After induction of cells with tetracycline for 3 days, 100μl of 1mM BrdU in 1XPBS was added gently to the cells and they were incubated for 1 hour at 37°C in a 5% CO2 incubator. To keep the floating cells, the media were transferred to appropriate 15ml falcon tubes and cells were harvested and transferred to 15ml falcon tubes (for more information regarding harvesting the cells see Chapter 2, method section). Falcon tubes were then centrifuged at 1000rpm for 2 minutes and the supernatant was removed. 1ml 1XPBS was added to the cell pellets and mixed gently. Cells were then transferred to FACS (fluorescence activated cell sorting) tubes. According to the manufacturer's instructions (apoptosis, DNA damage and cell proliferation kit, BD Biosciences), cells were fixed and permeabilized using BD cytofix/cytoperm fixation/permeabilization solution, and then incubated with BD

cytofix/cytoperm plus permeabilization buffer. The next steps were re-fixation of cells and their treatment with DNase. Fluorescent antibodies were used for staining BrdU, cleaved PARP and H2AX. Also, DAPI solution (1µg/ml) was used for staining DNA. Finally, 1ml of 1XPBS was added to resuspend cells. Unstained induced Flp-In HEK293 cells were also used as the control to set up the instrument. All samples were performed in triplicate. A flow cytometer (BD FACSCanto II flow cytometer) was used for cell sorting. BD Facsdiva version 6.1.3 and Excel software (Student's *t*-test) were used for analyzing the data. Setting up the instrument for analyzing the samples was carried out by Mr Ian Dimmick, flowcytometry core facility manager (Newcastle University, Bioscience centre, International centre for life). Table 30 shows the components of the apoptosis, DNA damage and cell proliferation kit.

Kit components	Working stock solution
BrdU (10 mg/ml)	10 µl of 1mM BrdU/ml
PerCP-Cy™5.5 Mouse Anti-BrdU	5µl/test
PE Mouse Anti-Cleaved PARP (Asp214) Antibody	5µl/test
Alexa Fluor 647 Mouse Anti-H2AX (pS139)	5µl/test
BD Cytofix/Cytoperm™ Fixation/Permeabilization Solution	100 µl/test
BD Perm/Wash™ Buffer (10X)	100 µl/test
BD Cytofix/Cytoperm™ Plus Permeabilization Buffer	100 µl/test
DNase (1mg/ml)	300µg/ml
DAPI (1mg/ml)	1µg/ml

Table 30: Apoptosis, cell proliferation and DNA damage kit components.

4.2.7   Cell viability analysis

Seeding cells was carried out at a density of 2×10^5/ml. After induction of cells with tetracycline for 5 days, media were transferred to 15ml falcon tubes, and cells were harvested (as already described, Chapter 2, method section) and transferred to falcon tubes. This step was followed by centrifugation at 1000rpm, removing the media, adding fresh 10% FBS media and mixing cells. Then 500 μl of cell mixture was transferred to a sample cup and placed in the carousel position in the cell viability analyzer. Vi-CELL Cell Viability Analyzer (Beckman coulter) was used for checking the percentage of cell viability.

4.3   Results:

4.3.1   Cloning the open reading frame of RBMX, RBMXL2 and TRA2B into pcDNA5

Forward and reverse primers were used to PCR amplify the open reading frames of *RBMX* (encodes hnRNP G), *RBMXL2* (encodes hnRNP G-T) and *TRA2B* (encodes Tra2β). The cloning strategies for these three genes were similar to those used for *T-STAR* and *Sam68*, which are explained previously in Chapter 2 (method section). To amplify the open reading frame of *RBMX*, RBMX-BamhI-F, RMBX-SalI-R and image clone no 1891274 were used as forward primer, reverse primer and template respectively. PCR amplification resulted in a 1Kb product (figure 45-A). To amplify the open reading frame of *RBMXL2*, the forward primer RBMXL2-BamhI-F, the reverse primer RBMXL2-SalI-R and *RBMXl2* in the pGFP3 construct as template were used. PCR amplification resulted in a 1Kb product (figure 45-B). To amplify the open reading frame of *TRA2B*, Tra2B-FL-BAMHF, Tra2B-FL-SalI and *TRA2B* in the pGFP3 construct were used as forward primer, reverse primer and template respectively. PCR amplification resulted in a ~ 0.8Kb product (figure 45-C).
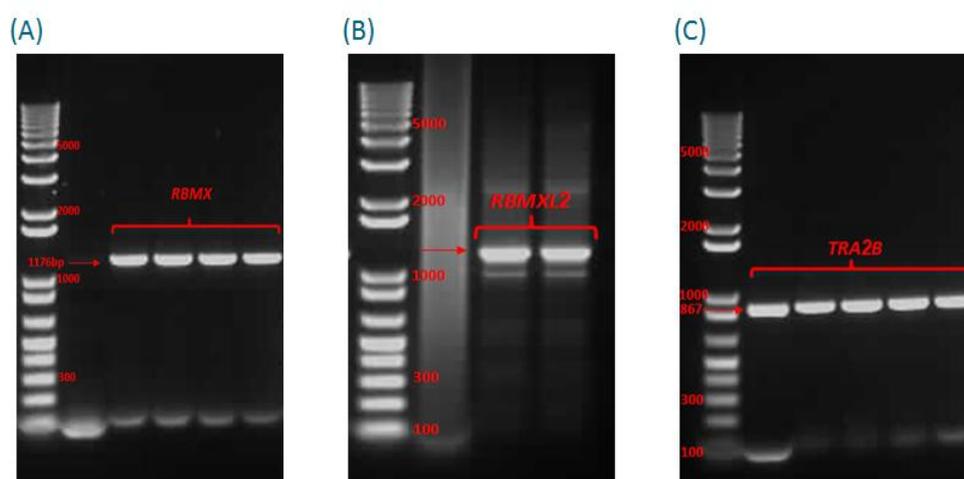
Figure 45: PCR amplification of *RBMX, RBMXL2 and TRA2B for cloning into the pcDNA5 vector.* PCR products were visualised using ethidium bromide stained agarose gels. A) The PCR product size of *RBMX* was 1176bp. B) Size of *RBMXL2* PCR product was 1179bp. C) The PCR product size of *TRA2B* was 867bp. The red arrows show PCR products. 1kb plus DNA ladder also has been shown in these pictures.

A combination of *BamH*I and *Sal*I were used for digestion of PCR amplified *RBMX*, *RBMXL2* and *TRA2B*. Also, *BamH*I and *Xho*I were used to digest pcDNA5 (vector). *Not*I and *Xho*I were used for the re-cleavage process to remove the ligated vectors which did not contain *RBMX* and *TRA2B* respectively. 2µl of dialysed ligation mixtures for cloning *RBMX*, *RBMXL2* and *TRA2B* were mixed separately with 20 µl of electro-competent cells and transferred to electroporation cuvettes. After electroporation, cells were incubated in 1 ml LB broth for 1 hour at 37°C enable them to recover. LB agar plates which contain 50µg/ml ampicillin were used for plating out competent cells. Plates were incubated at 37°C overnight to enable colonies to grow.

After overnight incubation, insert forward primers including: RBMX-BamhI-F, RBMXL2-BamhI-F, Tra2B-FL-BAMHF and vector reverse primer, BGHrev were used to screen the colonies by PCR. Colonies which contained the insert were grown in LB medium plus

50µg/ml ampicillin. After overnight incubation at 37°C on a shaker, plasmid purification was carried out. To confirm the correct presence of inserts in the pcDNA5 vector, 100ng/µl of constructs were sequenced. The results confirmed the presence of inserts containing the open reading frames of *RBMX*, *RBMXL2* and *TRA2B* within the constructs. The sequences of each construct are shown in Appendix A.

4.3.2   *Generation of stable cell lines expressing hnRNP G, hnRNP G-T and Tra2β*

The establishment of stable cell lines expressing hnRNP G, hnRNP G-T and Tra2β FLAG-tagged proteins were carried out similarly to how stable cell lines expressing T-STAR, wild type and P439R Sam68 were made (explained in Chapter 2, result section 3-3). Briefly, when Flp-In HEK293 cells were 60% confluent, the constructs containing the open reading frames of *RBMX*, *RBMXL2*, *TRA2B* and pcDNA5 (empty vector) were co-transfected with pOG44 plasmid (which encodes Flp recombinase) into the cells. Selection of stable transfected cells was carried out using hygromycin. After few weeks of selection by hygromycin B, colonies were re-plated. When the cells were confluent enough, tetracycline was added to the cell culture medium to induce expression of the genes of interest. To describe proteins, the names hnRNP G and hnRNP G-T are used instead of RBMX and RBMXL2 respectively (figure 46).
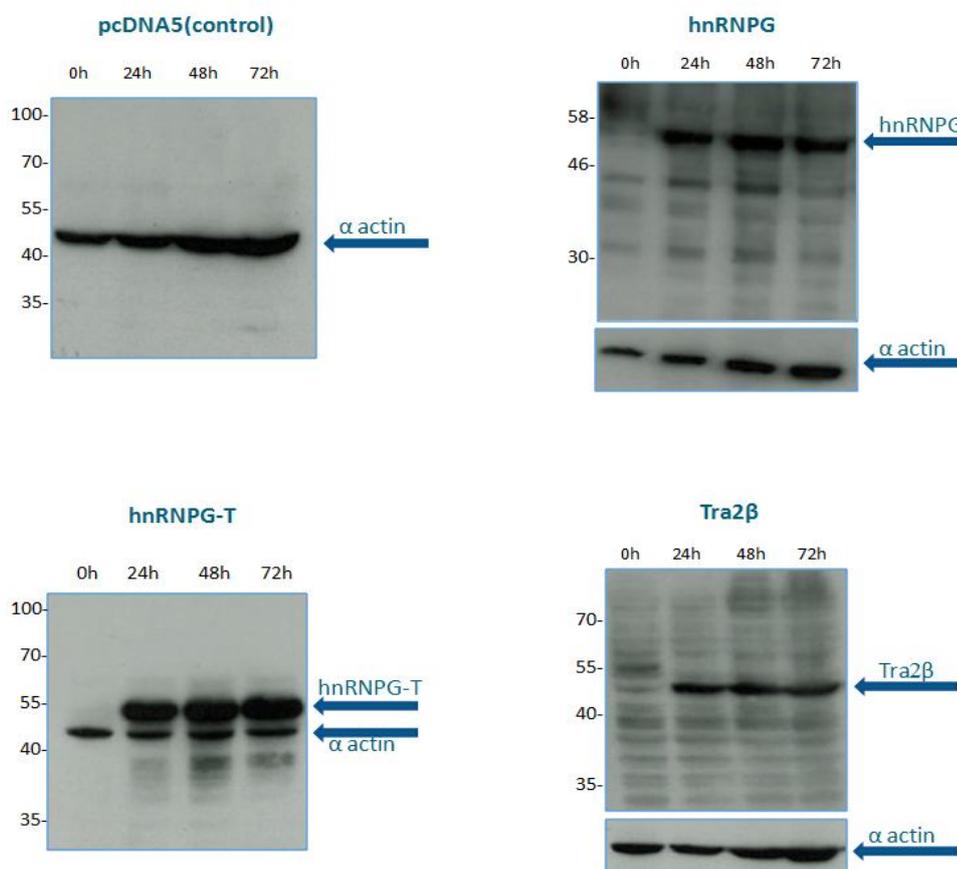
Figure 46: Induced expression of FLAG tagged proteins (hnRNP G, hnRNP G-T and Tra2β) monitored by Western blotting. The presence of the FLAG tagged proteins (hnRNP G, hnRNP G-T and Tra2β) was confirmed in the stable cell lines after 24, 48 and 72 hours induction in comparison with empty vector (pcDNA5) as the negative control. No expression was seen before induction (0h). Because there are FLAG tags and His tags contained within these recombinant proteins, their molecular weights were higher than endogenous proteins. Monoclonal anti FLAG and anti-mouse-HRP were used as the primary and secondary antibodies respectively to detect the presence of FLAG tagged proteins. Blots were re-probed with α-β actin and α-rabbit IgG (HRP) as the primary and secondary antibodies respectively to confirm that the samples had been loaded equally. Equal levels of actin loading control were present in each lane.

4.3.3   *Role of hnRNP G-T in cell morphology regulation*

While control cells (stable cell line expressing empty FLAG tag vector) were flat and attached to the cell culture surface, after induction of the hnRNP G-T expressing stable cell line with tetracycline, some cell morphological changes occurred. Most of the expressing cells became rounded, and floated into the media after tetracycline induction. Since the *RBMXL2* gene is derived from transcript retrotransposition of the *RBMX* gene (Elliott et al., 2000b)**,** hnRNP G was used as an extra control group to study these morphological changes. No morphological changes were observed in the control groups (expressing pcDNA5 and hnRNP G). The observed cell densities after tetracycline induction indicated that growth rates may have been reduced after hnRNP G, and specifically hnRNP G-T, over expression (figure 47).
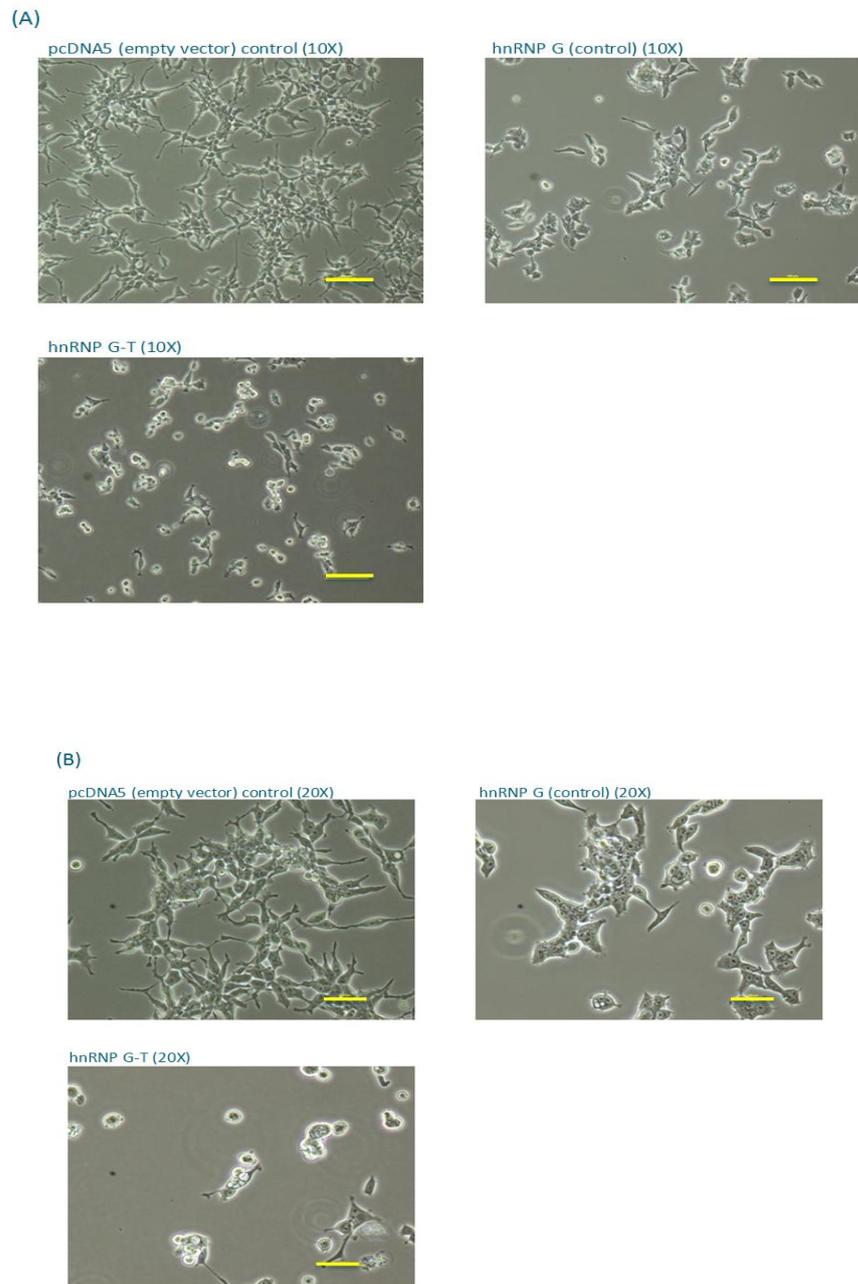
Figure 47: Morphological changes induced after 3 days tetracycline induction of the hnRNP G-T expressing stable cell line. After induction with tetracycline for 3 days some cell morphological changes happened. While the hnRNP G-T stable cell line became round and floated in the media, control cell lines expressing empty vector (pcDNA5) and hnRNP G did not show any changes. These latter two groups of cells remained attached to the cell culture

surface. In addition, cell growth seemed to be reduced in both cell lines expressing hnRNP G and hnRNP G-T in comparison with the cell line expressing the empty vector. The scale bars are equal to 100μm. A) 10X magnification B) 20X magnification.

Also the progress of these morphological cell changes was tested over a longer period of time. Cells were induced for 9 days with tetracycline and pictures were taken by light microscopy to check the effect of hnRNP G-T over this longer expression time. While control groups expressing pcDNA5 (empty vector) and hnRNP G cells remained attached to the cell culture surface and had normal shapes, cells expressing hnRNP G-T again were mostly circular and released within the media (figure 48).

(A)

pcDNA5 (empty vector) control (10x)

hnRNP G (control) (10x)

hnRNP G-T (10x)

(B)

pcDNA5 (empty vector) control (20x)

hnRNP G (control) (20x)

hnRNP G-T (20x)
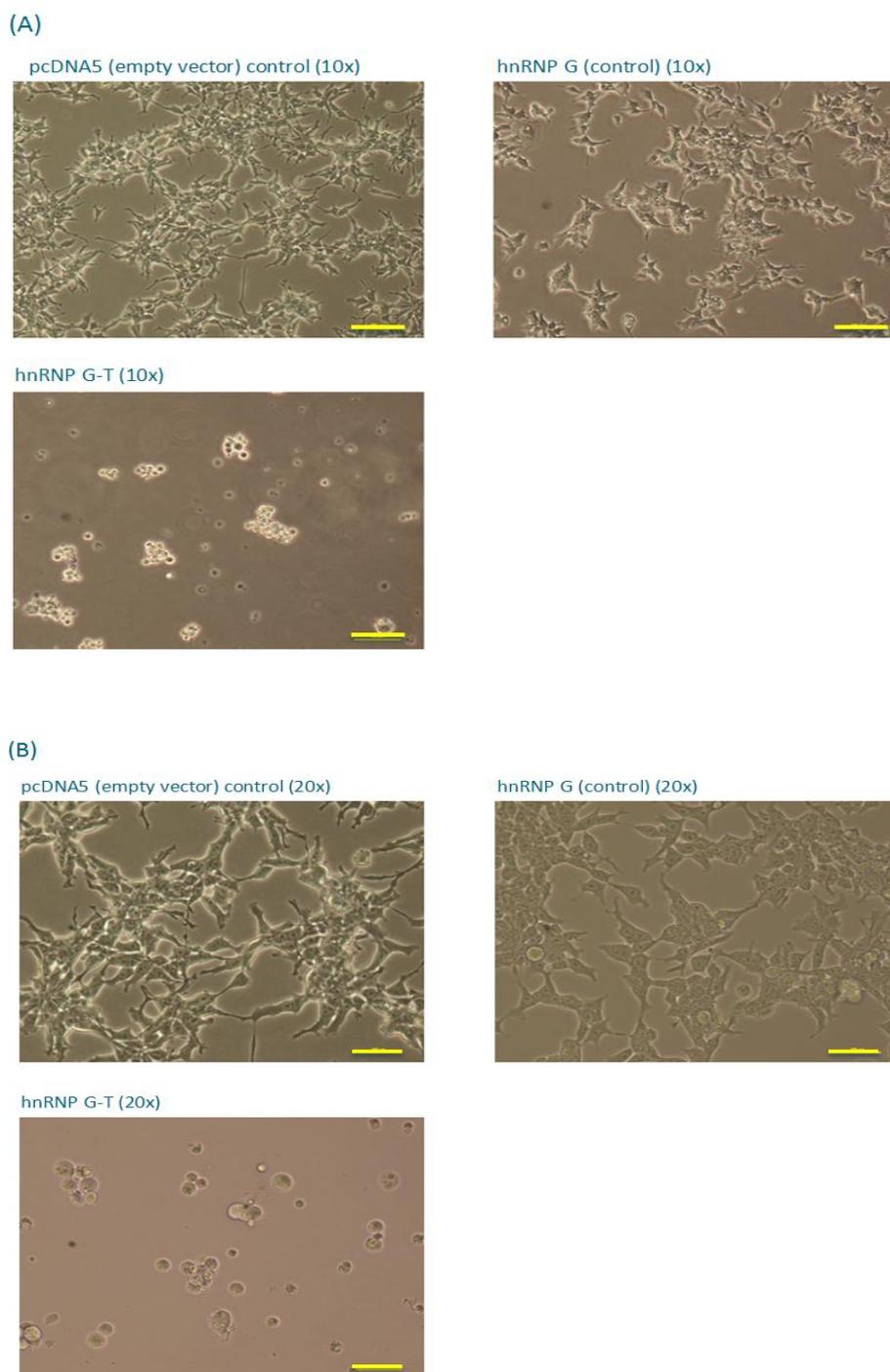
Figure 48: Morphological changes after 9 days induction of the hnRNP G-T stable cell line. After 9 days induction with tetracycline, while control cells including empty vector (pcDNA5) and hnRNP G had normal morphological shapes and were attached to the cell culture surface,

stable cell lines expressing hnRNP G-T had circular shapes and were released within the media. In addition, cell growth was hugely reduced in the hnRNP G-T group in comparison with the control cells. The scale bars are equal to 100μm. A) 10X magnification B) 20X magnification.

4.3.4 Testing role of the hnRNP G-T in cell viability of FLP-In HEK293 cells using the MTT assay

To test the cell viability rate of stable cell lines expressing empty FLAG tagged vector (pcDNA5), hnRNP G and hnRNP G-T proteins, 10000 cells/well were seeded in 100μl of 10% FBS DMEM media in six 96-well plates (plates 1-6, one plate for each day). Also, there was one row of wells in all the plates which contained only 10% FBS DMEM media to be considered as a background control.

Each day the cell density on one of the plates was measured and rest of other plates were induced for another 24 hours using tetracycline. Induction was carried out for 6 days, and cell viability of cells during this period of time was measured using a MTT assay. Using this method, the more cells present in the wells, the stronger a purple colour would appear. The cell viability rate in the hnRNP G-T expressing clone of cells was lower and significantly different in comparison with both control groups (pcDNA5 and hnRNP G) ($P<0.05$). In addition, between the two control groups (hnRNP G and pcDNA5), the cell confluence and colour strength of the hnRNP G group were less and significantly different in comparison with pcDNA5 ($P<0.05$). In agreement with the above observations, over-expression of hnRNP G and hnRNP G-T proteins reduced cell viability rates and this reduction was significantly different from cell lines expressing empty vector (pcDNA5) (figure 49).
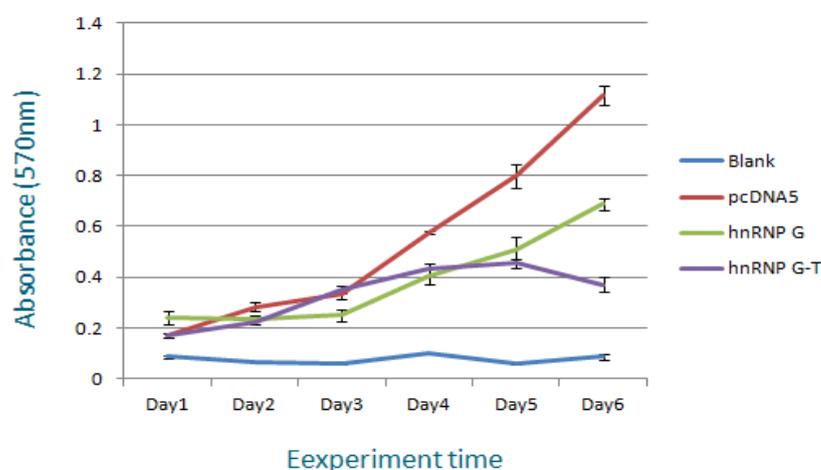
Figure 49: Cell viability of the stable cell lines using an MTT assay. The cell viability rate of stable cell lines expressing pcDNA5, hnRNP G and hnRNP G-T were measured using an MTT assay during 6 days after induction with tetracycline. The cell viability rate of the stable cell line expressing hnRNP G-T was less than the controls ($P<0.05$). Also the control groups (pcDNA5 and hnRNP G) were significantly different from each other in proliferation ($P<0.05$). Expression of hnRNP G and hnRNP G-T reduce the cell viability rate. The blank control only had media and MTT reagent and its absorbance was measured as background control. All samples were performed in triplicate.

4.3.5 HnRNP G-T has a role in apoptosis, DNA damage and cell proliferation in stable FLP-In HEK293 cells

To investigate more about the role of hnRNP G-T on morphological changes, and also to find out whether the circular floating cells expressing hnRNP G-T are apoptotic or DNA damaged, a kit (apoptosis, DNA damage and cell proliferation kit) was used. Cells were seeded and induced for 3 days with tetracycline. 100µl of 1mM BrdU (an analog of thymidine) in 1XPBS was added gently to the cells to test the cell cycle kinetics. After one hour incubation at 37°C in a 5% CO_2 incubator, both floating and attached cells

were harvested. According to the manufacturer's instructions cells were fixed, permeabilized and treated with DNase. Phosphorylated histone H2AX (phosphorylation of C-terminal serine 139) was used as a biomarker to measure cell response to DNA damage (Valdiglesias et al., 2013).

This kit also measured apoptosis, through a nuclear enzyme called PARP (poly [ADP-ribose] polymerase) which has role in transcription regulation and DNA repair, and which can be cleaved by caspase-3 and -7. After cleavage, the 116kDa PARP will be divided to two smaller fragments, and can be used as a marker of apoptosis (Kaufmann *et al.*, 1993; Lazebnik *et al.*, 1994; Nicholson *et al.*, 1995; Los *et al.*, 1997; Los *et al.*, 1999). Fluorescent antibodies were used for staining BrdU, cleaved PARP and H2AX to test cell proliferation, apoptosis and DNA damage. Also DAPI solution (1µg/ml) was used for staining DNA. Finally cells were resuspended in 1ml of 1XPBS and samples were analyzed by flow cytometry.

The percentage of cells in S phase in the hnRNP G-T expressing group of cells was lower and significantly different in comparison with both control groups (stable cell lines with pcDNA5 and hnRNP G) (P<0.05). The two control groups were significantly different from each other as well (P<0.05). These results showed that over expression of hnRNP G and hnRNP G-T causes reduction of cell proliferation. Furthermore, hnRNP G-T has stronger effect on this reduction since the percentage of cells in S phase in this group was significantly lower than percentage of S phase cells in hnRNP G-expressing group (P<0.05) (figure 50-B).

The total percentage of apoptotic cells in the hnRNP G-T expressing group of cells was less than 2% of the total cell population, but still was higher and significantly different in comparison with both control groups (pcDNA5 and hnRNP G) (P<0.05). In addition, the control groups (pcDNA5 and hnRNP G) were significantly different from each other

(P<0.05). These results showed that over-expression of hnRNP G and hnRNP G-T proteins can lead to a higher percentage of cell apoptosis (figure 50-C), although this effect is very small.

Also DNA damage percentage was investigated in the different clones of cells. Interestingly, while the control groups (expressing empty pcDNA5 and hnRNP G) were not significantly different (P<0.05), the percentage of cells with DNA damage in the hnRNP G-T group was less and significantly different in comparison with both control groups (pcDNA5 and hnRNP G) (P<0.05) (figure 50-D). Although statistically significant, this effect was very small.

In summary, expression of hnRNP G and specifically hnRNP G-T causes a reduction in cell growth rate and increases the level of apoptosis. Although cells expressing hnRNP G-T became circular and floated within the media, only very few cells (less than 2%) were apoptotic cells. The apoptosis number is very small so might not be biologically significant, even if it is statistically significant.

Specifically regarding to lower percentage of DNA damage within these cells in comparison with control groups, expression of the hnRNP G-T protein not only has no adverse effect on DNA stability, but might also lead to a reduction in the amount of DNA damage possibility, although this effect is very small. We could not therefore consider hnRNP G-T as an apoptosis inducer. However the S phase numbers might be more biologically significant and were much stronger.
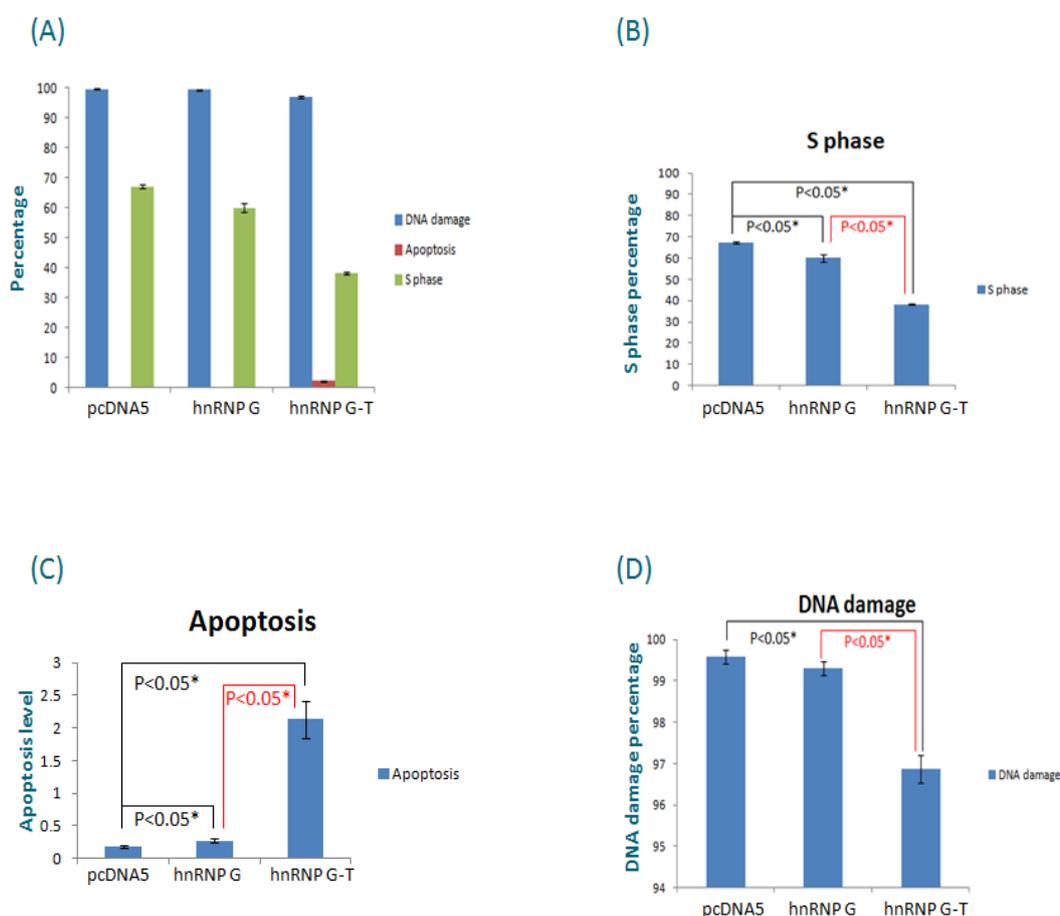
Figure 50: Percentages of S phase, apoptotic and DNA damaged cells in stable cell lines expressing pcDNA5, hnRNP G and hnRNP G-T. A) Summarised percentages of cells with DNA damage, apoptosis and in S phase from induced clones of cells expressing empty pcDNA5 vector, hnRNP G, hnRNP G-T. B) Percentage of cells in S phase in the cells stably expressing hnRNP G-T was significantly lower from both control groups ($P<0.05$). Also, the percentage of S phase cells in hnRNP G group was significantly different from the pcDNA5 group ($P<0.05$). These results indicate that expression of hnRNP G and hnRNP G-T reduce cell growth rate. C) Although less than 2% of total cells expressing hnRNP G-T were apoptotic, this group was significantly different from the two control groups ($P<0.05$). Also, the two control groups were significantly different from each other ($P<0.05$). It seems that an increasing level of apoptosis results from the over-expression of hnRNP G and hnRNP G-T. D) DNA damage was less but significantly different in hnRNP

G-T expressing cells in comparison with control groups (pcDNA5 and hnRNP G) (*P*<0.05). The control groups were not significantly different from each other (*P*>0.05). All samples were tested in triplicate. The asterisk shows that the difference between two groups was significant.

4.3.6   How many stable cells expressing hnRNP G-T are still alive?

The round, floating state of the hnRNP G-T expressing cells suggested they may be dead. To test the percentage of live cells expressing hnRNP G and hnRNP G-T after induction with tetracycline after 5 days, cells were harvested and transferred to falcon tubes. Cells were resuspended in 3mls of 10% FBS DMEM media and mixed. 500 µl of the cell mixture was transferred to a sample cup and placed in the carousel position in a cell viability analyzer. The percentage of cell viability was measured. Around 93.8% of cells expressing hnRNP G-T were viable but this percentage of live cells was lower and significantly different in comparison with both control groups (pcDNA5 [97.8%] and hnRNP G [98.9%]) (*P*<0.05). Control groups were not significantly different from each other in this case (*P*>0.05) (figure 51). Although the hnRNP G-T group was significantly different from the control groups, still around 94% of cells were viable and the difference was small. Hence, even though cells over-expressing hnRNP G-T were circular and released within the media the majority of them were still alive.
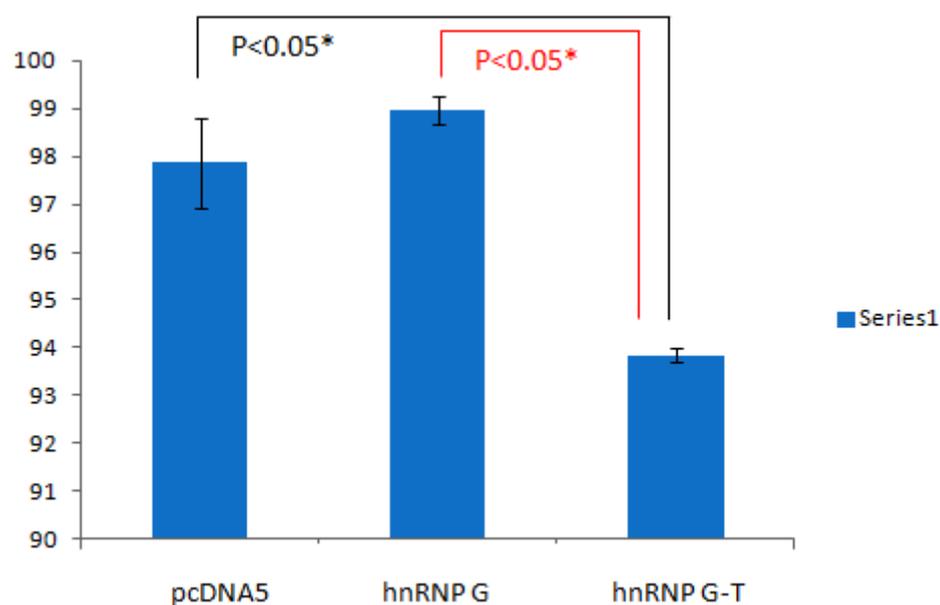
Figure 51: Cell viability percentage of stable cells expressing pcDNA5, hnRNP G and hnRNP G-T. Although cells expressing hnRNP G-T were significantly different from control groups ($P<0.05$), the percentage of viable cells was 93.8%. The two control groups (pcDNA5 and hnRNP G) were not significantly different from each other ($P>0.05$). All samples were tested in triplicate. The asterisk shows that the difference between two groups was significant.

4.3.7 Are there any effects of over-expression of hnRNP G and hnRNP G-T at the splicing level?

According to the important roles of hnRNP G and hnRNP G-T in splicing regulation and because cells stably expressing hnRNP G-T seemed to exhibit morphological changes rather than cell death, there was a question to be answered that what would happen at the splicing level after over-expression of hnRNP G-T. It seemed that after expression of hnRNP G-T cell motility and cell-cell adhesion would be changed. In response to Dr Julian P Venables' suggestion (Institute of Genetic Medicine), we analysed splicing patterns using a high throughput platform. After induction of stable

Chapter 4: Investigation into the cellular roles of hnRNP G (RBMX), hnRNP G-T (RBMXL2) and Tra2β using inducible cell lines

cell lines expressing pcDNA5, hnRNP G and hnRNP G-T with tetracycline, cells were harvested and RNA was extracted. The Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo scientific) was used to transcribe cDNA from RNA samples (see method section).

0.3μg/μl of the RNA samples were then sent to a lab in Canada (Laboratoire de génomique fonctionnelle de l'Université de Sherbrooke, Sherbrooke, Québec) for analysis. Alternative splicing patterns in 26 genes associated with EMT (epithelial–mesenchymal transitions) were monitored between the three groups of stably expressing clones (pcDNA5, hnRNP G and hnRNP G-T) using a high-throughput RT-PCR screen. Data was then analyzed by Dr Venables and four target genes including: *MINK1* (misshapen-Like Kinase 1), *ACOT9* (acyl-CoA thioesterase 9), *ENAH* (enabled homolog) and *MOYF* (myoferlin) were detected to be alternatively spliced in the presence of hnRNP G-T since their PSI (percentage of splicing inclusion) was different in the hnRNP G-T group in comparison with the control groups (figure 52).
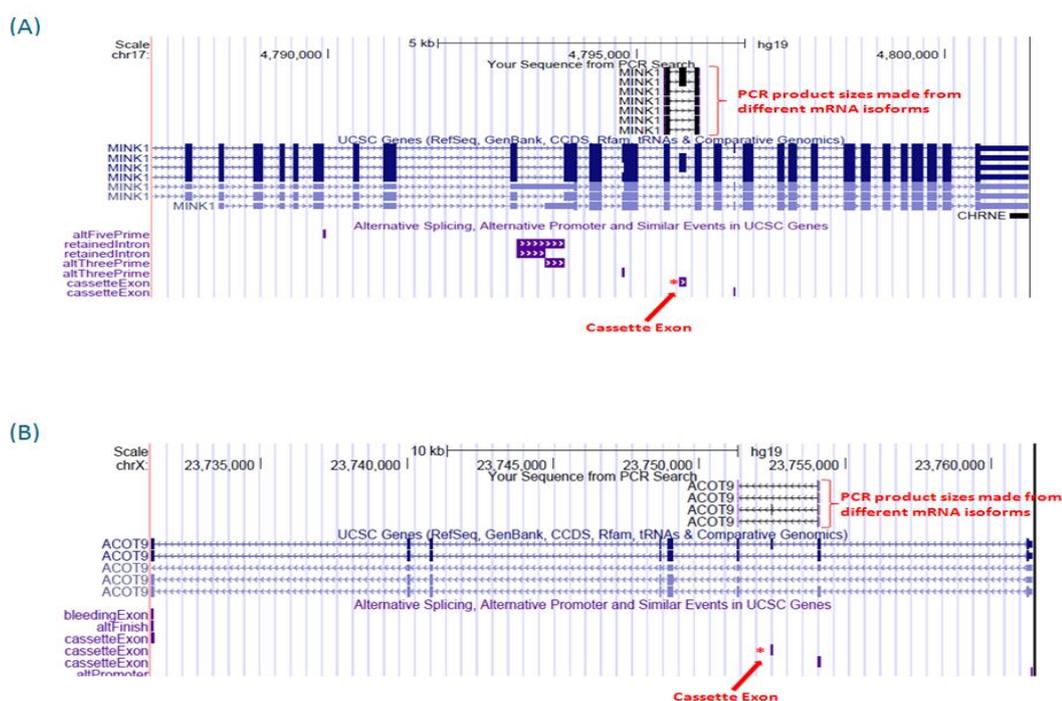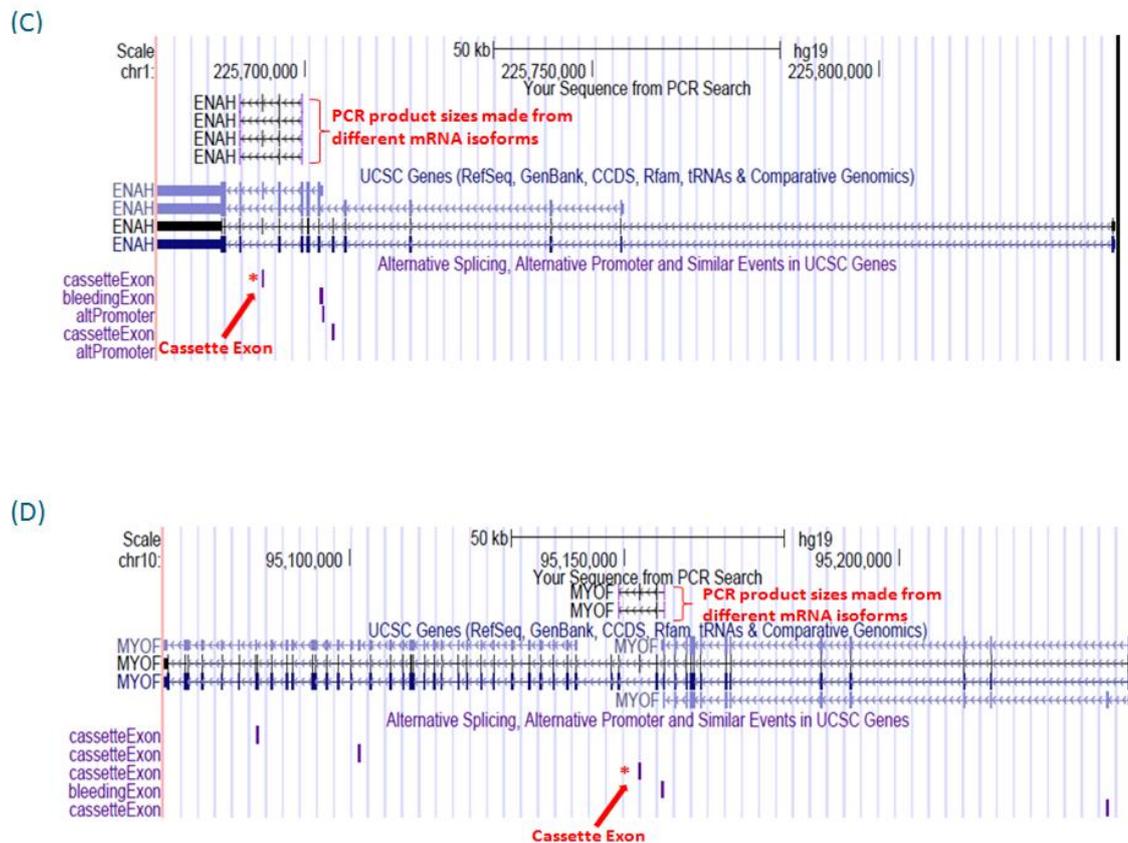
Figure 52: Alternative splicing pattern of detected cassette exons regulated by hnRNP G-T in *MINK1* (misshapen-Like Kinase 1), *ACOT9* (acyl-CoA thioesterase 9), *ENAH* (enabled homolog) and *MOYF* (myoferlin). Exons in these genes were detected to be alternatively spliced in the presence of hnRNP G-T. Here a screenshot from UCSC shows the sizes of the expected PCR products with or without these cassette exons. Cassette exons are shown by red asterisks. A) *MINK1* short form: 136bp, *MINK1* long form: 247bp. B) *ACOT 9* short form: 103bp,) *ACOT 9* long form: 130bp. C) *ENAH* short form: 133bp, *ENAH* long form: 196bp. D) *MYOF* short form: 216bp, *MYOF* long form: 255bp.

The next step was to reproduce these results in our lab. To do this, 1×10^{6} cells (1×10^{5}/ml) were seeded, induced with tetracycline antibiotic and then harvested. RNA extraction and RT-PCR using SuperScript® VILO™ cDNA Synthesis Kit (Invitrogen) were

carried out to make cDNA. 1µg/µl of sample RNA was then used for making cDNA. This step was followed by diluting the obtained cDNA (at a ratio of 1:4) and setting up 25 µl polymerase chain reactions using GoTaq polymerase. Forward and reverse primers (table 29) were used for PCR amplifications of *MINK1*, *ACOT9*, *ENAH* and *MYOF*. All RNA samples were analysed using three independently isolated RNA samples. The QIAXCEL multi-capillary electrophoresis system was used for analysing the relative quantities of the different PCR products. In each case the percentages of splicing inclusion of the variable exons in the *ACOT9* and *MYOF* genes were significantly different in stable cells expressing hnRNP G-T group from the control groups (expressing pcDNA5 and hnRNP G) (*P*<0.05). This suggests that the expression of hnRNP G-T has a role in preventing exon inclusion within these two genes (figure 53). These effects could be direct or indirect. If direct, then hnRNP G-T might directly bind to these pre-mRNAs and change their splicing pattern. If indirect, it could be that expressing hnRNP G-T affects the cells, which then affects the splicing pattern. It seems that hnRNP G-T is acting as a repressor (less exon is included in each of the three replicates), and the result is specific to hnRNP G-T (no effect seen for hnRNP G).
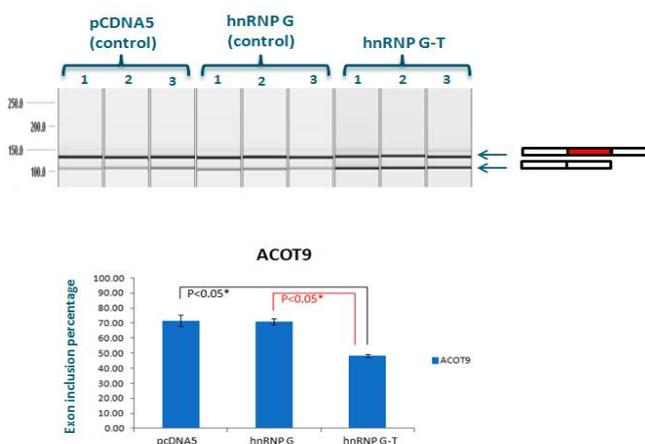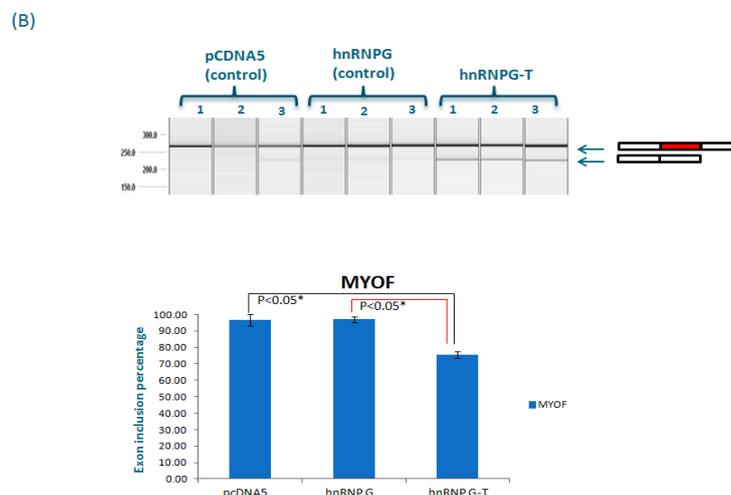
Figure 53: Role of hnRNP G-T in controlling exon inclusion of *ACOT9* and *MYOF*. A) While the percentage of exon inclusion values for the pcDNA5 and hnRNP G groups were 71.72% and 70.84% respectively, in the hnRNP G-T group this inclusion was significantly different from both control groups (*P*<0.05) and was 48.36%. The two control groups (stably expressing empty pcDNA5 or hnRNP G) were not significantly different from each other (*P*>0.05). These data show that hnRNP G-T has role to prevent exon inclusion of *ACOT9.* B) Percentage of exon inclusion in the stable cell lines expressing empty pcDNA5, hnRNP G and hnRNP G-T groups were 96.67%, 96.96% and 75.56% respectively. While the two control groups (empty pcDNA5 and hnRNP G) showed similar percentages of exon inclusion, the hnRNP G-T group was significantly different from both of these two groups (*P*<0.05). This showed that hnRNP G-T has role in exon exclusion of *MYOF* as well. All samples were performed in triplicate. The asterisk shows that the difference between two groups was significant.

Despite the high throughput results from the Canadian group showing the role of hnRNP G-T in splicing regulation of *MINK1,* in this study it was found out that the percentage of splicing inclusion of *MINK1* in stable cell lines over-expressing hnRNP G-

T were not significantly different from the control groups (expressing empty pcDNA5 and hnRNP G) (P>0.05) (figure 54-A).

Similar experiments was carried out to analyse *ENAH* splicing, also predicted to be an hnRNP G-T splicing target by the Canadian group. The percentage of splicing inclusion of *ENAH*, not only was significantly different in hnRNP G-T group from both control groups (pcDNA5 and hnRNP G) (P<0.05), but also the control groups were significantly different from each other (P<0.05). It means that both hnRNP G and hnRNP G-T have role in variable exon exclusion of *ENAH* (figure 54-B). Both hnRNP G and hnRNP G-T function as the splicing repressors for the *ENAH* exon.

(A)
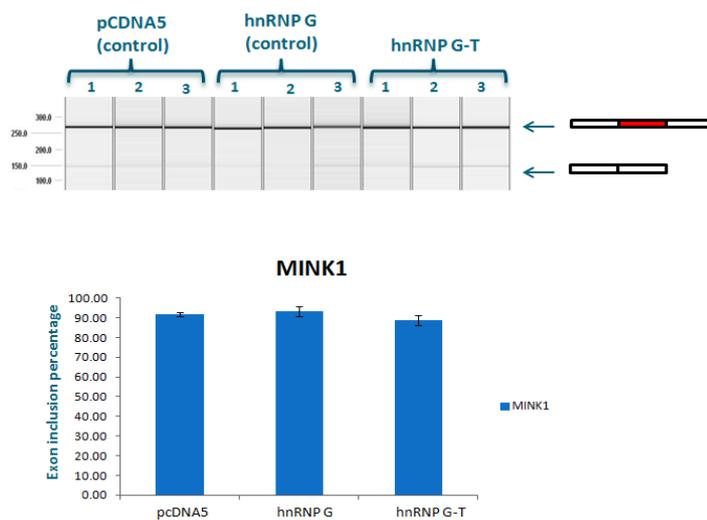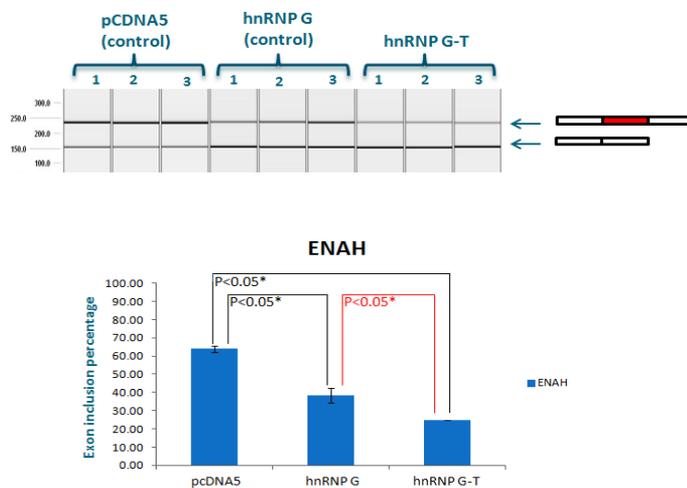
Figure 54: Role of hnRNP G-T in exon inclusion of *MINK1* and *ENAH.* A) HnRNP G-T has no role in regulation of *MINK1* exon inclusion. The percentages of exon inclusion in the stable cell lines expressing of pcDNA5, hnRNP G and hnRNP G-T were 91.73%, 93.21% and 88.93% respectively. These percentages were not significantly different from each other (*P*>0.05). B) Interestingly both hnRNP G and hnRNP G-T promote exon exclusion of the variable exon in *ENAH* (*P*<0.05). While the PSI value for this ENAH exon in the stable cell lines expressing pcDNA5 were 63.91%, the percentages in stable cell lines expressing hnRNP G and hnRNP G-T were 38.42% and 24.63% respectively. All samples were tested in triplicate. The asterisk shows that the difference between two groups was significant.

Chapter 4: Investigation into the cellular roles of hnRNP G (RBMX), hnRNP G-T (RBMXL2) and Tra2β using inducible cell lines

4.3.8   Presence of possible RNA binding sites within the detected cassette exons and their flanking introns

UCSC genome browser and RBPDB (database of RNA binding protein specificities) programmes were used to identify the presence of possible RNA binding sites within the detected alternatively spliced cassette exons and their flanking introns of target genes (*ACOT9*, *ENAH* and *MYOF)* after over-expression of hnRNP G and hnRNP G-T*.* Among the different RNA binding proteins, T-STAR, hnRNP G and some members of SR proteins such as SRSF1 and SRSF9 were detected to have potential binding sites specifically within the upstream and downstream flanking introns of detected cassette exons (figure 55). The full lists of RNA binding proteins which have potential binding sites have been shown in the Appendix B.

A) ACOT9

Upstream intron

Binding site	RNA binding protein
CUCAA	RBMY1A1
AGGAC	SRSF9
CCAU	hnRNP G/RBMX
AGGA	SRSF1
UCUAAU	KHDRBS3/T-STAR
UAUAAA	KHDRBS3/T-STAR
CAUAAA	KHDRBS3/T-STAR
GUUAAU	KHDRBS3/T-STAR
UGGA	SRSF1
UAUAAC	KHDRBS3/T-STAR
GUUAAA	KHDRBS3/T-STAR

Downstream intron

Binding site	RNA binding protein
AGGAG	SRSF9
AGCAC	SRSF9
CCAG	hnRNP G/RBMX
UUCUAACAAC	QK1
AGGA	SRSF1
UAUAAU	KHDRBS3/T-STAR
ACUAAU	KHDRBS3/T-STAR
AUUAAU	KHDRBS3/T-STAR
UAUAAA	KHDRBS3/T-STAR
AGUAAU	KHDRBS3/T-STAR
UGGA	SRSF1
ACUAAA	KHDRBS3/T-STAR
CUUAAA	KHDRBS3/T-STAR
GGUAAU	KHDRBS3/T-STAR
UCUAAC	KHDRBS3/T-STAR
AUUAAC	KHDRBS3/T-STAR
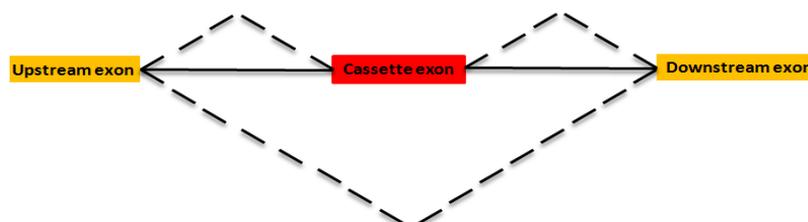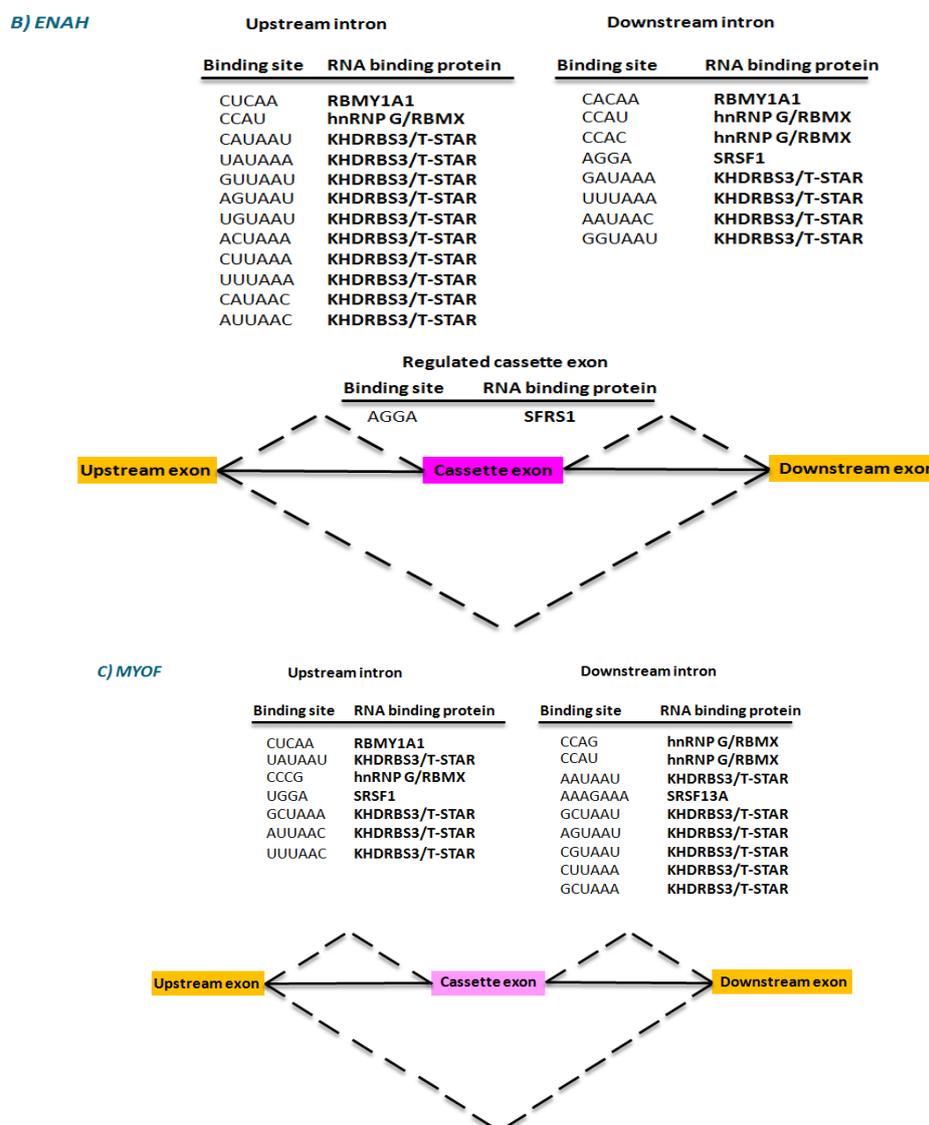
Figure 55: Presence of possible binding sites within the detected alternatively spliced cassette exons of *ACOT9*, *ENAH* and *MYOF* after over-expression of hnRNP G and hnRNP G-T. The binding sites and the names of RNA binding proteins which can bind to the cassette exons or flanking upstream and downstream introns of target genes including A) *ACOT9*, B) *ENAH*, C) *MYOF*. T-STAR, hnRNP G and some member of SR proteins are present among the RNA binding proteins which are able to bind these binding sites. The data are from RBPDB website: http://rbpdb.ccbr.utoronto.ca/ .

4.3.9   Purification of FLAG-tagged proteins (hnRNP G, hnRNP G-T and Tra2β) and their partners using immunoprecipitation

Same stable cell lines (Flp-In HEK293 cells) were used to identify proteins interacting with hnRNP G, hnRNP G-T and Tra2β. 72 hours after tetracycline induction of stably transfected Flp-In HEK293 cells, cells were harvested and lysed, followed by the immunoprecipitation of FLAG-tagged proteins (hnRNP G, hnRNP G-T and Tra2β) using α-FLAG affinity resin (for more information see Chapter2, method section). Briefly, 1.5 ml of IP +Benzonase lysis buffer was added to cell pellets followed by 20 minutes incubation on ice. Samples were then centrifuged at 4°C for 10 minutes. 30µl of the supernatant was kept as Input and rest of the supernatant was incubated with FLAG resin for an overnight step at 4°C. The FLAG tagged proteins and their interacting protein partners were pulled out by this way. Next day, samples were centrifuged at 1000rpm for 30 seconds and 30µl of the supernatant was kept and labelled as Sup. This step was followed by washing the FLAG resin with IP wash buffer and adding 2X SDS loading dye to samples. Immunoprecipitated samples were analysed by Western blotting. Monoclonal anti FLAG antibody was used as the primary antibody to probe the Western blots and analyse whether the immunoprecipitation was successful. IP lanes of FLAG tagged proteins in comparison with the control sample (pcDNA5, empty vector) showed that co-immunoprecipitations had been successful in retrieving the tagged proteins (figure 56).
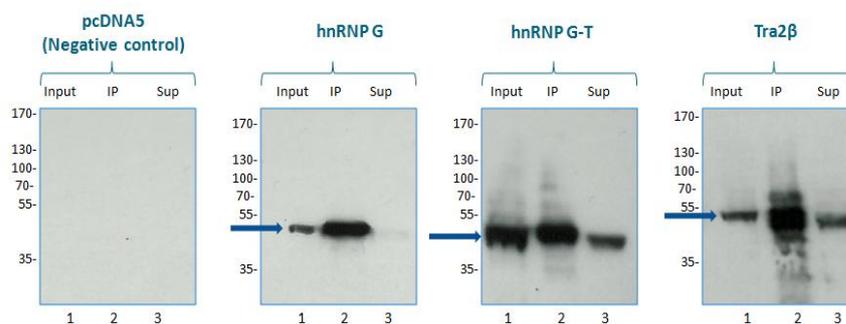
Figure 56: Immunoprecipitation of FLAG tagged proteins from stable cell lines expressing hnRNP G, hnRNP G-T and Tra2β. Western blotting was carried out to determine IP efficiency. Monoclonal anti FLAG antibody and HRP-conjugated anti-mouse IgG were used as the primary and secondary antibodies. In the above picture lanes 1, 2 and 3 show the Input, IP and Sup samples respectively. The IP lanes of FLAG tagged proteins in comparison with the control sample (cells stably transfected with pcDNA5) showed that the immunoprecipitations were successful. The size markers are in kDa.

4.3.10 Identification of candidate interacting protein partners of hnRNP G and Tra2β using mass spectrometry

After confirmation that the immunoprecipitation of the FLAG tagged proteins were successful, the IP samples were loaded onto SDS-PAGE gels and electrophoresed to identify any candidate interacting protein partners of hnRNP G, hnRNP G-T and Tra2β. This step was followed by staining the gels with Simply Blue stain for one hour. Gels were de-stained by using dH₂O overnight. Protein bands were cut and sent for mass spectrometry.

For the Tra2β experiment, no new band was detected in the IP lane in comparison with the control group (pcDNA5). However, there were some extra protein bands in the

hnRNP G IP lane (figure 57). The bands were cut and sent for analysis. Based on the sequencing results (table 31), six proteins including hnRNP G, RNA binding motif protein X-Linked-Like 1 (RBMXL1), hnRNP C, hnRNP CL1, RPL6 and hnRNP H3 were detected in hnRNP G lane by mass spectrometry. Table 31 shows information including Swiss-Prot entry name, Accession number, Gene name, Recommended Name and Nominal mass (Mr) of these proteins. Also, shown in figure 58 the peptide sequences of the detected proteins and matched peptides by mass spectrometry are in red.
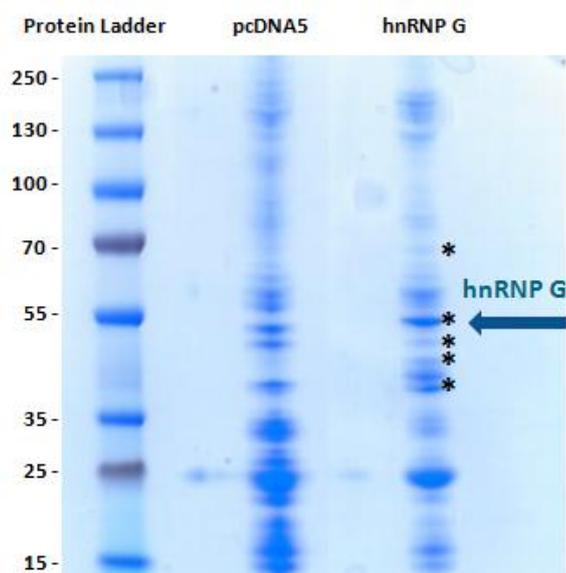
Figure 57: Simply Blue staining for detection of candidate interacting protein partners of hnRNP G. After loading the IP samples on 4-12% gradient gels and gel electrophoresis, Simply Blue safe stain was used for staining the gels. Stained gels were washed with dH$_2$O and protein bands were cut. Bands within the area highlighted by asterisks were cut and sent for sequencing by mass spectrometry. Based on mass spectrometry, hnRNP G, RNA binding motif protein X-Linked-Like 1 (RBMXL1), hnRNP C, hnRNP CL1, RPL6 and hnRNP H3 were detected. The presence of hnRNP G showed the immunoprecipitation had worked, since this is the protein immunoprecipitated.

Chapter 4: Investigation into the cellular roles of hnRNP G (RBMX), hnRNP G-T (RBMXL2) and Tra2β using inducible cell lines

	Swiss-Prot entry name	Accession number	Gene name	Recommended Name	Nominal mass(Mr)
1	RBMX_HUMAN	P38159	*RBMX*	RNA-binding motif protein, X chromosome 1	42275
2	RMXL1_HUMAN	Q96E39	*RBMXL1*	Heterogeneous nuclear ribonucleoprotein G-like 1	42173
3	HNRPC_HUMAN	P07910	*HNRNP C*	Heterogeneous nuclear ribonucleoproteins C1/C2	32004
4	HNRCL_HUMAN	O60812	*HNRNP CL1*	Heterogeneous nuclear ribonucleoprotein C-like 1	22695
5	RL6_HUMAN	Q02878	*RPL6*	60S ribosomal protein L6	32928
6	HNRH3_HUMAN	P31942	*HNRNP H3*	Heterogeneous nuclear ribonucleoprotein H3	31505

Table 31: Candidate protein partners of hnRNP G identified by FLAG immunoprecipitation. Based on mass spectrometry results, the presence of hnRNP G shows the IP has worked. In addition, some other candidate proteins including RNA binding motif protein X-linked-like 1 (RBMXL1), hnRNP C, hnRNP CL1, RPL6 and hnRNP H3 were detected. Information about these proteins such as : Swiss-Prot entry name, Accession number, Gene name, Recommended Name and Nominal mass (Mr) are from two websites: http://www.genecards.org/ and http://www.uniprot.org/

HNRNP G (RNA binding motif protein, X-linked)

Swiss-Prot entry name: RBMX_HUMAN   Accession number: P38159

MVEADRPGKLFIGGLNTETNEKALEAVFGKYGRIVEVLLMKDRETNKSRG
FAFVTFESPADAKDAARDMNGKSLDGKAIKVEQATKPSFESGRRGLPPPP
RSRGPPRGLRGGRGGSGGTRGPPSRGGHMDDGGYSMNFTMSSSRGPLPVK
RGPPPRSGGPPPKRSAPSGPVRSSSGLGGRAPVSRGRDGYGGPPRREPLP
SRRDVYLSPRDDGYSTKDSYSSREYPSSRDTRDYAPPPRDYTYRDYGHSS
SRDDYPSRGYSDRDGYGRDRDYSDHPSGGSYRDSYESYGNSRSAPPTRGP
PPSYGGSSRYDDYSSSRDGYGGSRDSYSSSRSDLYSSGRDRVGRQERGLP
PSMERGYPPPRDSYSSSSRGAPRGGGRGGSRSDRGGGRSRY

Heterogeneous nuclear ribonucleoprotein G-like 1 (RBMXL1)

Swiss-Prot entry name: RMXL1_HUMAN   Accession number: Q96E39

MVEADRPGKLFIGGLNTETNEKALETVFGKYGRIVEVLLIKDRETNKSRG
FAFVTFESPADAKDAARDMNGKSLDGKAIKVEQATKPSFERGRHGPPPPP
RSRGPPRGFGAGRGGSGGTRGPPSRGGHMDDGGYSMNFNMSSSRGPLPVK
RGPPPRSGGPSPKRSAPSGLVRSSSGMGGRAPLSRGRDSYGGPPRREPLP
SRRDVYLSPRDDGYSTKDSYSSRDYPSSRDTRDYAPPPRDYTYRDYGHSS
SRDDYPSRGYGDRDGYGRDRDYSDHPSGGSYRDSYESYGNSRSAPLTRGP
PPSYGGSSRYDDYSSSRDGYGGSRDSYSSSRSDLYSSCDRVGRQERGLPP
SVERGYPSSRDSYSSSSRGAPRGAGPGGSRSDRGGGRSRY

HnRNP 2H9B (60S ribosomal protein L6)

Swiss-Prot entry name: RL6_HUMAN   Accession number: Q02878

MDYQGRSTGEAFVQFASKEIAENALGKHKERIGHRYIEIFRSSRSEIKGF
YDPPRRLLGQRPGPYDRPIGGRGGYYGAGRGSMYDRMRRGGDGYDGGYGG
FDDYGGYNNYGYGNDGFDDRMRDGRGMGGHGYGGAGDASSGFHGGHFVHM
RGLPFRATENDIANFFSPLNPIRVHIDIGADGRATGEADVEFVTHEDAVA
AMSKDKNNMQHRYIELFLNSTPGGGSGMGGSGMGGYGRDGMDNQGGYGSV
GRMGMGNNYSGGYGTPDGLGGYGRGGGGSGGYYGQGGMSGGGWRGMY

Heterogeneous nuclear ribonucleoprotein H3 (DNA-binding protein TAXREB107)

Swiss-Prot entry name: HNRH3_HUMAN   Accession number: P31942

MAGEKVEKPDTKEKKPEAKKVDAGGKVKKGNLKAKKPKKGKPHCSRNPVL
VRGIGRYSRSAMYSRKAMYKRKYSAAKSKVEKKKKEKVLATVTKPVGGDK
NGGTRVVKLRKMPRYYPTEDVPRKLLSHGKKPFSQHVRKLRASITPGTIL
IILTGRRRGKNWVVFLKQLASGLLLVTGPLVLNRVPLRRTHQKFVIATST
KIDISNVKIPKHLTDAYFKKKKLRKPRHQEGEIFDTEKEKYEITEQRKID
QKAVDSQILPKIKAIPQLQGYLRSVFALTNGIYPHKLVF

Figure 58: MASCOT search results for candidate interacting protein partners of

hnRNP G. Swiss-Prot entry names and Accession numbers have been shown for all samples. In addition, matched peptides identified by mass spectrometry are shown in red. The peptide coverages on identified proteins were 27%, 6%, 10% and 16% for hnRNP G, RBMXL1, 60S ribosomal protein L6 and heterogenous nuclear ribonucleoprotein H3 respectively.

4.3.11 Identification of candidate interacting protein partners of hnRNP G-T using mass spectrometry

Similar to the experiments carried out for hnRNP G, the PAGE gel was stained with Simply Blue safe stain to analyse the hnRNP G-T immunoprecipitation. Some extra bands were detected in the IP lane of hnRNP G-T in comparison with the control group (pcDNA5 (figure 59). These bands were cut from the gel and sent for analysis. Based on the sequencing results (table 32), ten proteins were detected in the hnRNP G-T lane by mass spectrometry including hnRNP G-T, RNA binding motif protein X-Linked-Like 1 (RBMXL1), H2AFY, hnRNP C, hnRNP CL1, RPS6, RPL7, RPL7A, RPLP2 and RPS20. Table 31 shows some information for these candidate protein interaction partners including Swiss-Prot entry name, Accession number, Gene name, Recommended Name and Nominal mass (Mr) of these proteins. Also, in figure 60 the sequences of the detected proteins have been shown and matched peptides by mass spectrometry are in red.
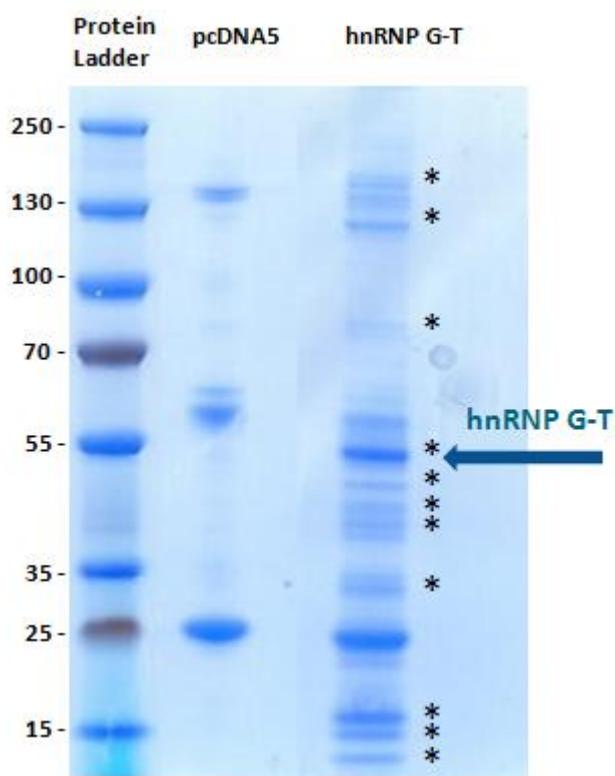
Figure 59: Simply Blue staining for detection of candidate interacting protein partners of hnRNP G-T. After loading IP samples on the 4-12% gradient gels and gel electrophoresis, Simply Blue safe stain was used for staining the gels. Stained gels were washed with dH_2O and the protein bands were cut and sent for sequencing. The bands highlighted by asterisks were cut out of the gel and sent for the sequencing by mass spectrometry. Based on mass spectrometry results, ten proteins were detected including: hnRNP G-T, RNA binding motif protein X-Linked-Like 1 (RBMXL1), H2AFY, hnRNP C, hnRNP CL1, RPS6, RPL7, RPL7A, RPLP2 and RPS20. The presence of hnRNP G-T in the IP lane showed that the immunoprecipitation had worked.

Chapter 4: Investigation into the cellular roles of hnRNP G (RBMX), hnRNP G-T (RBMXL2) and Tra2β using inducible cell lines

	Swiss-Prot entry name	Accession number	Gene name	Recommended Name	Nominal mass(Mr)
1	HNRGT_HUMAN	O75526	RBMXL2	RNA-binding motif protein, X-linked-like-2	42929
2	RBMXL_HUMAN	Q96E39	RBMXL1	Heterogeneous nuclear ribonucleoprotein G-like 1	42173
3	HNRPC_HUMAN	P07910	HNRNP C	Heterogeneous nuclear ribonucleoproteins C1/C2	32004
4	HNRCL_HUMAN	O60812	HNRNP CL1	Heterogeneous nuclear ribonucleoprotein C-like 1	22695
5	H2AY_HUMAN	O75367	H2AFY	Core histone macro-H2A.1	39748
6	RS6_HUMAN	P62753	RPS6	40S ribosomal protein S6	28842
7	RL7_HUMAN	P18124	RPL7	60S ribosomal protein L7	29221
8	RL7A_HUMAN	P62424	RPL7A	60S ribosomal protein L7a	30148
9	RLA2_HUMAN	P05387	RPLP2	60S acidic ribosomal protein P2	11658
10	RS20_HUMAN	P60866	RPS20	40S ribosomal protein S20	13478

Table 32: Candidate protein partners of hnRNP G-T identified by FLAG immunoprecipitation. Based on the mass spectrometry results, the presence of hnRNP G shows the IP has worked. In addition some other candidate proteins were detected including RNA binding motif protein X-Linked-Like 1 (RBMXL1), hnRNP C, hnRNP CL1, RPL6 and hnRNP H3. Information about these proteins such as : Swiss-Prot entry name, Accession number, Gene name, Recommended Name and Nominal mass (Mr) are from two websites: http://www.genecards.org/ and http://www.uniprot.org/

HNRNP G-T (RNA-binding motif protein, X-linked-like-2)

Swiss-Prot entry name: HNRGT_HUMAN   Accession number: O75526

MVEADRPGKLFIGGLNLETDEKALEAEFGKYGRIVEVLLMKDRETNKSRGFAFVTFESPADA
KAAARDMNGKSLDGKAIKVAQATKPAFESSRRGPPPPPRSRGRPRFLRGTRGGGGGPRRSPSR
GGPDDDGGYAADFDLRPSRAPMPMKRGPPPPRRVGPPPKRAAPSGPARSSGGGMRGRALAVRG
RDGYSGPPRREPLPPRRDPYLGPRDEGYSSRDGYSSRDYREPRGFAPSPGEYTHRDYGHSSV
RDDCPLRGYSDRDGYGGRDRDYGDHLSRGSHREPFESYGELRGAAPGRGTPPSYGGGGRYEE
YRGYSPDAYSGGRDSYSSSYGRSDRYSRGRHRVGRPDRGLSLSMERGCPPQRDSYSRSGCRV
PRGGGRLGGRLERGGGRSRY

Heterogeneous nuclear ribonucleoprotein G-like 1 (RBMXL1)

Swiss-Prot entry name: RBMXL_HUMAN   Accession number: Q96E39

MVEADRPGKLFIGGLNTETNEKALETVFGKYGRIVEVLLIKDRETNKSRGFAFVTFESPADA
KDAARDMNGKSLDGKAIKVEQATKPSFERGRHGPPPPPRSRGPPRGFGAGRGGSGGTRGPPS
RGGHMDDGGYSMNFNMSSSRGPLPVKRGPPPRSGGPSPKRSAPSGLVRSSSGMGGRAPLSRG
RDSYGGPPRREPLPSRRDVYLSPRDDGYSTKDSYSSRDYPSSRDTRDYAPPPRDYTYRDYGH
SSSRDDYPSRGYGDRDGYGRDRDYSDHPSGGSYRDSYESYGNSRSAPLTRGPPPSYGGSSRY
DDYSSSRDGYGGSRDSYSSSRSDLYSSCDRVGRQERGLPPSVERGYPSSRDSYSSSSRGAPR
GAGPGGSRSDRGGGRSRY

HNRNP C (heterogeneous nuclear ribonucleoproteins C1/C2)

Swiss-Prot entry name: HNRPC_HUMAN   Accession number: P07910

MASNVTNKTDPRSMNSRVFIGNLNTLVVKKSDVEAIFSKYGKIVGCSVHKGFAFVQYVNERN
ARAAVAGEDGRMIAGQVLDINLAAEPKVNRGKAGVKRSAAEMYGSSFDLDYDFQRDYYDRMY
SYPARVPPPPPIARAVVPSKRQRVSGNTSRRGKSGFNSKSGQRGSSKSGKLKGDDLQAIKKE
LTQIKQKVDSLLENLEKMEKEQSKQAVEMKNDKSEEEQSSSSVKKDETNVKMESEGGADDSA
EEGDLLDDDDNEDRGDDQLELIKDDEKEAEEGEDDRDSANGG

HNRNP CL1 (Heterogeneous nuclear ribonucleoprotein C-like 1)

Swiss-Prot entry name: HNRCL_HUMAN   Accession number: O60812

MASNVTNKMDPHSMNSRVFIGNLNTLVVKKSDVEAIFSKYGKIAGCSVHKGFAFVQYDKEKN
ARAAVAGEDGRMIASQVVDINLAAEPKVNRGNAGVKRSAAEMYGSSFDLDYGFQRDYYDGMY
SFPARVPPPPPIALAVVPSKRQRLSGNTSRRGKSGFNSKSGKRGSSKSGKLKGDDLQAIKQE
LTQIKQKVDSLLENLEKKKKK

Core histone macro-H2A.1

Swiss-Prot entry name: H2AY_HUMAN   Accession number: O75367

MSSRGGKKKSTKTSRSAKAGVIFPVGRMLRYIKKGHPKYRIGVGAPVYMAAVLEYLTAEIILE
LAGNAARDNKKGRVTPRHILLAVANDEELNQLLKGVTIASGGVLPNIHPELLAKKRGSKGKL
EAIITPPPAKKAKSPSQKKPVSKKAGGKKGARKSKKKQGEVSKAASADSTTEGTPADGFTVL
STKSLFLGQKLNLIHSEISNLAGFEVEAIINPTNADIDPKDDLGNTLEKKGGKEFVEAVLEL
RKKNGPLEVAGAAVSAGHGLPAKFVIHCNSPVWGADKCEELLEKTVKNCLALADDKKLKSIA
FPSIGSGRNGFPKQTAAQLILKAISSYFVSTMSSSIKTVYFVLFDSESIGIYVQEMAKLDAN

40S ribosomal protein S6

Swiss-Prot entry name: RS6_HUMAN   Accession number: P62753

MKLNISFPATGCQKLIEVDDERKLRTFYEKRMATEVAADALGEEWKGYVVRISGGNDKXGFP
MKQGVLTHGRVRLLLSKGHSCYRPRRTGERKRKSVRGCIVDANLSVLNLVIVKKGEKDIPGL
TDTTVPRRLGPKRASRIRKLFNLSKEDDVRECVVRKPLNKEGKRPRTKAPKIQRLVTPRVLQ
HKRRRIALKKQRTKKNKEEAAEYAKLLAKRMKQAKEKRQEQIAKRRRLSSLRASTSKSESSQ
K

60S ribosomal protein L7

Swiss-Prot entry name: RL7_HUMAN   Accession number: P18124

MEGVEEKKKEVPAVPETLKKKRRNFAELKIKRLRKKFAQKMLRKARRKLIYEKAKHYHKEYR
QMYRTEIRMARMARKAGNFYVPAEPKLAFVIRIRGINGVSPKVRKVLQLLRLRQIFNGTFVK
LNKASINMLRIVEPYIAWGYPNLKSVNELIYKLGYGKINKKRIALTDNALIARSLGKYGIIC
MEDLIHEIYTVGKRFKEANNFLWPFKLSSPRGGMKKKTTHFVEGGDAGNREDQINRLIRRMN

60S ribosomal protein L7a

Swiss-Prot entry name: RL7A_HUMAN   Accession number: P62424

MPKGKKAKGKKVAPAPAVVKKQEAKKVVNPLFEKRPKNFGIGQDIQPKRDLTRFVKWPRYIR
LQRQRAILYKRLKVPPAINQFTQALDRQTATQLLKLAHKYRPETKQEKKQRLLARAEKKAAG
KGDVPTKRPPVLRAGVNTVTTLVENKKAQLVVIAHDVDPIELVVFLPALCRKMGVPYCIIKG
KARLGRLVHRKTCTTVAFTQVNSEDKGALAKLVEAIRTNYNDRYDEIRRHWGGNVLGPKSVA
RIAKLEKAKAKELATKLG

60S acidic ribosomal protein P2

Swiss-Prot entry name: RLA2_HUMAN   Accession number: P05387

MRYVASYLLAALGGNSSPSAKDIKKILDSVGIEADDDRLNKVISELNGKNIEDVIAQGIGKL
ASVPAGGAVAVSAAPGSAAPAAGSAPAAAEEKKDEKKEESEESDDDMGFGLFD

40S ribosomal protein S20

Swiss-Prot entry name: RS20_HUMAN   Accession number: P60866

MAFKDTGKTPVEPEVAIHRIRITLTSRNVKSLEKVCADLIRGAKEKNLKVKGPVRMPTKTLR
ITTRKTPCGEGSKTWDRFQMRIHKRLIDLHSPSEIVKQITSISIEPGVEVEVTIADA

Figure 60: MASCOT search results for candidate interacting protein partners of hnRNP G-T. Swiss-Prot entry names and Accession numbers have been shown for all samples. In addition, matched peptides identified by mass spectrometry are shown in red. The percentage of the identified proteins that were covered by the peptide sequences were 35%, 6%, 15%, 12%, 10%, 14%, 12%, 12%, 39% and 19% for hnRNP G-T, RBMXL1, hnRNP C, hnRNP CL1, core histone maco-H2A.1, 40S ribosomal protein S6, 60S ribosomal protein L7, 60S ribosomal protein L7a, 60S acidic ribosomal protein P2 and 40S ribosomal protein S20 respectively.

4.3.12 Confirmation of the presence of hnRNP C in the hnRNP G and hnRNP G-T IP samples

HnRNP C was a strong candidate since it was detected by IP of both hnRNP G and hnRNP G-T. After identification of proteins from the SDS-PAGE gels using mass spectrometry, Western blotting was used to further test the presence of hnRNP C in both the hnRNP G and hnRNP G-T IP samples. Western blots of all samples were probed with anti hnRNP C1/C2 (4F4) antibody. The presence of hnRNP C protein in both of the IP samples from hnRNP G-FLAG and hnRNP G-T-FLAG were detected, but not for the empty pcDNA5 stable cell line (figure 61). Mouse hnRNP C1/C2 (4F4) antibody and anti-mouse-HRP were used as the primary and secondary antibodies

respectively. These results mean that endogenous hnRNP C interacts with both hnRNP G-FLAG and hnRNP G-T-FLAG proteins.
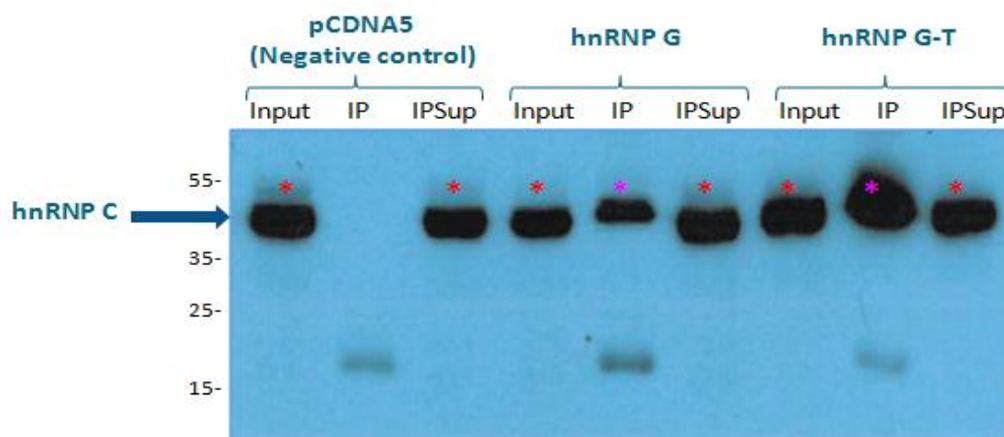
Figure 61: Confirmation of the presence of hnRNP C in the IP lanes of hnRNP G-FLAG and hnRNP G-T-FLAG samples. The presence of hnRNP C was confirmed in the input and IPsup lanes of all samples including the negative control (pcDNA5) (red asterisks). HnRNP C protein was also present in IP (immunoprecipitated) lanes of both hnRNP G and hnRNP G-T samples (pink asterisks) but not in the IP lane of pcDNA5. Mouse hnRNP C1/C2 (4F4) antibody and anti-mouse-HRP were used as the primary and secondary antibodies respectively.

4.4   Summary and discussion:

As a tool to investigate the functions of hnRNP G, hnRNP G-T and Tra2β, their open reading frames were cloned into the pcDNA5 vector so they produced FLAG tagged fusion proteins. After transfection of constructs into Flp-In HEK293 cells, hygromycin B was added to the cells for selection of transfected cells. The stable transfected cells had hygromycin resistance gene, therefore, they would survive after adding hygromycin to the media. Also, expression of the gene of interest was controlled by the tetracycline promoter. Therefore, a high level of gene expression would be achieved by adding tetracycline to the cell culture medium.

Interestingly, in this study it was found out that after induction of *RBMXL2* (name of the gene encoding hnRNP G-T protein), cells undergo morphological changes. These cells became round firstly and then released within the media. Since it has been shown that *RBMXL2* gene is derived from transcript retrotransposition of *RBMX* (name of the gene encoding hnRNP G protein) (Elliott *et al.*, 2000b) the cells expressing FLAG tagged hnRNP G were used as well as an extra control group expressing empty pcDNA5 to test morphological changes.

It was found out that morphological changes only happened within the cells expressing hnRNP G-T. At longer stages after tetracycline induction the observed morphological changes were more obvious. In addition, cell growth rates were less in both groups expressing hnRNP G and hnRNP G-T in comparison with cells expressing the empty vector (pcDNA5). To address the basis of these morphological changes, firstly an MTT assay was used to check the cell proliferation rate. Stable cells over-expressing hnRNP G-T protein had a lower and significantly different cell proliferation rate in comparison with both control groups (pcDNA5 and hnRNP G) ($P<0.05$). In addition, cells expressing hnRNP G had lower cell growth rate in comparison with the control group (pcDNA5)

(*P*<0.05). No such effects were seen for Sam68 or T-STAR suggesting these changes might not be only because of nonspecific overproduction of proteins.

To further examine the roles of hnRNP G and hnRNP G-T in the reduction of cell growth rate, percentage of apoptosis and DNA damage was analysed within the cells expressing hnRNP G and hnRNP G-T in comparison with the control group (cells expressing empty vector, pcDNA5).

BrdU (an analogue of thymidine) incorporation was measured in parallel to test the cell proliferation rate since it detects cells in S phase. The percentage of cells going through S phase in the hnRNP G-T group (cells expressing hnRNP G-T) was lower and significantly different in comparison with both control groups (expressing empty pcDNA5 and hnRNP G) (*P*<0.05). The control groups were significantly different from each other as well (over-expression of hnRNP G also slows cell growth, *P*<0.05). These results showed that over-expression of hnRNP G and hnRNP G-T each affects cell proliferation and growth. HnRNP G-T had stronger effect on this reduction since the percentage of cells in S phase in this group was significantly lower than percentage of S phase cells in the hnRNP G group (*P*<0.05).

Since cells expressing hnRNP G-T firstly became round and then released within the media I was important to test whether expression of hnRNP G-T might also lead to cell death. The percentage of apoptotic cells was monitored based on cleavage of PARP into two smaller fragments. Binding of a fluorescent antibody against the 89kDa PARP fragment was used as a marker of apoptosis (Kaufmann *et al.*, 1993; Lazebnik *et al.*, 1994; Nicholson *et al.*, 1995; Los *et al.*, 1997; Los *et al.*, 1999). If cells were apoptotic they would produce a smaller fragment which would be recognized by this α- PARP antibody.

The percentage of apoptotic cells in the group of cells stably expressing hnRNP G-T was less than two percent of total cell population but still was higher and significantly different in comparison with both control groups (pcDNA5 and hnRNP G) ($P<0.05$). In addition, control groups (pcDNA5 and hnRNP G) were significantly different from each other ($P<0.05$). These results showed the over expression of hnRNP G and hnRNP G-T do lead to a higher percentage of cell apoptosis but this was not a very strong effect.

The percentage of DNA damage was also tetsted after hnRNP G-T expression. Phosphorylated histone H2AX (phosphorylation of C-terminal serine 139) is a biomarker to show cell response to DNA damage (Valdiglesias et al., 2013). Phosphoinositide3-kinase-related protein kinases (PIKKs) including DNA-dependent protein kinase (DNA-PK), ataxia teleangiectasia mutated (ATM) and ATR (ATM and Rad3-related) can phosphorylate H2AX. In response to DNA double-strand break, ATM is one of the important kinases involved in H2AX phosphorylation (Kastan and Lim, 2000; Bakkenist and Kastan, 2003). Also, ATR plays role in H2AX phosphorylation once replication fork arrest or single-stranded DNA breaks occur (Ward and Chen, 2001; Ward et al., 2004). DNA-PK also phosphorylates H2AX in response to apoptosis DNA fragmentation and hypertonic conditions (Reitsema et al., 2005; Mukherjee et al., 2006). The presence of phosphorylated H2AX leads to localisation and accumulation of DNA damage repair proteins at the site of damage (Kouzarides, 2007; Taverna *et al.*, 2007). Therefore, if the cells expressing hnRNP G-T had DNA damage this would be detected using the kit antibody against H2AX. According to the percentage of DNA damage, cells expressing hnRNP G-T had less DNA damage than the other two lines of stable cells investigated in this Chapter (expressing pcDNA5 and hnRNP G) ($P<0.05$).

Based on a low percentage of apoptosis (less than 2%) and higher DNA integrity compared to control groups, these data suggest that hnRNP G-T expression does not lead to significant cell death. The reason for the small increase in cell death could be

because the cells expressing hnRNP G-T protein are released within the media, and these floating cells perhaps are more vulnerable to environmental factors leading to a higher percentage of cell death in comparison with the control groups. Furthermore, it seems that hnRNP G-T protein not only has no adverse effect on DNA stability, but also acts to reduce the amount of the DNA damage. Interestingly it has been shown that hnRNP G plays an important role in protection of DNA ends from nucleases (Shin et al., 2007).

The percentage of live cells was measured using cell viability analyzer as another complementary experiment to test if the cells expressing hnRNP G-T were viable. Cell viability analyzer machine used trypan blue to detect percentage of the dead cells. Viable cells have intact cell membrane which is not penetrable to dye and they will not be coloured. Only dead cells will absorb the dye when trypan blue is able to pass through the cell membrane.

Although the hnRNP G-T group was significantly different from the control groups, still around 94% of cells were viable. These results were consistent with previous findings of this study. It seems that when cells were circular and released within the media they were still alive.

Regarding to the cell morphological changes in hnRNP G-T group in comparison with control groups, it seemed that these cells were losing their cell-cell adhesions and also their attachments to the tissue culture surface after gene induction. These morphological changes were reminiscent of those described for epithelial–mesenchymal transitions (EMTs), an important event during embryonic development (Thiery, 2002).

Chapter 4: Investigation into the cellular roles of hnRNP G (RBMX), hnRNP G-T (RBMXL2) and Tra2β using inducible cell lines

Unlike mesenchymal cells, epithelial cells are not able to carry out migration and invasion. Epithelial-mesenchymal transformation (EMT) can lead to cell movement both in development and in cancer (Hay, 1995; Thiery, 2002; Yang et al., 2004). Since during EMT cells can migrate to other organs to be involved in organ formation, this transition plays an important role in development (Thiery, 2002). Also this phenomenon occurs in cancer metastasis as well. Metastatic cells can migrate from a tumour and invade other tissues (Yang et al., 2004).

Since hnRNP G-T has a role in alternative splicing regulation, there was a question ''are genes involved in EMT alternatively spliced by hnRNP G-T?''. To answer this question, the differences were tested between the alternative splicing regulation of some genes involved in cell adhesion and attachment in comparison with the control groups (cells expressing the FLAG tagged hnRNP G and empty vector, pcDNA5). Following Dr Julian P Venables' suggestion (Institute of Genetic Medicine, Newcastle University), after induction and harvesting the cells stably expressing pcDNA5, hnRNP G and hnRNP G-T, the RNA samples were extracted and cDNA synthesis was carried out to make cDNA from RNA. Samples were then sent to a lab in Canada (Venables *et al.*, 2009). Alternative splicing was monitored between three groups (pcDNA5, hnRNP G and hnRNP G-T) in genes associated with EMT using a high-throughput RT-PCR screen. Data was then analyzed by Dr Venables and four EMT target genes including: *MINK1* (misshapen-Like Kinase 1), *ACOT9* (acyl-CoA thioesterase 9), *ENAH* (enabled homolog) and *MOYF* (myoferlin) were detected to be alternatively spliced. PSI (percentage of splicing inclusion) was different in hnRNP G-T group in comparison with the control groups.

While MINK1 protein has important roles in cell morphology, motility, cytoskeletal organization and cytokinesis (Hu et al., 2004; Hyodo et al., 2012), the physiological roles of the ACOT proteins still need to be further investigated (Ohtomo et al., 2013).

ACOT9 belongs to the Acyl-CoA thioesterases (ACOTs) family which are involved in fatty acid metabolism (catalyzing the fatty acyl-CoA thioesters hydrolysis to coenzyme A and fatty acid). *ACOT9* is encoded by the X chromosome (Kirkby et al., 2010). This enzyme is localized within the mitochondria and associated with a mitochondrial disorder called MEHMO syndrome (Mental retardation, Epileptic seizures, Hypogenitalism, Microcephaly and Obesity) (Scharfe *et al.*, 2009).

The Enah (Enabled homolog) protein has an essential role in neural development (Gertler *et al.*, 1990; Gertler *et al.*, 1995). This protein belongs to Ena/VASP family which also include two other members: VASP and EVL (Krause *et al.*, 2002; Kwiatkowski *et al.*, 2003). This family is involved in linking the actin cytoskeleton remodelling to signal transduction (Urbanelli *et al.*, 2006).

The other target gene detected by the results was *MOYF* (myoferlin). Myoferlin belongs to the ferlin family (Xiong et al., 2013). Members of this family (myoferlin, otoferlin and dysferlin) are involved in calcium-mediated membrane fusion (Davis *et al.*, 2000). Also they have important roles in plasma membrane repair (Xiong et al., 2013). Levels of both myoferlin mRNA and protein are up-regulated in myofiber damage (Demonbreun *et al.*, 2010). Myoferlin protein is expressed in skeletal and cardiac muscle and has association with the plasma membrane (Davis *et al.*, 2000). Furthermore, myoferlin interactions with other proteins may have a role in membrane fusion for rearrangements of the cytoskeleton where cells have connections with each other (Doherty *et al.*, 2008). Therefore, all of these hnRNP G and hnRNP G-T target genes are involved in either plasma membrane or cytoskeleton rearrangement.

To investigate more about the role of hnRNP G-T in alternative splicing regulation of some of genes involved in EMT, the next step was to reproduce the results in our lab. Cells were induced with tetracycline, harvested and splicing analysed by RNA

extraction and cDNA synthesis. PCR reactions were set up and the percentages of splicing inclusion (PSI) were measured. The PSI of *ACOT9* and *MYOF* in the hnRNP G-T group were significantly different from the control groups (expressing pcDNA5 and hnRNP G) (*P*<0.05). This means that expression of hnRNP G-T has role in exon exclusion of these two genes. This role could be direct or indirect. It could be that expressing hnRNP G-T affects the cells, which then affects the splicing pattern. HnRNP G-T might be acting as a direct repressor – less exon is included in each of the three replicates. In the cases of *ACOT9* and *MYOF,* the results are specific to G-T (no effect was seen for G).

Although the results from the Canadian group predicted a role of hnRNP G-T in splicing regulation of *MINK1,* after reproducing the results it was found out that the percentage of splicing inclusion of *MINK1* in hnRNP G-T group was not significantly different from the control groups (expressing pcDNA5 and hnRNP G) (*P*>0.05). Interestingly, the percentage of splicing inclusion of *ENAH*, not only was significantly different in the hnRNP G-T group from both control groups (expressing pcDNA5 and hnRNP G) (*P*<0.05), but also the control groups were significantly different from each other as well (*P*<0.05). This suggests that both hnRNP G and hnRNP G-T have role in exon exclusion of *ENAH* target gene.

In summary, these data suggest that hnRNP G-T protein has important roles in regulation of alternative splicing of some genes involved in EMT including *ACOT9*, *ENAH* and *MOYF*. In addition, splicing of *ENAH* can be regulated by hnRNP G which means that this gene can be considered as a common target for both hnRNP G and hnRNP G-T. In future work it will be interesting to test whether these genes are direct or indirect targets of hnRNP G-T and hnRNP G. In addition, further steps can be generation of minigenes to check the effects on splicing regulation of RRM mutations, mapping binding sites and looking for more targets to investigate more about hnRNP G-T roles in splicing regulation.

Chapter 4: Investigation into the cellular roles of hnRNP G (RBMX), hnRNP G-T (RBMXL2) and Tra2β using inducible cell lines

One of the main aims of this project was to investigate more about new candidates as the protein partners for these RNA binding proteins (hnRNP G and hnRNP G-T and Tra2β). After confirming the efficiency of immunoprecipitation of the FLAG tagged proteins, IP samples were loaded on SDS-PAGE gels and electrophoresed. New bands were not detected in the Tra2β IP lane in comparison with the control group (pcDNA5). However, there were some extra protein bands in both the hnRNP G and hnRNP G-T IP lanes. Staining the gels with Simply Blue was followed by cutting the bands and sending them for mass spectrometry.

As a result, five and nine candidates were detected respectively as the new candidate protein partners for hnRNP G (RBMXL1, hnRNP C, hnRNP CL1, RPL6, hnRNP H3) and hnRNP G-T (RBMXL1, H2AFY , hnRNP C, hnRNP CL1, RPS6, RPL7, RPL7A, RPLP2, RPS20). Among these proteins RBMXL1, hnRNP C and hnRNP CL1 were common interacting proteins for both hnRNP G and hnRNP G-T.

HnRNP C belongs to the family of heterogeneous nuclear ribonucleoproteins . This protein has a nuclear localization within the cell but in tumour cells it can be cytoplasmic as well (Christian *et al.*, 2008). HnRNP C is important in splicing regulation, covering cryptic splice sites and mRNA stability (Dreyfuss *et al.*, 1993; Rajagopalan *et al.*, 1998; Shetty, 2005; Zarnack *et al.*, 2013). HnRNP C binding close to the splice site can lead to repression of alternative exon selection (Konig *et al.*, 2010). The role of hnRNP C in repression of cryptic exons seems to be associated with evolution and disease (Zarnack *et al.*, 2013).

In tumours, hnRNP C expression has the potential to be used as a prognostic marker (Sun *et al.*, 2007; Zhang *et al.*, 2011). HnRNP G is over-expressed in breast cancer and hepatocellular carcinoma (Sun *et al.*, 2007; Blume *et al.*, 2010). It has also been shown

that repression of hnRNP C expression can lead to tumour growth suppression (Hossain *et al.*, 2007; Meng *et al.*, 2008).

According to the important roles of hnRNP C and also considering this protein was detected as a common interacting partner for both hnRNP G and hnRNP G-T, the presence of hnRNP C was tested in IP samples of hnRNP G and hnRNP G-T by Western blot. In this study it was found out that hnRNP G and hnRNP G-T paly role as the repressor in splicing regulation of some target genes. Also, hnRNP C plays important roles in repression of alternative exon selection (Konig *et al.*, 2010). I was interested to test whether these proteins could interact with each other or not. Finally, it was confirmed that hnRNP C interacts with both hnRNP G and hnRNP G-T proteins. In future work it will be useful to test whether these proteins have interactions at the endogenous level of expression of hnRNP G and hnRNP G-T as well. In addition, it should be investigated that whether these proteins are involved in alternative splicing regulation of common target genes.

In summary, in this Chapter in addition to introducing the new detected interaction partners of hnRNP G and hnRNP G-T, some unique and important roles of hnRNP G and hnRNP G-T were described in alternative splicing regulation of some target genes. In vivo hnRNP G-T replaces hnRNP G expression when hnRNP G protein is inactivated after XY body formation during meiosis. However, the data in this study suggest some specific and unique roles which belong only to hnRNP G-T such as cell morphological changes after over-expression of this protein.

Chapter 5

5   Conclusion and Future work:

The main aim of this project was investigation of possible new interacting protein partners of the RNA binding proteins T-STAR, wild type and P439R Sam68, hnRNP G, hnRNP G-T and Tra2β. All of these proteins play important roles within the cells and almost all of them are involved in some kinds of disease including cancer, SMA and infertility. Therefore, understanding their roles and interaction networks may be a good step towards the finding of new treatments for these diseases. Regulating gene expression at the post-transcriptional level including mRNA processing has the potential to be considered as a future tool for cancer therapy. One of the procedures involved in mRNA processing is alternative splicing. During this procedure RNA binding proteins play important roles to regulate gene expression and production of protein diversity from one pre-mRNA. The main achievements of this project are listed below:

5.1   Establishment of stable cell lines expressing FLAG tagged proteins

The open reading frames of the RNA binding proteins *T-STAR*, wild type and P439R *Sam68*, *hnRNP G*, *hnRNP G-T* and *Tra2β* were cloned in to the pcDNA5 vector. This vector encodes a FLAG epitope. These constructs were then transfected into Flp-In HEK293 cells for integration into the genome. Hygromycin B was also used to select the transfected cells. The stable transfected cells had hygromycin resistance gene, therefore, they would survive after adding hygromycin to the media. In addition, tet repressor was expressed in these cells. Expression of the genes of interest was controlled by the tetracycline promoter. Therefore, a high level of gene expression would be achieved by adding tetracycline to the cell culture medium. After making

stable cell lines and inducing gene expression by using tetracycline antibiotic it was possible to express the protein of interest.

5.2   Confirmation of Wild type and P439R Sam68 localizations

These stable cell lines were firstly used for cell biological investigation of the localization of wild type and P439R mutant Sam68 after induction in the Flp-In HEK293 cells. Wild type Sam68 was nuclear and P439R mutant Sam68 protein was completely cytoplasmic. These results were similar to the previous study carried out by Reddy et al in HeLa cells (Reddy, 2000). This means that because of a point mutation (P439R, changing proline to arginine) in the tyrosine-rich C-terminal domain of Sam68, localization of this Sam68 protein will be changed, and as a result Sam68 protein will be completely cytoplasmic.

5.3   Investigation of RNA-protein interaction networks of wild type and P439R Sam68

By generating stable cell lines expressing wild type and P439R FLAG-tagged Sam68 which are nuclear and cytoplasmic protein isoforms respectively, more information regarding the interaction networks of Sam68 within the nucleus and cytoplasm were obtained. FLAG tagged proteins (wild type and P439R Sam68) and their partners were purified using immunoprecipitation and identified using mass spectrometry.

5.3.1   Detection of Sam68 and RHA interaction

In this project the interactions of wild type and P439R Sam68 with RHA were detected after Sam68 protein induction. Previously, an interaction between Sam68 (wild type) and RHA was already known (Reddy *et al.*, 2000) and in this project the interaction of RHA with completely cytoplasmic Sam68 (P439R) was further detected. Immunoprecipitation of endogenous Sam68 protein using protein A Dynabeads also

confirmed the interaction between wild type Sam68 and RHA at normal cellular levels of expression. Hence a major observation in this study is that RHA and Sam68 interact with each other both at endogenous and over-expressed levels of Sam68 protein.

5.3.2   Role of tyrosine-rich C-terminal domain in Sam68 and RHA interaction

In addition, by making short forms of Sam68 and testing if the protein-protein interactions still occur, it became clear that only deletion of tyrosine-rich C-terminal domain of Sam68 can disrupt the interaction with RHA. This domain therefore must play a fundamental role in enabling the interaction between these proteins. The tyrosine-rich C-terminal domain is also present within the structure of T-STAR as well. Tyrosine residues can be phosphorylated by protein kinases and then act like a dock for signalling proteins which contain SH2 domains (Ehrmann and Elliott, 2010). Both T-STAR and Sam68 interacted with RHA, and there is a possibility that these interactions may be related to signal transduction pathways.

5.3.3   Detection of Sam68 and CCAR1 interaction

Interactions between wild type and P439R Sam68 with CCAR1 (Cell cycle and apoptosis regulatory protein) were detected after over-expression of Sam68. Both Sam68 and CCAR1 proteins have roles in apoptosis, cancer and cell proliferation regulation. Sam68 and CCAR1 proteins only interact with each other after over-expression of Sam68, and not in the situation when Sam68 is expressed normally within the cell. It will be useful to check the influence of Sam68 expression on the level of expression of CCAR1 in breast cancer cells since both of these proteins are over-expressed and also involved in breast cancer (Kim et al., 2008; Elliott and Rajan, 2010). There is a possibility that by testing the relation between expression levels of Sam68 and CCAR1 more information will be available regarding the roles of these proteins in breast cancer.

In addition, in this study it was found out that RHA or CCAR1 were not present in SNBs, but endogenous RHA (mostly), CCAR1 and Sam68 were localised within the nucleoplasm. Since Sam68 was present both in SNBs and the cytoplasm it seems that

Sam68 interactions with RHA and CCAR1 must occur in the nucleoplasm instead of SNBs.

5.4   Investigation of RNA-protein interaction networks of T-STAR

FLAG tagged protein (T-STAR) and its partners were purified using immunoprecipitation and identified using mass spectrometry. In addition to proving the presence of T-STAR in the immunoprecipitates, two other proteins (hnRNP G and RNA helicase A (RHA)) were detected as T-STAR protein partners by mass spectrometry. This was the first time that RHA was detected as an interacting protein partner of T-STAR. The interaction of T-STAR and RNA helicase A was tested at the endogenous levels of T-STAR expression as well. Immunoprecipitation of endogenous T-STAR protein using protein A Dynabeads was carried out and immunoprecipitated (IP) samples were probed with appropriate antibodies, but no RHA was detected in the immunoprecipitated (IP) or input lanes. This might be because of negligible expression of T-STAR in kidney cells. Therefore, using other cell lines or tissues which have more T-STAR expression might be a better option to try to investigate more about the interaction between T-STAR and RHA.

5.5   Investigation of RNA-protein interaction networks of hnRNP G and hnRNP G-T

Potential protein interactions of hnRNP G and hnRNP G-T were investigated in this study. FLAG tagged proteins (hnRNP G and hnRNP G-T) and their partners were purified using immunoprecipitation and identified using mass spectrometry. The presence of RPL6 and hnRNP H3 were detect as the new candidate protein partners for hnRNP G after over-expression of this protein. H2AFY, RPS6, RPL7, RPL7A, RPLP2 and RPS20 were detected as candidate interacting partners for hnRNP G-T.

5.5.1 HnRNP C is a common protein partner of hnRNP G and hnRNP G-T

X-Linked-Like 1 (RBMXL1), hnRNP C, and hnRNP CL1 were detected to be common protein partners for both of hnRNP G and hnRNP G-T proteins. It was the first time that the interaction between hnRNP G-T and hnRNP C has been detected. HnRNP C plays a role in blocking U2AF65 (the core splicing factor) to prevent inclusion of cryptic exons and maintain transcriptome integrity (Zarnack et al., 2013). HnRNP proteins interact with other RNA binding proteins (see introduction). The role of hnRNP G protein is not fully understood in alternative regulation. This new protein interaction detected between hnRNP C and hnRNP G might help explain the role of hnRNP G protein in alternative splicing regulation in more details.

HnRNP G, hnRNP G-T have roles in some diseases such as SMA and male infertility respectively (Elliott, 2004) and hnRNP C maintains repression of cryptic exons which seems to be associated to disease including cancer (Sun et al., 2007; Blume et al., 2010; Zarnack et al., 2013). The diseases mentioned above are linked to splicing misregulation of some target genes. Therefore, the more data we could collect regarding these RNA binding proteins interactions, the more information will be achieved for designing future drugs and treatments such as gene therapy via splicing regulation of target genes involved in different diseases.

5.6 Use of stable cell lines to probe RNA splicing patterns

In this project, for the first time, a strong influence of hnRNP G-T over-expression on cell morphology was detected. HEK293 cells expressing hnRNP G-T developed firstly a circular shape, and then were released and floated within the media after increasing the level of expression of this protein. Alternative splicing is very important in cell morphology and also in changes which occur to cells in development and disease is called epithelial–mesenchymal transition (EMT).

5.6.1 Repressive roles of hnRNP G and hnRNP G-T in alternative splicing regulation of some EMT target genes

HnRNP G-T over-expression repressed splicing regulation of some target genes involved in epithelial–mesenchymal transitions (EMT) including *ACOT9* (acyl-CoA thioesterase 9), *ENAH* (enabled homolog) and *MOYF* (myoferlin). In addition, *ENAH* was detected to be the common target gene for both hnRNP G and hnRNP G-T.

Since in this project the interaction of hnRNP G and hnRNP G-T with hnRNP C were detected and because of the roles of hnRNP C and hnRNP G in alternative splicing regulation (Hofmann and Wirth, 2002; Nasim *et al.*, 2003; Zarnack *et al.*, 2013), there are some future experiments to be carried out. After confirming the endogenous interaction of hnRNP G and hnRNP G-T with hnRNP C, the next step will be testing the influence of these RNA binding proteins as direct or indirect regulators of target genes (*ACOT9, MOYF* and *ENAH*).

5.7 Discovering the role of hnRNP G-T in cell morphology regulation of cells expressing this protein

We know that EMT plays important role in development (Thiery, 2002). HnRNP G-T induces morphological changes for cells expressing this protein. Cells developed firstly a circular shape, and then were released and floated within the media after increasing the level of expression of hnRNP G-T. Also, in this project the role of hnRNP G-T was detected in splicing regulation of some target genes involved in EMT. HnRNP G-T has been shown to be only expressed in the nuclei of pachytene spermatocytes and round spermatids (Elliott et al., 2000b). Therefore, it seems that we could consider some unique roles for this protein in germ cell development. The question to be answered as future work is whether hnRNP G-T can play an important role in movement of spermatogonia from the basement membrane towards lumen of seminiferous tubules?

5.8 Future work

Following on from the results achieved in this study there are some questions to be answered. These questions are listed below:

5.8.1 Does endogenous T-STAR interact with RHA?

After confirmation of interaction of T-STAR with RHA at over-expressed levels of T-STAR protein, to confirm whether the interaction also happens at endogenous levels of T-STAR expression, we could use brain or testis tissues. If we could detect interaction it means that their association does not depend on a high level of T-STAR expression. If an interaction cannot be detected at the endogenous level of T-STAR expression, it will show us that higher level of T-STAR protein is needed for protein association.

5.8.2 Does over-expressed Sam68 interact with over-expressed CCAR1 protein in breast cancer cells?

The interaction of Sam68 was detected with CCAR1 only after over-expression of Sam68 protein and not at the endogenous level of Sam68 expression. Both Sam68 and CCAR1 proteins are involved in transcription regulation and also are over-expressed in breast cancer (Busa et al., 2007; Kim et al., 2008; Rajan et al., 2008b; Elliott and Rajan, 2010). Therefore, there could be a link between the expression level of Sam68 and its interaction with CCAR1. To discover if there is such a link, one option will be to uncover the influence of Sam68 on CCAR1 expression in breast cancer cell line. The question will be whether knocking down Sam68 in cancer cells can lead to decreased level of CCAR1 expression as well or not. If yes, will cell proliferation rate be reduced regarding to the roles of CCAR1 and Sam68 in regulation of cell cycle and apoptosis (Taylor *et al.*, 2004; Paronetto *et al.*, 2007; Ou *et al.*, 2009)?

5.8.3 Do hnRNP C and hnRNP G proteins interact at endogenous levels of protein expression?

As approved in this study, hnRNP G and hnRNP G-T interact with hnRNPC after over-expression of hnRNP G and hnRNP G-T proteins. Therefore, one of the future works could be testing that whether these interactions occour at endogenous expression levels of hnRNP G and hnRNP G-T. According to the important roles of these proteins in alternative splicing regulation, detection of new interacting partners and characterization of interaction situation will help us to understand the roles of these RNA binding proteins in splicing regulation of target genes.

5.8.4 Do hnRNP G and hnRNP G-T proteins interact with other identified partners from the proteomic screen?

In this project plus hnRNP C there were some other detected new protein candidates as the protein partners of hnRNP G and hnRNP G-T (table 31 and 32). These interactions occurred after over-expression of hnRNP G and hnRNP G-T proteins. Therefore, confirmation of these interactions at endogenous level of hnRNP G and hnRNP G-T protein expression level can be considered as one the future work to achieve more information related to interaction networks of these members of hnRNPs family.

5.8.5 Does hnRNP G-T protein regulate the splicing of identified target exons?

HnRNP G-T role was detected as a repressor in exon inclusion of some EMT target genes including *ACOT9*, *ENAH* and *MOYF*. However, the direct or indirect role of this protein in splicing regulation of these target genes needs to be uncovered. If hnRNP G-T is not effecting alternative splicing regulation of these target genes directly, it means that this protein might be playing an indirect role to recruit splicing factors involved in repression of these cassette exons' inclusions.

However, if we could detect hnRNP G-T acts as a direct repressor it means that this protein binds directly to the cassette exons and its presence is necessary for repression of cassette exons' inclusions of these target genes *(ACOT9*, *ENAH* and *MOYF)*. According to the detected interaction of hnRNP G-T with hnRNP C and their roles in splicing regulation, if we can confirm the direct effect of hnRNP G-T in splicing regulation of these EMT target genes, we could then test whether hnRNP C also plays any role in inclusion regulation of these cassette exons. The other question will be whether hnRNP C acts synergistically or antagonistically to hnRNP G-T? Do they collaborate with each other to regulate alternative splicing of these target genes or do they repress each other's activity? This question can be expanded to more common targets as well.

5.8.6   Testing detected EMT target gene (ENAH) is direct targets of hnRNP G protein

In addition a role for hnRNP G was detected in repression of cassette exon inclusion of *ENAH*. Therefore, similar to what has been mentioned in the previous section, one of the future works should be to test direct or indirect repressive roles of hnRNP G in splicing regulation of the *ENAH* cassette exon. If hnRNP G acts as a direct repressor, we could then test the role of hnRNP C as well in terms of possible collaboration and interaction between these two proteins to regulate alternative splicing of this target gene (*ENAH*).

Appendix A

(Sequencing results)

1. Sequencing results of *T-STAR*, wild type *Sam68* and P439R *Sam68*. ClustalW was used for generation of alignments. The stars show the identity. The stop codon is highlighted by red.

5' end of *T-STAR* insert showing the frame using CMVF primer

```
T-STAR      ------------------------------------------------------------
Construct   NNNNNGNNNNNNNNNNNNTNNNGCTNNTNTGACCTCCATAGNAGACACCGGGACCGATCCA 60


T-STAR      ------------------------------------------------------------
Construct   GCCTCCGGACTCTAGCGTTTAAACTTAAGCTTGGTACCGATCCACCATGGACTACAAAGA 120


T-STAR      ------------------------------------------------------------
Construct   CGATGACGATAAAGACTACAAAGACGATGACGATAAAGGTCACCCAGGATCACTGGAAGT 180


T-STAR      -----------------------------------------ATGGAGGAGAAGTACCT 17
Construct   TCTGTTCCAGGGGCCCCTGCATCACCACCATCACCATGGATCTATGGAGGAGAAGTACCT 240
                                                     ****************


T-STAR      GCCCGAGCTGATGGCGGAGAAGGACTCCCTGGACCCCTCCTTCACGCACGCCCTGCGCCT 77
Construct   GCCCGAGCTGATGGCGGAGAAGGACTCCCTGGACCCCTCCTTCACGCACGCCCTGCGCCT 300
            ************************************************************
```

3' end of *T-STAR* insert showing the frame using BGHrev primer

```
T-STAR      GGAGTCCCCCCAACTGGGTACAGACCTCCACCGCCACCCCCGACACAAGAGACTTATGGA 801
Construct   GGAGTCCCCCCAACTGGGTACAGACCTCCACCGCCACCCCCGACACAAGAGACTTATGGA 1139
            ************************************************************

T-STAR      GAATATGACTATGATGATGGATATGGCACTGCTTATGATGAACAGAGTTATGATTCCTAT 861
Construct   GAATATGACTATGATGATGGATATGGCACTGCTTATGATGAACAGAGTTATGATTCCTAT 1199
            ************************************************************

T-STAR      GATAACAGCTATAGCACCCCAGCCCAAAGTGGTGCTGATTACTATGATTACGGACATGGA 921
Construct   GATAACAGCTATAGCACCCCAGCCCAAAGTGGTGCTGATTACTATGATTACGGACATGGA 1259
            ************************************************************

T-STAR      CTCAGTGAGGAGACTTATGATTCCTACGGGCAAGAAGAGTGGACTAACTCAAGACACAAG 981
Construct   CTCAGTGAGGAGACTTATGATTCCTACGGGCAAGAAGAGTGGACTAACTCAAGACACAAG 1319
            ************************************************************

T-STAR      GCACCTTCAGCGAGGACAGCAAAGGGCGTCTACAGAGACCAGCCATATGGCAGATACTGA 1041
Construct   GCACCTTCAGCGAGGACAGCAAAGGGCGTCTACAGAGACCAGCCATATGGCAGATACTGA 1379
            ************************************************************
```

Stop codon↑

5' end of wild type *Sam68* insert showing the frame using CMVF primer


```
Sam68       ------------------------------------------------------------
Construct   NNNNNNNNNNNTCNCGCTGTTTTGACCTCCNTAGAAGACACCGGGACCGATCCAGCCTCCG 60


Sam68       ------------------------------------------------------------
Construct   GACTCTAGCGTTTAAACTTAAGCTTGGTACCGATCCACCATGGACTACAAAGACGATGAC 120


Sam68       ------------------------------------------------------------
Construct   GATAAAGACTACAAAGACGATGACGATAAAGGTCACCCAGGATCACTGGAAGTTCTGTTC 180


Sam68       ------------------------------------ATGCAGCGCCGGGACGATCCTGCC 24
Construct   CAGGGGCCCCTGCATCACCACCATCACCATGGATCCATGCAGCGCCGGGACGATCCTGCC 240
                                                ***********************


Sam68       TCGCGCCTCACCCGGTCCTCGGGCCGCAGCTGCTCCAAGGACCCGTCAGGTGCCCACCCC 84
Construct   TCGCGCCTCACCCGGTCCTCGGGCCGCAGCTGCTCCAAGGACCCGTCAGGTGCCCACCCC 300
            ************************************************************


Sam68       TCGGTGCGTCTGACCCCGTCTCGGCCGTCGCCGCTTCCTCACCGGCCCCGGGGAGGGGGA 144
Construct   TCGGTGCGTCTGACCCCGTCTCGGCCGTCGCCGCTTCCTCACCGGCCCCGGGGAGGGGGA 360
            ************************************************************
```

3' end of wild type *Sam68* insert showing the frame using pCDNArev primer


```
Sam68       GGGTGCTCCAACACCAAGAGCTCGGACAGCTGGGATTCAGAGAATACCTTTGCCTCCCAC 1079
Construct   GGGTGCTCCAACACCAAGAGCTCGGACAGCTGGGATTCAGAGAATACCTTTGCCTCCCAC 891
            ************************************************************


Sam68       ACCTGCACCAGAAACATACGAAGATTATGGATATGATGATACATACGCAGAACAGAGTTA 1139
Construct   ACCTGCACCAGAAACATACGAAGATTATGGATATGATGATACATACGCAGAACAGAGTTA 951
            ************************************************************


Sam68       CGAAGGCTATGAAGGCTATTACAGCCAGAGTCAAGGGGAGTCAGAGTATTATGACTATGG 1199
Construct   CGAAGGCTATGAAGGCTATTACAGCCAGAGTCAAGGGGAGTCAGAGTATTATGACTATGG 1011
            ************************************************************


Sam68       ACATGGGGAGCTCCAAGATTCTTACGAAGCCTACGGACAAGATGACTGGAATGGGACCAG 1259
Construct   ACATGGGGAGCTCCAAGATTCTTACGAAGCCTACGGACAAGATGACTGGAATGGGACCAG 1071
            ************************************************************


Sam68       GCCATCACTGAAGGCTCCTCCAGCTAGGCCAGTGAAGGGAGCATACAGAGAGCATCCATA 1319
Construct   GCCATCACTGAAGGCTCCTCCAGCTAGGCCAGTGAAGGGAGCATACAGAGAGCATCCATA 1131
            ************************************************************


Sam68       TGGACGTTATTAA----------------------------------------------- 1332
Construct   TGGACGTTATTAAGTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGT 1191
            *************
```

Stop codon↑

5' end of P439R mutant *Sam68* insert showing the frame using CMVF primer

```
Sam68       ------------------------------------------------------------
Construct   NNNNNNNNNNNNNNNCNNNCTGTTTTGACCTCCNTAGAAGACACCGGGACCGATCCAGCCT 60


Sam68       ------------------------------------------------------------
Construct   CCGGACTCTAGCGTTTAAACTTAAGCTTGGTACCGATCCACCATGGACTACAAAGACGAT 120


Sam68       ------------------------------------------------------------
Construct   GACGATAAAGACTACAAAGACGATGACGATAAAGGTCACCCAGGATCACTGGAAGTTCTG 180


Sam68       ----------------------------------------ATGCAGCGCCGGGACGATCCT 21
Construct   TTCCAGGGGCCCCTGCATCACCACCATCACCATGGATCCATGCAGCGCCGGGACGATCCT 240
                                                    ********************

Sam68       GCCTCGCGCCTCACCCGGTCCTCGGGCCGCAGCTGCTCCAAGGACCCGTCAGGTGCCCAC 81
Construct   GCCTCGCGCCTCACCCGGTCCTCGGGCCGCAGCTGCTCCAAGGACCCGTCAGGTGCCCAC 300
            ************************************************************

Sam68       CCCTCGGTGCGTCTGACCCCGTCTCGGCCGTCGCCGCTTCCTCACCGGCCCCGGGGAGGG 141
Construct   CCCTCGGTGCGTCTGACCCCGTCTCGGCCGTCGCCGCTTCCTCACCGGCCCCGGGGAGGG 360
            ************************************************************
```

3' end of P439R mutant *Sam68* insert showing the frame using pCDNArev primer

Instead of CCA (highlighted by green) which encodes proline, there is CGA (highlighted by blue) which encodes arginine.

```
Sam68       GGTGCTCCAACACCAAGAGCTCGGACAGCTGGGATTCAGAGAATACCTTTGCCTCCCACA 1080
Construct   GGTGCTCCAACACCAAGAGCTCGGACAGCTGGGATTCAGAGAATACCTTTGCCTCCCACA 901
            ************************************************************

Sam68       CCTGCACCAGAAACATACGAAGATTATGGATATGATGATACATACGCAGAACAGAGTTAC 1140
Construct   CCTGCACCAGAAACATACGAAGATTATGGATATGATGATACATACGCAGAACAGAGTTAC 961
            ************************************************************

Sam68       GAAGGCTATGAAGGCTATTACAGCCAGAGTCAAGGGGAGTCAGAGTATTATGACTATGGA 1200
Construct   GAAGGCTATGAAGGCTATTACAGCCAGAGTCAAGGGGAGTCAGAGTATTATGACTATGGA 1021
            ************************************************************

Sam68       CATGGGGAGCTCCAAGATTCTTACGAAGCCTACGGACAAGATGACTGGAATGGGACCAGG 1260
Construct   CATGGGGAGCTCCAAGATTCTTACGAAGCCTACGGACAAGATGACTGGAATGGGACCAGG 1081
            ************************************************************

Sam68       CCATCACTGAAGGCTCCTCCAGCTAGGCCAGTGAAGGGAGCATACAGAGAGCATCCATAT 1320
Construct   CCATCACTGAAGGCTCCTCCAGCTAGGCCAGTGAAGGGAGCATACAGAGAGCATCGATAT 1141
            ******************************************************* ****
```

                                                              Mutation↑

```
Sam68       GGACGTTATTAA----------------------------------------------- 1332
Construct   GGACGTTATTAAGTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTG 1201
            ************
```

    Stop codon↑

2. Sequencing results of different short forms of *Sam68*. ClustalW was used for generation of alignments. The stars show the identity. The stop codon is highlighted by gray.

Dark green: extended N-terminal domain of Sam68; green: proline rich sequences (SH3 binding sits); red: N-terminus; **bold and underlined**: Qua1 domain; **bold, blue and underlined**: KH Domain; blue font color with under line: Qua2 domain; yellow: RG-rich domain; purple: tyrosine-rich C-terminal domain.

Sam68 short form1 (N-KH) insert showing the frame using BGHrev primer

```
Sam68    ------------------------------------------------ATGCAGCGCCG  11
N-KH     GGAAGTTCTGTTCCAGGGGCCCCTGCATCACCACCATCACCATGGATCCATGCAGCGCCG  660
                                                             **********


Sam68    GGACGATCCTGCCTCGCGCCTCACCCGGTCCTCGGGCCGCAGCTGCTCCAAGGACCCGTC  71
N-KH     GGACGATCCTGCCTCGCGCCTCACCCGGTCCTCGGGCCGCAGCTGCTCCAAGGACCCGTC  720
         ************************************************************


Sam68    AGGTGCCCACCCCTCGGTGCGTCTGACCCCGTCTCGGCCGTCGCCGCTTCCTCACCGGCC  131
N-KH     AGGTGCCCACCCCTCGGTGCGTCTGACCCCGTCTCGGCCGTCGCCGCTTCCTCACCGGCC  780
         ************************************************************


Sam68    CCGGGGAGGGGGAGGTGGGCCCAGAGGAGGCGCTCGGGCCTCGCCCGCCACCCAGCCGCC  191
N-KH     CCGGGGAGGGGGAGGTGGGCCCAGAGGAGGCGCTCGGGCCTCGCCCGCCACGCAGCCGCC  840
         *************************************************** ********


Sam68    GCCGCTGCTGCCTCCCTCCACCCCTGGTCCCGACGCGACGGTGGTGGGTTCCGCGCCGAC  251
N-KH     GCCGCTGCTGCCTCCCTCCACCCCGGGTCCCGACGCGACGGTGGTGGGTTCCGCGCCGAC  900
         ************************ ***********************************


Sam68    CCCGCTGCTGCCCCCGTCAGCCACAGCCGCGGTCAAGATGGAGCCGGAGAATAAGTACCT  311
N-KH     CCCGCTGCTGCCCCCGTCAGCCACAGCCGCGGTCAAGATGGAGCCGGAGAATAAGTACCT  960
         ************************************************************


Sam68    GCCTGAACTCATGGCCGAGAAGGACTCGCTCGACCCGTCCTTCACTCACGCCATGCAGCT  371
N-KH     GCCTGAACTCATGGCCGAGAAGGACTCGCTCGACCCGTCCTTCACTCACGCCATGCAGCT  1020
         ************************************************************


Sam68    GCTGTCCGTAGAAATTGAGAAGATTCAGAAGGGAGAGTCAAAAAAAGATGACGAGGAGAA  431
N-KH     GCTGTCCGTAGAAATTGAGAAGATTCAGAAGGGAGAGTCAAAAAAAGATGACGAGGAGAA  1080
         ************************************************************
```

195

```
Sam68    TTATTTGGATTTATTTTCTCATAAGAACATGAAGCTGAAAGAACGCGTGCTGATACCTGT 491
N-KH     TTATTTGGATTTATTTTCTCATAAGAACATGAAGCTGAAAGAACGCGTGCTGATACCTGT 1140
         ************************************************************


Sam68    CAAGCAGTATCCAAAGTTCAATTTTGTGGGGAAGATTCTTGGACCACAAGGAAATACAAT 551
N-KH     CAAGCAGTATCCAAAGTTCAATTTTGTGGGGAAGATTCTTGGACCACAAGGAAATACAAT 1200
         ************************************************************


Sam68    CAAAAGACTCCAGGAAGAGACTGGTGCAAAGATCTCTGTCTTGGGGAAGGGTTCAATGAG 611
N-KH     CAAAAGACTCCAGGAAGAGACTGGTGCAAAGATCTCTGTCTTGGGGAAGGGTTCAATGAG 1260
         ************************************************************


Sam68    AGACAAAGCCAAGGAGGAAGAGTTGCGCAAGGGTGGAGACCCCAAATATGCCCATTTAAA 671
N-KH     AGACAAAGCCAAGGAGGAATAAGT------------------------------------ 1284
         ******************* *. *
```

Stop codon↑

5' end of *Sam68* short form2 (N-KH-RG) insert showing the frame using CMVF primer

```
Sam68    ----------------------------------------ATGCAGCGCCGGGACGATC 19
NKHRG    TGTTCCAGGGGCCCCTGCATCACCACCATCACCATGGATCCATGCAGCGCCGGGACGATC 240
                                                 ******************


Sam68    CTGCCTCGCGCCTCACCCGGTCCTCGGGCCGCAGCTGCTCCAAGGACCCGTCAGGTGCCC 79
NKHRG    CTGCCTCGCGCCTCACCCGGTCCTCGGGCCGCAGCTGCTCCAAGGACCCGTCAGGTGCCC 300
         ************************************************************


Sam68    ACCCCTCGGTGCGTCTGACCCCGTCTCGGCCGTCGCCGCTTCCTCACCGGCCCCGGGGAG 139
NKHRG    ACCCCTCGGTGCGTCTGACCCCGTCTCGGCCGTCGCCGCTTCCTCACCGGCCCCGGGGAG 360
         ************************************************************


Sam68    GGGGAGGTGGGCCCAGAGGAGGCGCTCGGGCCTCGCCCGCCACCCAGCCGCCGCCGCTGC 199
NKHRG    GGGGAGGTGGGCCCAGAGGAGGCGCTCGGGCCTCGCCCGCCACGCAGCCGCCGCCGCTGC 420
         *******************************************  ***************


Sam68    TGCCTCCCTCCACCCCTGGTCCCGACGCGACGGTGGTGGGTTCCGCGCCGACCCCGCTGC 259
NKHRG    TGCCTCCCTCCACCCCGGGTCCCGACGCGACGGTGGTGGGTTCCGCGCCGACCCCGCTGC 480
         ***************  *******************************************


Sam68    TGCCCCCGTCAGCCACAGCCGCGGTCAAGATGGAGCCGGAGAATAAGTACCTGCCTGAAC 319
NKHRG    TGCCCCCGTCAGCCACAGCCGCGGTCAAGATGGAGCCGGAGAATAAGTACCTGCCTGAAC 540
         ************************************************************


Sam68    TCATGGCCGAGAAGGACTCGCTCGACCCGTCCTTCACTCACGCCATGCAGCTGCTGTCCG 379
NKHRG    TCATGGCCGAGAAGGACTCGCTCGACCCGTCCTTCACTCACGCCATGCAGCTGCTGTCCG 600
         ************************************************************
```

3' end of *Sam68* short form2 (N-KH-RG) insert showing the frame using BGHrev primer

```
Sam68    AGTACCTGAACCCTCTCGTGGTCGTGGGGTATCTGTGAGAGGACGAGGAGCTGCCCCTCC  887
NKHRG    AGTACCTGAACCCTCTCGTGGTCGTGGGGTATCTGTGAGAGGACGAGGAGCTGCCCCTCC  1080
         ************************************************************


Sam68    TCCTCCACCTGTTCCCAGAGGACGTGGTGTTGGACCACCTAGAGGAGCTTTGGTTCGTGG  947
NKHRG    TCCTCCACCTGTTCCCAGAGGACGTGGTGTTGGACCACCTAGAGGAGCTTTGGTTCGTGG  1140
         ************************************************************


Sam68    AACCCCAGTGAGAGGCTCCATCACCAGAGGTGCCACTGTGACTCGAGGAGTGCCACCCCC  1007
NKHRG    AACCCCAGTGAGAGGCTCCATCACCAGAGGTGCCACTGTGACTCGAGGAGTGCCACCCCC  1200
         ************************************************************


Sam68    ACCTACTGTGAGGGGTGCTCCAACACCAAGAGCTCGGACAGCTGGGATTCAGAGAATACC  1067
NKHRG    ACCTACTGTGAGGGGTGCTCCAACACCAAGAGCTCGGACATAAG----------------  1244
         ****************************************  .:*
```

Stop codon↑

Sam68 short form 3 (KH-RG) insert showing the frame using BGHrev primer

```
Sam68    -TTCAATTTTGTGGGGAAGATTCTTGGACCACAAGGAAATACAATCAAAAGACTCCAGGA  139
KHRG     CTTCAATTTTGTGGGGAAGATTCTTGGACCACAAGGAAATACAATCAAAAGACTCCAGGA  780
          ***********************************************************


Sam68    AGAGACTGGTGCAAAGATCTCTGTCTTGGGGAAGGGTTCAATGAGAGACAAAGCCAAGGA  199
KHRG     AGAGACTGGTGCAAAGATCTCTGTCTTGGGGAAGGGTTCAATGAGAGACAAAGCCAAGGA  840
         ************************************************************


Sam68    GGAAGAGTTGCGCAAGGGTGGAGACCCCAAATATGCCCATTTAAATATGGATCTGCATGT  259
KHRG     GGAAGAGTTGCGCAAGGGTGGAGACCCCAAATATGCCCATTTAAATATGGATCTGCATGT  900
         ************************************************************


Sam68    CTTCATTGAAGTCTTTGGACCCCCGTGTGAAGCTTATGCTCTTATGGCCCATGCTATGGA  319
KHRG     CTTCATTGAAGTCTTTGGACCCCCGTGTGAAGCTTATGCTCTTATGGCCCATGCTATGGA  960
         ************************************************************


Sam68    AGAAGTCAAGAAGTTCCTAGTACCAGATATGATGGATGATATCTGTCAGGAGCAGTTTCT  379
KHRG     AGAAGTCAAGAAGTTCCTAGTACCAGATATGATGGATGATATCTGTCAGGAGCAGTTTCT  1020
         ************************************************************
```

```
Sam68   AGAATTGTCCTACTTGAACGGAGTACCTGAACCCTCTCGTGGTCGTGGGGTATCTGTGAG 439
KHRG    AGAATTGTCCTACTTGAACGGAGTACCTGAACCCTCTCGTGGTCGTGGGGTATCTGTGAG 1080
        ************************************************************


Sam68   AGGACGAGGAGCTGCCCCTCCTCCTCCACCTGTTCCCAGAGGACGTGGTGTTGGACCACC 499
KHRG    AGGACGAGGAGCTGCCCCTCCTCCTCCACCTGTTCCCAGAGGACGTGGTGTTGGACCACC 1140
        ************************************************************


Sam68   TAGAGGAGCTTTGGTTCGTGGAACCCCAGTGAGAGGCTCCATCACCAGAGGTGCCACTGT 559
KHRG    TAGAGGAGCTTTGGTTCGTGGAACCCCAGTGAGAGGCTCCATCACCAGAGGTGCCACTGT 1200
        ************************************************************


Sam68   GACTCGAGGAGTGCCACCCCCACCTACTGTGAGGGGTGCTCCAACACCAAGAGCTCGGAC 619
KHRG    GACTCGAGGAGTGCCACCCCCACCTACTGTGAGGGGTGCTCCAACACCAAGAGCTCGGAC 1260
        ************************************************************


Sam68   AGCTGGGATTCAGAGAATACCTTTGCCTCCCACACCTGCACCAGAAACATACGAAGATTA 679
KHRG    ATAAG----TCG-------------------------------------AGTNTA 1274
        *  .:*     **.                                    **:.**
```

Stop codon↑

Sam68 short form4 (RG-tyrosine) inser showing the frame using CMVF primer

```
Sam68   ------CGGAG----------------------------TACCTGAACCCTCTCGTGG 31
RGTY    CTGTTCCAGGGGCCCCTGCATCACCACCATCACCATGGATCCCCCTGAACCCTCTCGTGG 240
            *.*.*                              .****************


Sam68   TCGTGGGGTATCTGTGAGAGGACGAGGAGCTGCCCCTCCTCCTCCACCTGTTCCCAGAGG 91
RGTY    TCGTGGGGTATCTGTGAGAGGACGAGGAGCTGCCCCTCCTCCTCCACCTGTTCCCAGAGG 300
        ************************************************************


Sam68   ACGTGGTGTTGGACCACCTAGAGGAGCTTTGGTTCGTGGAACCCCAGTGAGAGGCTCCAT 151
RGTY    ACGTGGTGTTGGACCACCTAGAGGAGCTTTGGTTCGTGGAACCCCAGTGAGAGGCTCCAT 360
        ************************************************************


Sam68   CACCAGAGGTGCCACTGTGACTCGAGGAGTGCCACCCCCACCTACTGTGAGGGGTGCTCC 211
RGTY    CACCAGAGGTGCCACTGTGACTCGAGGAGTGCCACCCCCACCTACTGTGAGGGGTGCTCC 420
        ************************************************************


Sam68   AACACCAAGAGCTCGGACAGCTGGGATTCAGAGAATACCTTTGCCTCCCACACCTGCACC 271
RGTY    AACACCAAGAGCTCGGACAGCTGGGATTCAGAGAATACCTTTGCCTCCCACACCTGCACC 480
        ************************************************************


Sam68   AGAAACATACGAAGATTATGGATATGATGATACATACGCAGAACAGAGTTACGAAGGCTA 331
```

198

```
RGTY          AGAAACATACGAAGATTATGGATATGATGATACATACGCAGAACAGAGTTACGAAGGCTA 540
              ************************************************************


Sam68         TGAAGGCTATTACAGCCAGAGTCAAGGGGAGTCAGAGTATTATGACTATGGACATGGGGA 391
RGTY          TGAAGGCTATTACAGCCAGAGTCAAGGGGAGTCAGAGTATTATGACTATGGACATGGGGA 600
              ************************************************************


Sam68         GCTCCAAGATTCTTACGAAGCCTACGGACAAGATGACTGGAATGGGACCAGGCCATCACT 451
RGTY          GCTCCAAGATTCTTACGAAGCCTACGGACAAGATGACTGGAATGGGACCAGGCCATCACT 660
              ************************************************************


Sam68         GAAGGCTCCTCCAGCTAGGCCAGTGAAGGGAGCATACAGAGAGCATCCATATGGACGTTA 511
RGTY          GAAGGCTCCTCCAGCTAGGCCAGTGAAGGGAGCATACAGAGAGCATCCATATGGACGTTA 720
              ************************************************************


Sam68         TTAA-------------------------------------------------------- 515
RGTY          TTAAGTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAG 780
              ****
```

Stop codon↑

3. Sequencing results of *RHA*, *RBM*, *RBMXL2* and *TRA2B*. ClustalW was used for generation of alignments. Stars show the identity. The stop codon is highlighted by red.

5' end of *RHA* construct insert the frame using CMVF primer

```
RHA           --------------------------------------ATGGGTGACGTTAAAAATTT 20
RHA-a         GTTCCAGGGGCCCCTGCATCACCACCATCACCATGGATCCATGGGTGACGTTAAAAATTT 240
                                                  ******************


RHA           TCTGTATGCCTGGTGTGGCAAAAGGAAGATGACCCCATCCTATGAAATTAGAGCAGTGGG 80
RHA-a         TCTGTATGCCTGGTGTGGCAAAAGGAAGATGACCCCATCCTATGAAATTAGAGCAGTGGG 300
              ************************************************************


RHA           GAACAAAAACAGGCAGAAATTCATGTGTGAGGTTCAGGTGGAAGGTTATAATTACACTGG 140
RHA-a         GAACAAAAACAGGCAGAAATTCATGTGTGAGGTTCAGGTGGAAGGTTATAATTACACTGG 360
              ************************************************************


RHA           CATGGGAAATTCCACCAATAAAAAAGATGCACAAAGCAATGCTGCCAGAGACTTTGTTAA 200
RHA-a         CATGGGAAATTCCACCAATAAAAAAGATGCACAAAGCAATGCTGCCAGAGACTTTGTTAA 420
              ************************************************************


RHA           CTATTTGGTTCGAATAAATGAAATAAAGAGTGAAGAAGTTCCAGCTTTTGGGGTAGCATC 260
RHA-a         CTATTTGGTTCGAATAAATGAAATAAAGAGTGAAGAAGTTCCAGCTTTTGGGGTAGCATC 480
              ************************************************************
```

```
RHA       TCCGCCCCCACTTACTGATACTCCTGACACTACAGCAAATGCTGAAGGAGATTTACCAAC 320
RHA-a     TCCGCCCCCACTTACTGATACTCCTGACACTACAGCAAATGCTGAAGGAGATTTACCAAC 540
          ************************************************************


RHA       AACCATGGGAGGACCTCTTCCTCCACATCTGGCTCTCAAAGCAGAAAATAATTCTGAGGT 380
RHA-a     AACCATGGGAGGACCTCTTCCTCCACATCTGGCTCTCAAAGCAGAAAATAATTCTGAGGT 600
          ************************************************************
```

3' end of *RHA* insert showing the frame using BGHrev primer

```
RHA       CGTCCTCCCAAGATGGCCCGATACGACAATGGAAGCGGATATAGAAGGGGAGGTTCTAGT 3537
RHA-a     CGTCCTCCCAAGATGGCCCGATACGACAATGGAAGCGGATATAGAAGGGGAGGTTCTAGT 746
          ************************************************************


RHA       TACAGTGGTGGAGGCTATGGCGGTGGCTATAGCAGTGGAGGCTATGGTAGCGGAGGCTAT 3597
RHA-a     TACAGTGGTGGAGGCTATGGCGGTGGCTATAGCAGTGGAGGCTATGGTAGCGGAGGCTAT 806
          ************************************************************


RHA       GGTGGCAGCGCCAACTCCTTTCGGGCAGGATATGGTGCAGGTGTTGGTGGAGGCTATAGA 3657
RHA-a     GGTGGCAGCGCCAACTCCTTTCGGGCAGGATATGGTGCAGGTGTTGGTGGAGGCTATAGA 866
          ************************************************************


RHA       GGAGTTTCCCGAGGTGGCTTTAGAGGCAACTCTGGAGGAGACTACAGAGGGCCTAGTGGA 3717
RHA-a     GGAGTTTCCCGAGGTGGCTTTAGAGGCAACTCTGGAGGAGACTACAGAGGGCCTAGTGGA 926
          ************************************************************


RHA       GGCTACAGAGGATCTGGGGGATTCCAGCGAGGAGGTGGTAGGGGGGCCTATGGAACTGGC 3777
RHA-a     GGCTACAGAGGATCTGGGGGATTCCAGCGAGGAGGTGGTAGGGGGGCCTATGGAACTGGC 986
          ************************************************************


RHA       TACTTTGGACAGGGAAGAGGAGGTGGCGGCTATTAA------------------------ 3813
RHA-a     TACTTTGGACAGGGAAGAGGAGGTGGCGGCTATTAAGTCGAGTCTNNNGNNNNNNNNNNN 1046
          ***********************************
```

Stop codon↑

5' end of *RBMX* insert showing the frame using CMVF primer

```
RBMX        ---------------------------------------ATGGTTGAAGCAGATCGCCC 20
Construct   GTTCCAGGGGCCCCTGCATCACCACCATCACCATGGATCCATGGTTGAAGCAGATCGCCC 240
                                                 ******************

RBMX        AGGAAAGCTCTTTATTGGTGGGCTTAATACAGAGACGAATGAGAAAGCCCTTGAGGCAGT 80
Construct   AGGAAAGCTCTTTATTGGTGGGCTTAATACAGAGACGAATGAGAAAGCCCTTGAGGCAGT 300
            ************************************************************

RBMX        GTTTGGCAAATATGGACGAATAGTGGAAGTTCTTTTGATGAAGGACCGAGAAACGAATAA 140
Construct   GTTTGGCAAATATGGACGAATAGTGGAAGTTCTTTTGATGAAGGACCGAGAAACGAATAA 360
            ************************************************************

RBMX        GTCAAGAGGATTCGCTTTTGTCACTTTTGAAAGCCCAGCAGATGCAAAGGATGCTGCCAG 200
Construct   GTCAAGAGGATTCGCTTTTGTCACTTTTGAAAGCCCAGCAGATGCAAAGGATGCTGCCAG 420
            ************************************************************

RBMX        AGACATGAATGGAAAGTCCTTAGATGGGAAAGCCATCAAGGTGGAGCAAGCCACCAAACC 260
Construct   AGACATGAATGGAAAGTCCTTAGATGGGAAAGCCATCAAGGTGGAGCAAGCCACCAAACC 480
            ************************************************************
```

3' end of *RBMX* insert showing the frame using BGHrev primer

```
RBMX        TTATGGTAACTCACGTAGTGCTCCACCTACACGAGGGCCCCCGCCATCTTATGGTGGAAG 920
Construct   TTATGGTAACTCACGTAGTGCTCCACCTACACGAGGGCCCCCGCCATCTTATGGTGGAAG 952
            ************************************************************

RBMX        CAGTCGCTATGATGATTACAGCAGCTCACGTGACGGATATGGTGGAAGTCGAGACAGTTA 980
Construct   CAGTCGCTATGATGATTACAGCAGCTCACGTGACGGATATGGTGGAAGTCGAGACAGTTA 1012
            ************************************************************

RBMX        CTCAAGCAGCCGAAGTGATCTCTACTCAAGTGGTCGTGATCGCGTGGGCAGACAAGAAAG 1040
Construct   CTCAAGCAGCCGAAGTGATCTCTACTCAAGTGGTCGTGATCGCGTGGGCAGACAAGAAAG 1072
            ************************************************************

RBMX        AGGGCTTCCCCCTTCTATGGAAAGGGGGTACCCTCCTCCACGTGATTCCTACAGCAGTTC 1100
Construct   AGGGCTTCCCCCTTCTATGGAAAGGGGGTACCCTCCTCCACGTGATTCCTACAGCAGTTC 1132
            ************************************************************

RBMX        AAGCCGCGGAGCACCAAGAGGTGGTGGCCGTGGAGGAAGCCGATCTGATAGAGGGGGAGG 1160
Construct   AAGCCGCGGAGCACCAAGAGGTGGTGGCCGTGGAGGAAGCCGATCTGATAGAGGGGGAGG 1192
            ************************************************************

RBMX        CAGAAGCAGATACTAG-------------------------- 1176
Construct   CAGAAGCAGATACTAGGTCGAGTCTAGAGNNNNNNNNNNNNNNNN 1235
            ****************
```

<div align="center">

Stop codon↑

</div>

5' end of *RBMXL2* insert showing the frame using CMVF primer

```
RBMXL2      -----------------------------------------ATGGTTG 7
Construct   TCTGTTCCAGGGGCCCCTGCATCACCACCATCACCATGGATCCATGGTTG 150
                                                     *******

RBMXL2      AAGCGGATCGCCCGGGGGAAGCTGTTCATTGGGGGCCTCAACCTCGAAACC 57
Construct   AAGCGGATCGCCCGGGGGAAGCTGTTCATTGGGGGCCTCAACCTCGAAACC 200
            **************************************************

RBMXL2      GACGAGAAAGCCCTCGAAGCCGAGTTTGGCAAGTATGGCCGCATCGTCGA 107
Construct   GACGAGAAAGCCCTCGAAGCCGAGTTTGGCAAGTATGGCCGCATCGTCGA 250
            **************************************************
```

3' end of *RBMXL2* insert showing the frame using BGHrev primer

```
RBMXL2      CGCGACCGTGACTACGGGGATCATCTGAGCAGAGGCTCCCATCGAGAGCCCTTTGAGAGC 855
Construct   CGCGACCGTGACTACGGGGATCATCTGAGCAGAGGCTCCCATCGAGAGCCCTTTGAGAGC 896
            ************************************************************


RBMXL2      TACGGAGAGCTGCGCGGCGCCGCCCCAGGACGGGGGACACCGCCATCTTACGGAGGAGGA 915
Construct   TACGGAGAGCTGCGCGGCGCCGCCCCAGGACGGGGGACACCGCCATCTTACGGAGGAGGA 956
            ************************************************************


RBMXL2      GGCCGCTACGAGGAGTACCGGGGCTACTCACCCGATGCCTACAGCGGCGGCCGCGACAGT 975
Construct   GGCCGCTACGAGGAGTACCGGGGCTACTCACCCGATGCCTACAGCGGCGGCCGCGACAGT 1016
            ************************************************************


RBMXL2      TACAGCAGCAGTTATGGCCGGAGCGACCGCTACTCGAGGGGCCGACACCGGGTGGGCAGA 1035
Construct   TACAGCAGCAGTTATGGCCGGAGCGACCGCTACTCGAGGGGCCGACACCGGGTGGGCAGA 1076
            ************************************************************


RBMXL2      CCAGATCGTGGGCTCTCTCTGTCCATGGAAAGGGGCTGCCCTCCCCAGCGTGATTCTTAC 1095
Construct   CCAGATCGTGGGCTCTCTCTGTCCATGGAAAGGGGCTGCCCTCCCCAGCGTGATTCTTAC 1136
            ************************************************************


RBMXL2      AGCCGGTCAGGCTGCAGGGTGCCCAGGGGCGGAGGCCGTCTAGGAGGCCGCTTGGAGAGA 1155
Construct   AGCCGGTCAGGCTGCAGGGTGCCCAGGGGCGGAGGCCGTCTAGGAGGCCGCTTGGAGAGA 1196
            ************************************************************


RBMXL2      GGAGGAGGCCGGAGCAGATACTAA-------------------------- 1179
Construct   GGAGGAGGCCGGAGCAGATACTAAGTCGAGTCTAGAGNNNNNNNNNNNNNNNN 1247
            ***********************
```

Stop codon↑

5' end of *TRA2B* insert showing the frame using CMVF primer

```
TRA2B       ------------------------------------ATGAGCGACAGCGGCGAGCAGA 22
Construct   TCCAGGGGCCCCTGCATCACCACCATCACCATGGATCCATGAGCGACAGCGGCGAGCAGA 240
                                                ********************

TRA2B       ACTACGGCGAGCGGGAATCCCGTTCTGCTTCCAGAAGTGGAAGTGCTCACGGATCGGGGA 82
Construct   ACTACGGCGAGCGGGAATCCCGTTCTGCTTCCAGAAGTGGAAGTGCTCACGGATCGGGGA 300
            ************************************************************

TRA2B       AATCTGCAAGGCATACCCCTGCAAGGTCTCGCTCCAAGGAAGATTCCAGGCGTTCCAGAT 142
Construct   AATCTGCAAGGCATACCCCTGCAAGGTCTCGCTCCAAGGAAGATTCCAGGCGTTCCAGAT 360
            ************************************************************

TRA2B       CAAAGTCCAGGTCCCGATCTGAATCTAGGTCTAGATCCAGAAGAAGCTCCCGAAGGCATT 202
Construct   CAAAGTCCAGGTCCCGATCTGAATCTAGGTCTAGATCCAGAAGAAGCTCCCGAAGGCATT 420
            ************************************************************

TRA2B       ATACCCGGTCACGGTCTCGCTCCCGCTCCCATAGACGATCACGTAGCAGGTCTTACAGTC 262
Construct   ATACCCGGTCACGGTCTCGCTCCCGCTCCCATAGACGATCACGTAGCAGGTCTTACAGTC 480
            ************************************************************
```

3' end of *TRA2B* insert showing the frame using pCDNArev

```
TRA2B       TAGGCGTTCAAGAGGATTTGCCTTTGTATATTTTGAAAATGTAGATGATGCCAAGGAAGC 524
Construct   TAGGCGTTCAAGAGGATTTGCCTTTGTATATTTTGAAAATGTAGATGATGCCAAGGAAGC 780
            ************************************************************

TRA2B       TAAAGAACGTGCCAATGGAATGGAGCTTGATGGGCGTAGGATCAGAGTTGATTTCTCTAT 584
Construct   TAAAGAACGTGCCAATGGAATGGAGCTTGATGGGCGTAGGATCAGAGTTGATTTCTCTAT 840
            ************************************************************

TRA2B       AACAAAAAGACCACATACGCCAACACCAGGAATTTACATGGGGAGACCTACCTATGGCAG 644
Construct   AACAAAAAGACCACATACGCCAACACCAGGAATTTACATGGGGAGACCTACCTATGGCAG 900
            ************************************************************

TRA2B       CTCTCGCCGTCGGGATTACTATGACAGAGGATATGATCGGGGCTATGATGATCGGGACTA 704
Construct   CTCTCGCCGTCGGGATTACTATGACAGAGGATATGATCGGGGCTATGATGATCGGGACTA 960
            ************************************************************

TRA2B       CTATAGCAGATCATACAGAGGAGGAGGTGGAGGAGGAGGAGGATGGAGAGCTGCCCAAGA 764
Construct   CTATAGCAGATCATACAGAGGAGGAGGTGGAGGAGGAGGAGGATGGAGAGCTGCCCAAGA 1020
            ************************************************************

TRA2B       CAGGGATCAGATTTATAGAAGGCGGTCACCTTCTCCTTACTATAGTCGTGGAGGATACAG 824
Construct   CAGGGATCAGATTTATAGAAGGCGGTCACCTTCTCCTTACTATAGTCGTGGAGGATACAG 1080
            ************************************************************
```

```
TRA2B        ATCACGTTCCAGATCTCGATCATACTCACCTCGTCGCTATTAA----------------  867
Construct    ATCACGTTCCAGATCTCGATCATACTCACCTCGTCGCTATTAAGTCGAGTCTAGAGGGCC 1140
             ******************************************
```

Stop codon↑

Appendix B

(The predicted binding sites of RNA binding proteins within the alternatively spliced exons of *ACOT9* and *MYOF*)

ACOT9:

Number	Name	Sequence	Start	End	Number	Name	Sequence	Start	End
1	A2BP1	UGCAUG	41	46	34	KHDRBS3	UAUAAU	473	478
2	PABPC1	AAAAA	10	14	35	KHDRBS3	ACUAAU	332	337
3	PABPC1	AAAAA	149	153	36	KHDRBS3	UCUAAU	222	227
4	RBMY1A1	CUCAA	162	166	37	KHDRBS3	AUUAAU	467	472
5	a2bp1	GCAUG	250	254	38	ELAVL1	GUUU	1	4
6	a2bp1	GCAUG	42	46	39	ELAVL1	GUUU	201	204
7	RBMY1A1	CUCAA	162	166	40	ELAVL1	GUUU	66	69
8	PABPC1	ACUAAUA	332	338	41	KHDRBS3	UAUAAA	51	56
9	FUS	GGUG	277	280	42	KHDRBS3	CAUAAA	156	161
10	SNRPA	CUUGCAC	343	349	43	KHDRBS3	UAUAAA	451	456
11	Pum2	UGUA	128	131	44	KHDRBS3	GUUAAU	107	112
12	Pum2	UGUA	325	328	45	KHDRBS3	AGUAAU	518	523
13	Pum2	UGUA	398	401	46	SRSF1	UGGA	491	494
14	SRSF9	AGGAC	88	92	47	SRSF1	UGGA	123	126
15	SRSF9	AGGAG	444	448	48	KHDRBS3	ACUAAA	368	373
16	MBNL1	UGCU	213	216	49	KHDRBS3	CUUAAA	424	429
17	MBNL1	UGCU	100	103	50	KHDRBS3	CUUAAA	316	321
18	MBNL1	UGCU	408	411	51	KHDRBS3	GGUAAU	286	291
19	MBNL1	UGCU	142	145	52	KHDRBS3	UAUAAC	117	122
20	EIF4B	GGAA	124	127	53	KHDRBS3	GUUAAA	34	39
21	EIF4B	GGAA	147	150	54	ELAVL1	AUUU	21	24
22	MBNL1	CGCU	341	344	55	ELAVL1	AUUU	414	417
23	SRSF9	AGCAC	329	333	56	ELAVL1	AUUU	61	64
24	RBMX	CCAG	442	445	57	ELAVL1	AUUU	111	114
25	EIF4B	GGAC	89	92	58	ELAVL1	AUUU	321	324
26	QKI	UUCUAACAAC	433	442	59	ELAVL1	AUUU	522	525
27	Vts1	GCUGGUU	214	220	60	ELAVL1	AUUU	230	233
28	YTHDC1	UACUAC	181	186	61	ELAVL1	AUUU	189	192
29	RBMX	CCAU	19	22	62	ELAVL1	AUUU	133	136
30	SRSF1	AGGA	444	447	63	ELAVL1	AUUU	173	176
31	SRSF1	AGGA	146	149	64	ELAVL1	AUUU	488	491
32	SRSF1	AGGA	88	91	65	KHDRBS3	UCUAAC	434	439
33	SRSF1	AGGA	95	98	66	KHDRBS3	AUUAAC	364	369

MOYF:

Number	Name	Sequence	Start	End
1	RBMY1A1	CUCAA	394	398
2	RBMY1A1	CUCAA	394	398
3	FUS	GGUG	15	18
4	FUS	GGUG	22	25
5	FUS	GGUG	46	49
6	FUS	GGUG	143	146
7	Pum2	UGUA	352	355
8	Pum2	UGUA	88	91
9	Pum2	UGUA	427	430
10	PTBP1	AGAACUUUUUUUUGA	368	382
11	MBNL1	UGCU	295	298
12	MBNL1	UGCU	502	505
13	MBNL1	UGCU	388	391
14	MBNL1	UGCU	17	20
15	MBNL1	UGCU	308	311
16	MBNL1	UGCU	148	151
17	MBNL1	UGCU	488	491
18	MBNL1	UGCU	407	410
19	MBNL1	UGCU	474	477
20	EIF4B	GGAA	265	268
21	KHSRP	GUCC	496	499
22	KHSRP	GUCC	445	448
23	YTHDC1	UAAUGC	203	208
24	RBMX	CCAG	228	231
25	RBMX	CCAG	480	483
26	RBMX	CCAU	499	502
27	RBMX	CCAU	435	438
28	RBMX	CCAU	522	525
29	RBMX	CCAU	492	495
30	SRSF13A	AAAGGGU	101	107
31	KHDRBS3	AAUAAU	416	421

Number	Name	Sequence	Start	End
34	KHDRBS3	UAUAAU	201	206
35	SRSF13A	AAAGAAA	465	471
36	ELAVL1	GUUU	222	225
37	ELAVL1	GUUU	403	406
38	ELAVL1	GUUU	184	187
39	ELAVL1	GUUU	290	293
40	ELAVL1	GUUU	197	200
41	RBMX	CCCG	208	211
42	RBMX	CCCG	56	59
43	KHDRBS3	GCUAAU	330	335
44	KHDRBS3	AGUAAU	413	418
45	SRSF1	UGGA	170	173
46	KHDRBS3	CGUAAU	300	305
47	KHDRBS3	CUUAAA	461	466
48	KHDRBS3	GCUAAA	510	515
49	KHDRBS3	GCUAAA	408	413
50	KHDRBS3	GCUAAA	149	154
51	ELAVL1	AUUU	128	131
52	ELAVL1	AUUU	236	239
53	ELAVL1	AUUU	251	254
54	ELAVL1	AUUU	304	307
55	ELAVL1	AUUU	173	176
56	ELAVL1	AUUU	64	67
57	ELAVL1	AUUU	384	387
58	ELAVL1	AUUU	420	423
59	ELAVL1	AUUU	71	74
60	ELAVL1	AUUU	135	138
61	ELAVL1	AUUU	529	532
62	KHDRBS3	AUUAAC	178	183
63	KHDRBS3	UUUAAC	72	77
64	KHDRBS3	UUUAAC	223	228

(The predicted binding sites of RNA binding proteins within the alternatively spliced exons of *ENAH*)

ENAH:

Number	Name	Sequence	Start	End		Number	Name	Sequence	Start	End
1	NONO	AGGGA	408	412		34	SRSF1	AGGA	364	367
2	PABPC1	AAAAA	444	448		35	KHDRBS3	CAUAAU	187	192
3	PABPC1	AAAAA	22	26		36	ELAVL1	GUUU	235	238
4	PABPC1	AAAAA	114	118		37	ELAVL1	GUUU	152	155
5	PABPC1	AAAAA	267	271		38	ELAVL1	GUUU	279	282
6	RBMY1A1	CUCAA	142	146		39	ELAVL1	GUUU	157	160
7	RBMY1A1	CUCAA	142	146		40	KHDRBS3	UAUAAA	19	24
8	PTBP1	AGAUUUUUUUUUCUU	412	426		41	KHDRBS3	UAUAAA	174	179
9	RBMY1A1	CACAA	404	408		42	KHDRBS3	GUUAAU	133	138
10	sap-49	UUGUGA	333	338		43	KHDRBS3	AGUAAU	86	91
11	Pum2	UGUA	59	62		44	SFRS1	UGGA	389	392
12	Pum2	UGUA	219	222		45	SFRS1	UGGA	469	472
13	Pum2	UGUA	535	538		46	KHDRBS3	UGUAAU	37	42
14	Pum2	UGUA	553	556		47	KHDRBS3	ACUAAA	80	85
15	Pum2	UGUA	37	40		48	KHDRBS3	GAUAAA	479	484
16	PABPC1	ACAAAUG	486	492		49	KHDRBS3	UUUAAA	348	353
17	MBNL1	UGCU	323	326		50	KHDRBS3	UUUAAA	441	446
18	MBNL1	UGCU	160	163		51	KHDRBS3	CUUAAA	92	97
19	EIF4B	GGAA	265	268		52	KHDRBS3	UUUAAA	47	52
20	EIF4B	GGAA	519	522		53	KHDRBS3	UUUAAA	204	209
21	EIF4B	GGAA	112	115		54	KHDRBS3	AAUAAC	352	357
22	KHSRP	GUCC	293	296		55	KHDRBS3	GGUAAU	395	400
23	YTHDC1	GAAUAC	433	438		56	KHDRBS3	CAUAAC	76	81
24	QKI	UAUUAACAAC	163	172		57	ELAVL1	AUUU	338	341
25	QKI	AAAUAACAUG	351	360		58	ELAVL1	AUUU	125	128
26	RBMX	CCAU	454	457		59	ELAVL1	AUUU	44	47
27	RBMX	CCAU	548	551		60	ELAVL1	AUUU	414	417
28	RBMX	CCAU	31	34		61	ELAVL1	AUUU	428	431
29	RBMX	CCAC	403	406		62	ELAVL1	AUUU	550	553
30	RBMX	CCAC	530	533		63	ELAVL1	AUUU	200	203
31	SRSF1	AGGA	264	267		64	ELAVL1	AUUU	53	56
32	SRSF1	AGGA	477	480		65	KHDRBS3	AUUAAC	164	169

Refrences

Abdelhaleem, M. (2004) 'Do human RNA helicases have a role in cancer?', Biochim Biophys Acta, 1704(1), pp. 37-46.

Alzhanova-Ericsson, A.T., Sun, X., Visa, N., Kiseleva, E., Wurtz, T. and Daneholt, B. (1996) 'A protein of the SR family of splicing factors binds extensively to exonic Balbiani ring pre-mRNA and accompanies the RNA from the gene to the nuclear pore', Genes Dev, 10(22), pp. 2881-93.

Amann, R.P. and Howards, S.S. (1980) 'Daily spermatozoal production and epididymal spermatozoal reserves of the human male', J Urol, 124(2), pp. 211-5.

Anderson, S.F., Schlegel, B.P., Nakajima, T., Wolpin, E.S. and Parvin, J.D. (1998) 'BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A', Nat Genet, 19(3), pp. 254-6.

Azubel, M., Habib, N., Sperling, R. and Sperling, J. (2006) 'Native spliceosomes assemble with pre-mRNA to form supraspliceosomes', J Mol Biol, 356(4), pp. 955-66.

Azubel, M., Wolf, S.G., Sperling, J. and Sperling, R. (2004) 'Three-dimensional structure of the native spliceosome by cryo-electron microscopy', Mol Cell, 15(5), pp. 833-9.

Babic, I., Cherry, E. and Fujita, D.J. (2006) 'SUMO modification of Sam68 enhances its ability to repress cyclin D1 expression and inhibits its ability to induce apoptosis', Oncogene, 25(36), pp. 4955-64.

Babic, I., Jakymiw, A. and Fujita, D.J. (2004) 'The RNA binding protein Sam68 is acetylated in tumor cell lines, and its acetylation correlates with enhanced RNA binding activity', Oncogene, 23(21), pp. 3781-9.

Baehrecke, E.H. (1997) 'who encodes a KH RNA binding protein that functions in muscle development', Development, 124(7), pp. 1323-32.

Bakkenist, C.J. and Kastan, M.B. (2003) 'DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation', Nature, 421(6922), pp. 499-506.

Bannister, A.J. and Kouzarides, T. (1996) 'The CBP co-activator is a histone acetyltransferase', Nature, 384(6610), pp. 641-3.

Barlat, I., Maurier, F., Duchesne, M., Guitard, E., Tocque, B. and Schweighoffer, F. (1997) 'A role for Sam68 in cell cycle progression antagonized by a spliced variant within the KH domain', J Biol Chem, 272(6), pp. 3129-32.

Bauren, G., Belikov, S. and Wieslander, L. (1998) 'Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3'-end formation and excision of the 3'-terminal intron', Genes Dev, 12(17), pp. 2759-69.

Beil, B., Screaton, G. and Stamm, S. (1997) 'Molecular cloning of htra2-beta-1 and htra2-beta-2, two human homologs of tra-2 generated by alternative splicing', DNA Cell Biol, 16(6), pp. 679-90.

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

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

Best, A., Dagliesh, C., Ehrmann, I., Kheirollahi-Kouhestani, M., Tyson-Capper, A. and Elliott, D.J. (2013) 'Expression of Tra2 beta in Cancer Cells as a Potential Contributory Factor to Neoplasia and Metastasis', Int J Cell Biol, 2013, p. 843781.

Beyer, A.L., Christensen, M.E., Walker, B.W. and LeStourgeon, W.M. (1977) 'Identification and characterization of the packaging proteins of core 40S hnRNP particles', Cell, 11(1), pp. 127-38.

Beyer, A.L. and Osheim, Y.N. (1988) 'Splice site selection, rate of splicing, and alternative splicing on nascent transcripts', Genes Dev, 2(6), pp. 754-65.

Bielli, P., Busa, R., Paronetto, M.P. and Sette, C. (2011) 'The RNA-binding protein Sam68 is a multifunctional player in human cancer', Endocr Relat Cancer, 18(4), pp. R91-R102.

Birney, E., Kumar, S. and Krainer, A.R. (1993) 'Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors', Nucleic Acids Res, 21(25), pp. 5803-16.

Black, D.L. (2003) 'Mechanisms of alternative pre-messenger RNA splicing', Annu Rev Biochem, 72, pp. 291-336.

Blobel, G.A. (2000) 'CREB-binding protein and p300: molecular integrators of hematopoietic transcription', Blood, 95(3), pp. 745-55.

Blume, S.W., Jackson, N.L., Frost, A.R., Grizzle, W.E., Shcherbakov, O.D., Choi, H. and Meng, Z. (2010) 'Northwestern profiling of potential translation-regulatory proteins in human breast epithelial cells and malignant breast tissues: evidence for pathological activation of the IGF1R IRES', Exp Mol Pathol, 88(3), pp. 341-52.

Borgeson, C.D. and Samson, M.L. (2005) 'Shared RNA-binding sites for interacting members of the Drosophila ELAV family of neuronal proteins', Nucleic Acids Res, 33(19), pp. 6372-83.

Bourgeois, C.F., Lejeune, F. and Stevenin, J. (2004) 'Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA', Prog Nucleic Acid Res Mol Biol, 78, pp. 37-88.

Bratt, E. and Ohman, M. (2003) 'Coordination of editing and splicing of glutamate receptor pre-mRNA', RNA, 9(3), pp. 309-18.

Bray, M., Prasad, S., Dubay, J.W., Hunter, E., Jeang, K.T., Rekosh, D. and Hammarskjold, M.L. (1994) 'A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent', Proc Natl Acad Sci U S A, 91(4), pp. 1256-60.

Bullock, A.N., Das, S., Debreczeni, J.E., Rellos, P., Fedorov, O., Niesen, F.H., Guo, K., Papagrigoriou, E., Amos, A.L., Cho, S., Turk, B.E., Ghosh, G. and

Knapp, S. (2009) 'Kinase domain insertions define distinct roles of CLK kinases in SR protein phosphorylation', Structure, 17(3), pp. 352-62.

Busa, R., Paronetto, M.P., Farini, D., Pierantozzi, E., Botti, F., Angelini, D.F., Attisani, F., Vespasiani, G. and Sette, C. (2007) 'The RNA-binding protein Sam68 contributes to proliferation and survival of human prostate cancer cells', Oncogene, 26(30), pp. 4372-82.

Busa, R. and Sette, C. (2010) 'An emerging role for nuclear RNA-mediated responses to genotoxic stress', RNA Biol, 7(4), pp. 390-6.

Busch, A. and Hertel, K.J. (2012) 'Evolution of SR protein and hnRNP splicing regulatory factors', Wiley Interdiscip Rev RNA, 3(1), pp. 1-12.

Caceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L. and Krainer, A.R. (1997) 'Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity', J Cell Biol, 138(2), pp. 225-38.

Caceres, J.F., Screaton, G.R. and Krainer, A.R. (1998) 'A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm', Genes Dev, 12(1), pp. 55-66.

Chabot, B. (1996) 'Directing alternative splicing: cast and scenarios', Trends Genet, 12(11), pp. 472-8.

Chan, H.M. and La Thangue, N.B. (2001) 'p300/CBP proteins: HATs for transcriptional bridges and scaffolds', J Cell Sci, 114(Pt 13), pp. 2363-73.

Chawla, G., Lin, C.H., Han, A., Shiue, L., Ares, M., Jr. and Black, D.L. (2009) 'Sam68 regulates a set of alternatively spliced exons during neurogenesis', Mol Cell Biol, 29(1), pp. 201-13.

Chen, A., Mannen, H. and Li, S.S. (1998) 'Characterization of mouse ubiquitin-like SMT3A and SMT3B cDNAs and gene/pseudogenes', Biochem Mol Biol Int, 46(6), pp. 1161-74.

Chen, C.D., Kobayashi, R. and Helfman, D.M. (1999a) 'Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene', Genes Dev, 13(5), pp. 593-606.

Chen, T., Boisvert, F.M., Bazett-Jones, D.P. and Richard, S. (1999b) 'A role for the GSG domain in localizing Sam68 to novel nuclear structures in cancer cell lines', Mol Biol Cell, 10(9), pp. 3015-33.

Chen, T., Damaj, B.B., Herrera, C., Lasko, P. and Richard, S. (1997) 'Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain', Mol Cell Biol, 17(10), pp. 5707-18.

Chen, W. and Moore, M.J. (2014) 'The spliceosome: disorder and dynamics defined', Curr Opin Struct Biol, 24C, pp. 141-149.

Chen, X., Guo, L., Lin, W. and Xu, P. (2003) 'Expression of Tra2beta isoforms is developmentally regulated in a tissue- and temporal-specific pattern', Cell Biol Int, 27(6), pp. 491-6.

Chen, Y. and Varani, G. (2005) 'Protein families and RNA recognition', FEBS J, 272(9), pp. 2088-97.

Chou, M.Y., Rooke, N., Turck, C.W. and Black, D.L. (1999) 'hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells', Mol Cell Biol, 19(1), pp. 69-77.

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

Christian, K.J., Lang, M.A. and Raffalli-Mathieu, F. (2008) 'Interaction of heterogeneous nuclear ribonucleoprotein C1/C2 with a novel cis-regulatory element within p53 mRNA as a response to cytostatic drug treatment', Mol Pharmacol, 73(5), pp. 1558-67.

Clery, A., Blatter, M. and Allain, F.H. (2008) 'RNA recognition motifs: boring? Not quite', Curr Opin Struct Biol, 18(3), pp. 290-8.

Colwill, K., Feng, L.L., Yeakley, J.M., Gish, G.D., Caceres, J.F., Pawson, T. and Fu, X.D. (1996a) 'SRPK1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors', J Biol Chem, 271(40), pp. 24569-75.

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C. and Duncan, P.I. (1996b) 'The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution', EMBO J, 15(2), pp. 265-75.

Cooper, D.L. and Dougherty, G.J. (1995) 'To metastasize or not? Selection of CD44 splice sites', Nat Med, 1(7), pp. 635-7.

Cote, J., Boisvert, F.M., Boulanger, M.C., Bedford, M.T. and Richard, S. (2003) 'Sam68 RNA binding protein is an in vivo substrate for protein arginine N-methyltransferase 1', Mol Biol Cell, 14(1), pp. 274-87.

Craig, A.M. and Kang, Y. (2007) 'Neurexin-neuroligin signaling in synapse development', Curr Opin Neurobiol, 17(1), pp. 43-52.

Davis, D.B., Delmonte, A.J., Ly, C.T. and McNally, E.M. (2000) 'Myoferlin, a candidate gene and potential modifier of muscular dystrophy', Hum Mol Genet, 9(2), pp. 217-26.

de Hoog, C.L., Foster, L.J. and Mann, M. (2004) 'RNA and RNA binding proteins participate in early stages of cell spreading through spreading initiation centers', Cell, 117(5), pp. 649-62.

Delbridge, M.L., Harry, J.L., Toder, R., O'Neill, R.J., Ma, K., Chandley, A.C. and Graves, J.A. (1997) 'A human candidate spermatogenesis gene, RBM1, is conserved and amplified on the marsupial Y chromosome', Nat Genet, 15(2), pp. 131-6.

Demonbreun, A.R., Lapidos, K.A., Heretis, K., Levin, S., Dale, R., Pytel, P., Svensson, E.C. and McNally, E.M. (2010) 'Myoferlin regulation by NFAT in muscle injury, regeneration and repair', J Cell Sci, 123(Pt 14), pp. 2413-22. Denegri, M., Chiodi, I., Corioni, M., Cobianchi, F., Riva, S. and Biamonti, G. (2001) 'Stress-induced nuclear bodies are sites of accumulation of pre-mRNA processing factors', Mol Biol Cell, 12(11), pp. 3502-14.

Derry, J.J., Richard, S., Valderrama Carvajal, H., Ye, X., Vasioukhin, V., Cochrane, A.W., Chen, T. and Tyner, A.L. (2000) 'Sik (BRK) phosphorylates Sam68 in the nucleus and negatively regulates its RNA binding ability', Mol Cell Biol, 20(16), pp. 6114-26.

Di Fruscio, M., Chen, T., Bonyadi, S., Lasko, P. and Richard, S. (1998) 'The identification of two Drosophila K homology domain proteins. Kep1 and SAM are members of the Sam68 family of GSG domain proteins', J Biol Chem, 273(46), pp. 30122-30.

Di Fruscio, M., Chen, T. and Richard, S. (1999) 'Characterization of Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis', Proc Natl Acad Sci U S A, 96(6), pp. 2710-5.

Dichmann, D.S., Fletcher, R.B. and Harland, R.M. (2008) 'Expression cloning in Xenopus identifies RNA-binding proteins as regulators of embryogenesis and Rbmx as necessary for neural and muscle development', Dev Dyn, 237(7), pp. 1755-66.

Doherty, K.R., Demonbreun, A.R., Wallace, G.Q., Cave, A., Posey, A.D., Heretis, K., Pytel, P. and McNally, E.M. (2008) 'The endocytic recycling protein EHD2 interacts with myoferlin to regulate myoblast fusion', J Biol Chem, 283(29), pp. 20252-60.

Dreumont, N., Bourgeois, C.F., Lejeune, F., Liu, Y., Ehrmann, I.E., Elliott, D.J. and Stevenin, J. (2010) 'Human RBMY regulates germline-specific splicing events by modulating the function of the serine/arginine-rich proteins 9G8 and Tra2-{beta}', J Cell Sci, 123(Pt 1), pp. 40-50.

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

Dreyfuss, G., Matunis, M.J., Pinol-Roma, S. and Burd, C.G. (1993) 'hnRNP proteins and the biogenesis of mRNA', Annu Rev Biochem, 62, pp. 289-321. Duncan, P.I., Howell, B.W., Marius, R.M., Drmanic, S., Douville, E.M. and Bell, J.C. (1995) 'Alternative splicing of STY, a nuclear dual specificity kinase', J Biol Chem, 270(37), pp. 21524-31.

Edenfeld, G., Volohonsky, G., Krukkert, K., Naffin, E., Lammel, U., Grimm, A., Engelen, D., Reuveny, A., Volk, T. and Klambt, C. (2006) 'The splicing factor crooked neck associates with the RNA-binding protein HOW to control glial cell maturation in Drosophila', Neuron, 52(6), pp. 969-80.

Ehrmann, I., Dalgliesh, C., Liu, Y., Danilenko, M., Crosier, M., Overman, L., Arthur, H.M., Lindsay, S., Clowry, G.J., Venables, J.P., Fort, P. and Elliott, D.J. (2013) 'The tissue-specific RNA binding protein T-STAR controls regional splicing patterns of neurexin pre-mRNAs in the brain', PLoS Genet, 9(4), p. e1003474.

Ehrmann, I., Dalgliesh, C., Tsaousi, A., Paronetto, M.P., Heinrich, B., Kist, R., Cairns, P., Li, W., Mueller, C., Jackson, M., Peters, H., Nayernia, K., Saunders, P., Mitchell, M., Stamm, S., Sette, C. and Elliott, D.J. (2008) 'Haploinsufficiency of the germ cell-specific nuclear RNA binding protein hnRNP G-T prevents functional spermatogenesis in the mouse', Hum Mol Genet, 17(18), pp. 2803-18.

Ehrmann, I. and Elliott, D.J. (2010) 'Expression and functions of the star proteins Sam68 and T-STAR in mammalian spermatogenesis', Adv Exp Med Biol, 693, pp. 67-81.

Elliott, D.J. (2000) 'RBMY genes and AZFb deletions', J Endocrinol Invest, 23(10), pp. 652-8.

Elliott, D.J. (2004) 'The role of potential splicing factors including RBMY, RBMX, hnRNPG-T and STAR proteins in spermatogenesis', Int J Androl, 27(6), pp. 328-34.

Elliott, D.J., Best, A., Dalgliesh, C., Ehrmann, I. and Grellscheid, S. (2012) 'How does Tra2beta protein regulate tissue-specific RNA splicing?', Biochem Soc Trans, 40(4), pp. 784-8.

Elliott, D.J., Bourgeois, C.F., Klink, A., Stevenin, J. and Cooke, H.J. (2000a) 'A mammalian germ cell-specific RNA-binding protein interacts with ubiquitously expressed proteins involved in splice site selection', Proc Natl Acad Sci U S A, 97(11), pp. 5717-22.

Elliott, D.J. and Grellscheid, S.N. (2006) 'Alternative RNA splicing regulation in the testis', Reproduction, 132(6), pp. 811-9.

Elliott, D.J., Ma, K., Kerr, S.M., Thakrar, R., Speed, R., Chandley, A.C. and Cooke, H. (1996) 'An RBM homologue maps to the mouse Y chromosome and is expressed in germ cells', Hum Mol Genet, 5(7), pp. 869-74.

Elliott, D.J., Millar, M.R., Oghene, K., Ross, A., Kiesewetter, F., Pryor, J., McIntyre, M., Hargreave, T.B., Saunders, P.T., Vogt, P.H., Chandley, A.C. and Cooke, H. (1997) 'Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm', Proc Natl Acad Sci U S A, 94(8), pp. 3848-53.

Elliott, D.J. and Rajan, P. 'The role of the RNA-binding protein Sam68 in mammary tumourigenesis', J Pathol, 222(3), pp. 223-6.

Elliott, D.J. and Rajan, P. (2010) 'The role of the RNA-binding protein Sam68 in mammary tumourigenesis', J Pathol, 222(3), pp. 223-6.

Elliott, D.J., Venables, J.P., Newton, C.S., Lawson, D., Boyle, S., Eperon, I.C. and Cooke, H.J. (2000b) 'An evolutionarily conserved germ cell-specific hnRNP is encoded by a retrotransposed gene', Hum Mol Genet, 9(14), pp. 2117-24.

Ernst, R.K., Bray, M., Rekosh, D. and Hammarskjold, M.L. (1997) 'A structured retroviral RNA element that mediates nucleocytoplasmic export of intron-containing RNA', Mol Cell Biol, 17(1), pp. 135-44.

Fischer, D.C., Noack, K., Runnebaum, I.B., Watermann, D.O., Kieback, D.G., Stamm, S. and Stickeler, E. (2004) 'Expression of splicing factors in human ovarian cancer', Oncol Rep, 11(5), pp. 1085-90.

Fondell, J.D., Ge, H. and Roeder, R.G. (1996) 'Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex', Proc Natl Acad Sci U S A, 93(16), pp. 8329-33.

Friedemann, J., Grosse, F. and Zhang, S. (2005) 'Nuclear DNA helicase II (RNA helicase A) interacts with Werner syndrome helicase and stimulates its exonuclease activity', J Biol Chem, 280(35), pp. 31303-13.

Fu, X.D. (1995) 'The superfamily of arginine/serine-rich splicing factors', RNA, 1(7), pp. 663-80.

Fu, X.D. (2004) 'Towards a splicing code', Cell, 119(6), pp. 736-8.

Fujita, H., Ohshima, T., Oishi, T., Aratani, S., Fujii, R., Fukamizu, A. and Nakajima, T. (2005) 'Relevance of nuclear localization and functions of RNA helicase A', Int J Mol Med, 15(4), pp. 555-60.

Fuller-Pace, F.V. (2006) 'DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation', Nucleic Acids Res, 34(15), pp. 4206-15.

Fumagalli, S., Totty, N.F., Hsuan, J.J. and Courtneidge, S.A. (1994) 'A target for Src in mitosis', Nature, 368(6474), pp. 871-4.

Fung, E.T., Lanahan, A., Worley, P. and Huganir, R.L. (1998) 'Identification of a Torpedo homolog of Sam68 that interacts with the synapse organizing protein rapsyn', FEBS Lett, 437(1-2), pp. 29-33.

Fusaki, N., Iwamatsu, A., Iwashima, M. and Fujisawa, J. (1997) 'Interaction between Sam68 and Src family tyrosine kinases, Fyn and Lck, in T cell receptor signaling', J Biol Chem, 272(10), pp. 6214-9.

Gabriel, B., Zur Hausen, A., Bouda, J., Boudova, L., Koprivova, M., Hirschfeld, M., Jager, M. and Stickeler, E. (2009) 'Significance of nuclear hTra2-beta1 expression in cervical cancer', Acta Obstet Gynecol Scand, 88(2), pp. 216-21.

Gallouzi, I.E. and Steitz, J.A. (2001) 'Delineation of mRNA export pathways by the use of cell-permeable peptides', Science, 294(5548), pp. 1895-901.
Gelinas, R.E. and Roberts, R.J. (1977) 'One predominant 5'-undecanucleotide in adenovirus 2 late messenger RNAs', Cell, 11(3), pp. 533-44.

Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D., Du, J., Korbel, J.O., Emanuelsson, O., Zhang, Z.D., Weissman, S. and Snyder, M. (2007) 'What is a gene, post-ENCODE? History and updated definition', Genome Res, 17(6), pp. 669-81.

Gertler, F.B., Comer, A.R., Juang, J.L., Ahern, S.M., Clark, M.J., Liebl, E.C. and Hoffmann, F.M. (1995) 'enabled, a dosage-sensitive suppressor of mutations in the Drosophila Abl tyrosine kinase, encodes an Abl substrate with SH3 domain-binding properties', Genes Dev, 9(5), pp. 521-33.

Gertler, F.B., Doctor, J.S. and Hoffmann, F.M. (1990) 'Genetic suppression of mutations in the Drosophila abl proto-oncogene homolog', Science, 248(4957), pp. 857-60.

Ghosh, G. and Adams, J.A. (2011) 'Phosphorylation mechanism and structure of serine-arginine protein kinases', FEBS J, 278(4), pp. 587-97.

Gilbert, S.F. (2000) Developmental biology. 6th edn. Sunderland, Mass.: Sinauer Associates.

Glisovic, T., Bachorik, J.L., Yong, J. and Dreyfuss, G. (2008) 'RNA-binding proteins and post-transcriptional gene regulation', FEBS Lett, 582(14), pp. 1977-86.

Good, P.J. (1995) 'A conserved family of elav-like genes in vertebrates', Proc Natl Acad Sci U S A, 92(10), pp. 4557-61.

Gooding, C., Kemp, P. and Smith, C.W. (2003) 'A novel polypyrimidine tract-binding protein paralog expressed in smooth muscle cells', J Biol Chem, 278(17), pp. 15201-7.

Goodman, R.H. and Smolik, S. (2000) 'CBP/p300 in cell growth, transformation, and development', Genes Dev, 14(13), pp. 1553-77.

Gorlach, M., Burd, C.G., Portman, D.S. and Dreyfuss, G. (1993) 'The hnRNP proteins', Mol Biol Rep, 18(2), pp. 73-8.

Gornemann, J., Kotovic, K.M., Hujer, K. and Neugebauer, K.M. (2005) 'Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex', Mol Cell, 19(1), pp. 53-63.

Grabowski, P.J. (2000) 'Genetic evidence for a Nova regulator of alternative splicing in the brain', Neuron, 25(2), pp. 254-6.

Graveley, B.R. (2000) 'Sorting out the complexity of SR protein functions', Rna, 6(9), pp. 1197-211.

Grellscheid, S., Dalgliesh, C., Storbeck, M., Best, A., Liu, Y., Jakubik, M., Mende, Y., Ehrmann, I., Curk, T., Rossbach, K., Bourgeois, C.F., Stevenin, J., Grellscheid, D., Jackson, M.S., Wirth, B. and Elliott, D.J. (2011)

'Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2beta in development', PLoS Genet, 7(12), p. e1002390.

Grishin, N.V. (2001) 'KH domain: one motif, two folds', Nucleic Acids Res, 29(3), pp. 638-43.

Gruter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B.K. and Izaurralde, E. (1998) 'TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus', Mol Cell, 1(5), pp. 649-59.

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

Guttmacher, A.E. and Collins, F.S. (2002) 'Genomic medicine--a primer', N Engl J Med, 347(19), pp. 1512-20.

Haegebarth, A., Heap, D., Bie, W., Derry, J.J., Richard, S. and Tyner, A.L. (2004) 'The nuclear tyrosine kinase BRK/Sik phosphorylates and inhibits the RNA-binding activities of the Sam68-like mammalian proteins SLM-1 and SLM-2', J Biol Chem, 279(52), pp. 54398-404.

Han, S.P., Tang, Y.H. and Smith, R. 'Functional diversity of the hnRNPs: past, present and perspectives', Biochem J, 430(3), pp. 379-92.

Han, S.P., Tang, Y.H. and Smith, R. (2010) 'Functional diversity of the hnRNPs: past, present and perspectives', Biochem J, 430(3), pp. 379-92.

Hansen, D. and Schedl, T. (2006) 'The regulatory network controlling the proliferation-meiotic entry decision in the Caenorhabditis elegans germ line', Curr Top Dev Biol, 76, pp. 185-215.

Hartman, T.R., Qian, S., Bolinger, C., Fernandez, S., Schoenberg, D.R. and Boris-Lawrie, K. (2006) 'RNA helicase A is necessary for translation of selected messenger RNAs', Nat Struct Mol Biol, 13(6), pp. 509-16.

Hartmann, A.M., Nayler, O., Schwaiger, F.W., Obermeier, A. and Stamm, S. (1999) 'The interaction and colocalization of Sam68 with the splicing-

associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59(fyn)', Mol Biol Cell, 10(11), pp. 3909-26.

Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M. and Luhrmann, R. (2002) 'Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method', Proc Natl Acad Sci U S A, 99(26), pp. 16719-24.

Hay, E.D. (1995) 'An overview of epithelio-mesenchymal transformation', Acta Anat (Basel), 154(1), pp. 8-20.

Heinrich, B., Zhang, Z., Raitskin, O., Hiller, M., Benderska, N., Hartmann, A.M., Bracco, L., Elliott, D., Ben-Ari, S., Soreq, H., Sperling, J., Sperling, R. and Stamm, S. (2009) 'Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regions in pre-mRNA', J Biol Chem, 284(21), pp. 14303-15.

Heyd, F. and Lynch, K.W. (2011) 'Degrade, move, regroup: signaling control of splicing proteins', Trends Biochem Sci, 36(8), pp. 397-404.

Hofmann, Y., Lorson, C.L., Stamm, S., Androphy, E.J. and Wirth, B. (2000) 'Htra2-beta 1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2)', Proc Natl Acad Sci U S A, 97(17), pp. 9618-23.

Hofmann, Y. and Wirth, B. (2002) 'hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1', Hum Mol Genet, 11(17), pp. 2037-49.

Hong, W., Resnick, R.J., Rakowski, C., Shalloway, D., Taylor, S.J. and Blobel, G.A. (2002) 'Physical and functional interaction between the transcriptional cofactor CBP and the KH domain protein Sam68', Mol Cancer Res, 1(1), pp. 48-55.

Hossain, M.N., Fuji, M., Miki, K., Endoh, M. and Ayusawa, D. (2007) 'Downregulation of hnRNP C1/C2 by siRNA sensitizes HeLa cells to various stresses', Mol Cell Biochem, 296(1-2), pp. 151-7.

Hu, Y., Leo, C., Yu, S., Huang, B.C., Wang, H., Shen, M., Luo, Y., Daniel-Issakani, S., Payan, D.G. and Xu, X. (2004) 'Identification and functional

characterization of a novel human misshapen/Nck interacting kinase-related kinase, hMINK beta', J Biol Chem, 279(52), pp. 54387-97.

Huo, L., Wang, Y.N., Xia, W., Hsu, S.C., Lai, C.C., Li, L.Y., Chang, W.C., Wang, Y., Hsu, M.C., Yu, Y.L., Huang, T.H., Ding, Q., Chen, C.H., Tsai, C.H. and Hung, M.C. (2010) 'RNA helicase A is a DNA-binding partner for EGFR-mediated transcriptional activation in the nucleus', Proc Natl Acad Sci U S A, 107(37), pp. 16125-30.

Huranova, M., Jablonski, J.A., Benda, A., Hof, M., Stanek, D. and Caputi, M. (2009) 'In vivo detection of RNA-binding protein interactions with cognate RNA sequences by fluorescence resonance energy transfer', RNA, 15(11), pp. 2063-71.

Hyodo, T., Ito, S., Hasegawa, H., Asano, E., Maeda, M., Urano, T., Takahashi, M., Hamaguchi, M. and Senga, T. (2012) 'Misshapen-like kinase 1 (MINK1) is a novel component of striatin-interacting phosphatase and kinase (STRIPAK) and is required for the completion of cytokinesis', J Biol Chem, 287(30), pp. 25019-29.

Iijima, T., Wu, K., Witte, H., Hanno-Iijima, Y., Glatter, T., Richard, S. and Scheiffele, P. (2011) 'SAM68 regulates neuronal activity-dependent alternative splicing of neurexin-1', Cell, 147(7), pp. 1601-14.

Ishidate, T., Yoshihara, S., Kawasaki, Y., Roy, B.C., Toyoshima, K. and Akiyama, T. (1997) 'Identification of a novel nuclear localization signal in Sam68', FEBS Lett, 409(2), pp. 237-41.

Ito, M., Yuan, C.X., Malik, S., Gu, W., Fondell, J.D., Yamamura, S., Fu, Z.Y., Zhang, X., Qin, J. and Roeder, R.G. (1999) 'Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators', Mol Cell, 3(3), pp. 361-70.

Izaurralde, E., Jarmolowski, A., Beisel, C., Mattaj, I.W., Dreyfuss, G. and Fischer, U. (1997) 'A role for the M9 transport signal of hnRNP A1 in mRNA nuclear export', J Cell Biol, 137(1), pp. 27-35.

Jiang, Y., Puliyappadamba, V.T., Zhang, L., Wu, W., Wali, A., Yaffe, M.B., Fontana, J.A. and Rishi, A.K. (2010) 'A novel mechanism of cell growth

regulation by Cell Cycle and Apoptosis Regulatory Protein (CARP)-1', J Mol Signal, 5, p. 7.

Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R. and Shoemaker, D.D. (2003) 'Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays', Science, 302(5653), pp. 2141-4.

Kahn, C.R. (1994) 'Banting Lecture. Insulin action, diabetogenes, and the cause of type II diabetes', Diabetes, 43(8), pp. 1066-84.

Kajita, K., Kuwano, Y., Kitamura, N., Satake, Y., Nishida, K., Kurokawa, K., Akaike, Y., Honda, M., Masuda, K. and Rokutan, K. (2013) 'Ets1 and heat shock factor 1 regulate transcription of the Transformer 2beta gene in human colon cancer cells', J Gastroenterol.

Kamitani, T., Kito, K., Nguyen, H.P., Fukuda-Kamitani, T. and Yeh, E.T. (1998) 'Characterization of a second member of the sentrin family of ubiquitin-like proteins', J Biol Chem, 273(18), pp. 11349-53.

Kamma, H., Portman, D.S. and Dreyfuss, G. (1995) 'Cell type-specific expression of hnRNP proteins', Exp Cell Res, 221(1), pp. 187-96.

Kan, Z., Garrett-Engele, P.W., Johnson, J.M. and Castle, J.C. (2005) 'Evolutionarily conserved and diverged alternative splicing events show different expression and functional profiles', Nucleic Acids Res, 33(17), pp. 5659-66.

Kang, Y.K., Guermah, M., Yuan, C.X. and Roeder, R.G. (2002) 'The TRAP/Mediator coactivator complex interacts directly with estrogen receptors alpha and beta through the TRAP220 subunit and directly enhances estrogen receptor function in vitro', Proc Natl Acad Sci U S A, 99(5), pp. 2642-7.

Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D. and Krainer, A.R. (2007) 'The gene encoding the splicing factor SF2/ASF is a proto-oncogene', Nat Struct Mol Biol, 14(3), pp. 185-93.

Kastan, M.B. and Lim, D.S. (2000) 'The many substrates and functions of ATM', Nat Rev Mol Cell Biol, 1(3), pp. 179-86.

Kataoka, N., Bachorik, J.L. and Dreyfuss, G. (1999) 'Transportin-SR, a nuclear import receptor for SR proteins', J Cell Biol, 145(6), pp. 1145-52.

Kaufmann, S.H., Desnoyers, S., Ottaviano, Y., Davidson, N.E. and Poirier, G.G. (1993) 'Specific proteolytic cleavage of poly(ADP-ribose) polymerase: an early marker of chemotherapy-induced apoptosis', Cancer Res, 53(17), pp. 3976-85.

Keren, H., Lev-Maor, G. and Ast, G. (2010) 'Alternative splicing and evolution: diversification, exon definition and function', Nat Rev Genet, 11(5), pp. 345-55.

Kerscher, O. (2007) 'SUMO junction-what's your function? New insights through SUMO-interacting motifs', EMBO Rep, 8(6), pp. 550-5.

Kielkopf, C.L., Lucke, S. and Green, M.R. (2004) 'U2AF homology motifs: protein recognition in the RRM world', Genes Dev, 18(13), pp. 1513-26.

Kim, J.H., Li, H. and Stallcup, M.R. (2003) 'CoCoA, a nuclear receptor coactivator which acts through an N-terminal activation domain of p160 coactivators', Mol Cell, 12(6), pp. 1537-49.

Kim, J.H., Yang, C.K., Heo, K., Roeder, R.G., An, W. and Stallcup, M.R. (2008) 'CCAR1, a key regulator of mediator complex recruitment to nuclear receptor transcription complexes', Mol Cell, 31(4), pp. 510-9.

Kimble, J. and Crittenden, S.L. (2007) 'Controls of germline stem cells, entry into meiosis, and the sperm/oocyte decision in Caenorhabditis elegans', Annu Rev Cell Dev Biol, 23, pp. 405-33.

Kirkby, B., Roman, N., Kobe, B., Kellie, S. and Forwood, J.K. (2010) 'Functional and structural properties of mammalian acyl-coenzyme A thioesterases', Prog Lipid Res, 49(4), pp. 366-77.

Kitagawa, H., Yanagisawa, J., Fuse, H., Ogawa, S., Yogiashi, Y., Okuno, A., Nagasawa, H., Nakajima, T., Matsumoto, T. and Kato, S. (2002) 'Ligand-selective potentiation of rat mineralocorticoid receptor activation function 1 by a CBP-containing histone acetyltransferase complex', Mol Cell Biol, 22(11), pp. 3698-706.

Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Luhrmann, R., Garcia-Blanco, M.A. and Manley, J.L. (1994) 'Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors', Nature, 368(6467), pp. 119-24.

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. and Ule, J. (2010) 'iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution', Nat Struct Mol Biol, 17(7), pp. 909-15.

Kool, J., van Zaane, W., van der Eb, A.J. and Terleth, C. (2001) 'Down-regulation of T-STAR, a growth inhibitory protein, after SV40-mediated immortalization', Cell Growth Differ, 12(11), pp. 535-41.

Kotani, M., Harada, A., Odawara, J., Azuma, M., Okada, S., Nishiyama, Y., Nakamura, M., Tachibana, T. and Ohkawa, Y. (2010) 'Monoclonal antibody specific for Dhx9/NDHII/RHA', Hybridoma (Larchmt), 29(3), pp. 259-61.

Kouzarides, T. (2007) 'Chromatin modifications and their function', Cell, 128(4), pp. 693-705.

Krause, M., Bear, J.E., Loureiro, J.J. and Gertler, F.B. (2002) 'The Ena/VASP enigma', J Cell Sci, 115(Pt 24), pp. 4721-6.

Kuroda, M.I., Kernan, M.J., Kreber, R., Ganetzky, B. and Baker, B.S. (1991) 'The maleless protein associates with the X chromosome to regulate dosage compensation in Drosophila', Cell, 66(5), pp. 935-47.

Kwiatkowski, A.V., Gertler, F.B. and Loureiro, J.J. (2003) 'Function and regulation of Ena/VASP proteins', Trends Cell Biol, 13(7), pp. 386-92.

Lacadie, S.A. and Rosbash, M. (2005) 'Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast', Mol Cell, 19(1), pp. 65-75.

Lahn, B.T. and Page, D.C. (1997) 'Functional coherence of the human Y chromosome', Science, 278(5338), pp. 675-80.

Lai, M.C., Lin, R.I., Huang, S.Y., Tsai, C.W. and Tarn, W.Y. (2000) 'A human importin-beta family protein, transportin-SR2, interacts with the

phosphorylated RS domain of SR proteins', J Biol Chem, 275(11), pp. 7950-7.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. (2001) 'Initial sequencing and analysis of the human genome', Nature, 409(6822), pp. 860-921.

Larocque, D. and Richard, S. (2005) 'QUAKING KH domain proteins as regulators of glial cell fate and myelination', RNA Biol, 2(2), pp. 37-40.

Lazebnik, Y.A., Kaufmann, S.H., Desnoyers, S., Poirier, G.G. and Earnshaw, W.C. (1994) 'Cleavage of poly(ADP-ribose) polymerase by a proteinase with properties like ICE', Nature, 371(6495), pp. 346-7.

Lee, C.G. and Hurwitz, J. (1992) 'A new RNA helicase isolated from HeLa cells that catalytically translocates in the 3' to 5' direction', J Biol Chem, 267(7), pp. 4398-407.

Lee, C.G. and Hurwitz, J. (1993) 'Human RNA helicase A is homologous to the maleless protein of Drosophila', J Biol Chem, 268(22), pp. 16822-30.

Lee, J. and Burr, J.G. (1999) 'Salpalpha and Salpbeta, growth-arresting homologs of Sam68', Gene, 240(1), pp. 133-47.

Lee, M.S., Henry, M. and Silver, P.A. (1996) 'A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export', Genes Dev, 10(10), pp. 1233-46.

Lewis, J.D. and Tollervey, D. (2000) 'Like attracts like: getting RNA processing together in the nucleus', Science, 288(5470), pp. 1385-9.

Li, Z., Zhang, Y., Li, D. and Feng, Y. (2000) 'Destabilization and mislocalization of myelin basic protein mRNAs in quaking dysmyelination lacking the QKI RNA-binding proteins', J Neurosci, 20(13), pp. 4944-53.

Lin, Q., Taylor, S.J. and Shalloway, D. (1997) 'Specificity and determinants of Sam68 RNA binding. Implications for the biological function of K homology domains', J Biol Chem, 272(43), pp. 27274-80.

Lingenfelter, P.A., Delbridge, M.L., Thomas, S., Hoekstra, H.E., Mitchell, M.J., Graves, J.A. and Disteche, C.M. (2001) 'Expression and conservation of processed copies of the RBMX gene', Mamm Genome, 12(7), pp. 538-45.
Liu, X. and Mertz, J.E. (1995) 'HnRNP L binds a cis-acting RNA sequence element that enables intron-dependent gene expression', Genes Dev, 9(14), pp. 1766-80.

Liu, Y., Bourgeois, C.F., Pang, S., Kudla, M., Dreumont, N., Kister, L., Sun, Y.H., Stevenin, J. and Elliott, D.J. (2009) 'The germ cell nuclear proteins hnRNP G-T and RBMY activate a testis-specific exon', PLoS Genet, 5(11), p. e1000707.

Liu, Y. and Elliott, D.J. (2010) 'Coupling genetics and post-genomic approaches to decipher the cellular splicing code at a systems-wide level', Biochem Soc Trans, 38(Pt 1), pp. 237-41.

Lo, P.C. and Frasch, M. (1997) 'A novel KH-domain protein mediates cell adhesion processes in Drosophila', Dev Biol, 190(2), pp. 241-56.

Lodish, H.F. (2004) Molecular cell biology. 5th edn. New York, N.Y. ; Basingstoke: W. H. Freeman.

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

Los, M., Herr, I., Friesen, C., Fulda, S., Schulze-Osthoff, K. and Debatin, K.M. (1997) 'Cross-resistance of CD95- and drug-induced apoptosis as a consequence of deficient activation of caspases (ICE/Ced-3 proteases)', Blood, 90(8), pp. 3118-29.

Los, M., Wesselborg, S. and Schulze-Osthoff, K. (1999) 'The role of caspases in development, immunity, and apoptotic signal transduction: lessons from knockout mice', Immunity, 10(6), pp. 629-39.

Loya, C.M., Van Vactor, D. and Fulga, T.A. (2010) 'Understanding neuronal connectivity through the post-transcriptional toolkit', Genes Dev, 24(7), pp. 625-35.

Lu, C.K., Lai, Y.C., Lin, Y.F., Chen, H.R. and Chiang, M.K. (2012) 'CCAR1 is required for Ngn3-mediated endocrine differentiation', Biochem Biophys Res Commun, 418(2), pp. 307-12.

Lu, Y., Ryan, S.L., Elliott, D.J., Bignell, G.R., Futreal, P.A., Ellison, D.W., Bailey, S. and Clifford, S.C. (2009) 'Amplification and overexpression of Hsa-miR-30b, Hsa-miR-30d and KHDRBS3 at 8q24.22-q24.23 in medulloblastoma', PLoS One, 4(7), p. e6159.

Lu, Z., Ku, L., Chen, Y. and Feng, Y. (2005) 'Developmental abnormalities of myelin basic protein expression in fyn knock-out brain reveal a role of Fyn in posttranscriptional regulation', J Biol Chem, 280(1), pp. 389-95.

Lukong, K.E., Chang, K.W., Khandjian, E.W. and Richard, S. (2008) 'RNA-binding proteins in human genetic disease', Trends Genet, 24(8), pp. 416-25.

Lukong, K.E., Larocque, D., Tyner, A.L. and Richard, S. (2005) 'Tyrosine phosphorylation of sam68 by breast tumor kinase regulates intranuclear localization and cell cycle progression', J Biol Chem, 280(46), pp. 38639-47.
Lukong, K.E. and Richard, S. (2003) 'Sam68, the KH domain-containing superSTAR', Biochim Biophys Acta, 1653(2), pp. 73-86.

Lukong, K.E. and Richard, S. (2007) 'Targeting the RNA-binding protein Sam68 as a treatment for cancer?', Future Oncol, 3(5), pp. 539-44.

Lunde, B.M., Moore, C. and Varani, G. (2007) 'RNA-binding proteins: modular design for efficient function', Nat Rev Mol Cell Biol, 8(6), pp. 479-90.

Lynch, K.W. (2007) 'Regulation of alternative splicing by signal transduction pathways', Adv Exp Med Biol, 623, pp. 161-74.

Mahadevaiah, S.K., Odorisio, T., Elliott, D.J., Rattigan, A., Szot, M., Laval, S.H., Washburn, L.L., McCarrey, J.R., Cattanach, B.M., Lovell-Badge, R. and Burgoyne, P.S. (1998) 'Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities', Hum Mol Genet, 7(4), pp. 715-27.

Maniatis, T. and Tasic, B. (2002) 'Alternative pre-mRNA splicing and proteome expansion in metazoans', Nature, 418(6894), pp. 236-43.

Manley, J.L. and Tacke, R. (1996) 'SR proteins and splicing control', Genes Dev, 10(13), pp. 1569-79.

Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J.F., Revil, T. and Chabot, B. (2007) 'hnRNP proteins and splicing control', Adv Exp Med Biol, 623, pp. 123-47.

Massiello, A. and Chalfant, C.E. (2006) 'SRp30a (ASF/SF2) regulates the alternative splicing of caspase-9 pre-mRNA and is required for ceramide-responsiveness', J Lipid Res, 47(5), pp. 892-7.

Matera, A.G. and Wang, Z. (2014) 'A day in the life of the spliceosome', Nat Rev Mol Cell Biol, 15(2), pp. 108-21.

Matlin, A.J., Clark, F. and Smith, C.W. (2005) 'Understanding alternative splicing: towards a cellular code', Nat Rev Mol Cell Biol, 6(5), pp. 386-98.

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

Matunis, M.J., Coutavas, E. and Blobel, G. (1996) 'A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating

protein RanGAP1 between the cytosol and the nuclear pore complex', J Cell Biol, 135(6 Pt 1), pp. 1457-70.

Maymon, B.B., Paz, G., Elliott, D.J., Lifschitz-Mercer, B., Yogev, L., Kleiman, S.E., Botchan, A., Hauser, R., Schreiber, L. and Yavetz, H. (2002) 'Localization of the germ cell-specific protein, hnRNP G-T, in testicular biopsies of azoospermic men', Acta Histochem, 104(3), pp. 255-61.

Mazeyrat, S., Saut, N., Mattei, M.G. and Mitchell, M.J. (1999) 'RBMY evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene', Nat Genet, 22(3), pp. 224-6.

McElreavey, K. and Krausz, C. (1999) 'Sex Chromosome Genetics '99. Male infertility and the Y chromosome', Am J Hum Genet, 64(4), pp. 928-33.

Mende, Y., Jakubik, M., Riessland, M., Schoenen, F., Rossbach, K., Kleinridders, A., Kohler, C., Buch, T. and Wirth, B. (2010) 'Deficiency of the splicing factor Sfrs10 results in early embryonic lethality in mice and has no impact on full-length SMN/Smn splicing', Hum Mol Genet, 19(11), pp. 2154-67.

Meng, Z., Jackson, N.L., Choi, H., King, P.H., Emanuel, P.D. and Blume, S.W. (2008) 'Alterations in RNA-binding activities of IRES-regulatory proteins as a mechanism for physiological variability and pathological dysregulation of IGF-IR translational control in human breast tumor cells', J Cell Physiol, 217(1), pp. 172-83.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W. and et al. (1994) 'A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1', Science, 266(5182), pp. 66-71.

Mili, S. and Steitz, J.A. (2004) 'Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses', RNA, 10(11), pp. 1692-4.

Min, H., Chan, R.C. and Black, D.L. (1995) 'The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event', Genes Dev, 9(21), pp. 2659-71.

Missler, M. and Sudhof, T.C. (1998) 'Neurexins: three genes and 1001 products', Trends Genet, 14(1), pp. 20-6.

Mootz, D., Ho, D.M. and Hunter, C.P. (2004) 'The STAR/Maxi-KH domain protein GLD-1 mediates a developmental switch in the translational control of C. elegans PAL-1', Development, 131(14), pp. 3263-72.

Mukherjee, B., Kessinger, C., Kobayashi, J., Chen, B.P., Chen, D.J., Chatterjee, A. and Burma, S. (2006) 'DNA-PK phosphorylates histone H2AX during apoptotic DNA fragmentation in mammalian cells', DNA Repair (Amst), 5(5), pp. 575-90.

Mullan, P.B., Quinn, J.E. and Harkin, D.P. (2006) 'The role of BRCA1 in transcriptional regulation and cell cycle control', Oncogene, 25(43), pp. 5854-63.

Naar, A.M., Beaurang, P.A., Zhou, S., Abraham, S., Solomon, W. and Tjian, R. (1999) 'Composite co-activator ARC mediates chromatin-directed transcriptional activation', Nature, 398(6730), pp. 828-32.

Nagai, K., Oubridge, C., Ito, N., Avis, J. and Evans, P. (1995) 'The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA', Trends Biochem Sci, 20(6), pp. 235-40.

Najib, S., Martin-Romero, C., Gonzalez-Yanes, C. and Sanchez-Margalet, V. (2005) 'Role of Sam68 as an adaptor protein in signal transduction', Cell Mol Life Sci, 62(1), pp. 36-43.

Nakajima, T., Uchida, C., Anderson, S.F., Lee, C.G., Hurwitz, J., Parvin, J.D. and Montminy, M. (1997) 'RNA helicase A mediates association of CBP with RNA polymerase II', Cell, 90(6), pp. 1107-12.

Nasim, M.T., Chernova, T.K., Chowdhury, H.M., Yue, B.G. and Eperon, I.C. (2003) 'HnRNP G and Tra2beta: opposite effects on splicing matched by antagonism in RNA binding', Hum Mol Genet, 12(11), pp. 1337-48.

Nayler, O., Cap, C. and Stamm, S. (1998) 'Human transformer-2-beta gene (SFRS10): complete nucleotide sequence, chromosomal localization, and generation of a tissue-specific isoform', Genomics, 53(2), pp. 191-202.

Newman, A. (2001) 'Molecular biology. RNA enzymes for RNA splicing', Nature, 413(6857), pp. 695-6.

Ngo, J.C., Chakrabarti, S., Ding, J.H., Velazquez-Dones, A., Nolen, B., Aubol, B.E., Adams, J.A., Fu, X.D. and 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(1), pp. 77-89.

Nicholson, D.W., Ali, A., Thornberry, N.A., Vaillancourt, J.P., Ding, C.K., Gallant, M., Gareau, Y., Griffin, P.R., Labelle, M., Lazebnik, Y.A. and et al. (1995) 'Identification and inhibition of the ICE/CED-3 protease necessary for mammalian apoptosis', Nature, 376(6535), pp. 37-43.

Niswender, K.D. and Schwartz, M.W. (2003) 'Insulin and leptin revisited: adiposity signals with overlapping physiological and intracellular signaling capabilities', Front Neuroendocrinol, 24(1), pp. 1-10.

Ohtomo, T., Nakao, C., Sumiya, M., Kaminuma, O., Abe, A., Mori, A., Inaba, N., Kato, T. and Yamada, J. (2013) 'Identification of acyl-CoA thioesterase in mouse mesenteric lymph nodes', Biol Pharm Bull, 36(5), pp. 866-71.

Ou, C.Y., Kim, J.H., Yang, C.K. and Stallcup, M.R. (2009) 'Requirement of cell cycle and apoptosis regulator 1 for target gene activation by Wnt and beta-catenin and for anchorage-independent growth of human colon carcinoma cells', J Biol Chem, 284(31), pp. 20629-37.

Pan, Q., Shai, O., Misquitta, C., Zhang, W., Saltzman, A.L., Mohammad, N., Babak, T., Siu, H., Hughes, T.R., Morris, Q.D., Frey, B.J. and Blencowe, B.J. (2004) 'Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform', Mol Cell, 16(6), pp. 929-41.

Paronetto, M.P., Achsel, T., Massiello, A., Chalfant, C.E. and Sette, C. (2007) 'The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x', J Cell Biol, 176(7), pp. 929-39.

Paronetto, M.P., Messina, V., Bianchi, E., Barchi, M., Vogel, G., Moretti, C., Palombi, F., Stefanini, M., Geremia, R., Richard, S. and Sette, C. (2009) 'Sam68 regulates translation of target mRNAs in male germ cells, necessary for mouse spermatogenesis', J Cell Biol, 185(2), pp. 235-49.

Paronetto, M.P., Zalfa, F., Botti, F., Geremia, R., Bagni, C. and Sette, C. (2006) 'The nuclear RNA-binding protein Sam68 translocates to the cytoplasm and associates with the polysomes in mouse spermatocytes', Mol Biol Cell, 17(1), pp. 14-24.

Pasquinelli, A.E., Ernst, R.K., Lund, E., Grimm, C., Zapp, M.L., Rekosh, D., Hammarskjold, M.L. and Dahlberg, J.E. (1997) 'The constitutive transport element (CTE) of Mason-Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway', EMBO J, 16(24), pp. 7500-10.

Pellizzoni, L., Charroux, B., Rappsilber, J., Mann, M. and Dreyfuss, G. (2001) 'A functional interaction between the survival motor neuron complex and RNA polymerase II', J Cell Biol, 152(1), pp. 75-85.

Perez-Canadillas, J.M. and Varani, G. (2001) 'Recent advances in RNA-protein recognition', Curr Opin Struct Biol, 11(1), pp. 53-8.

Pinol-Roma, S., Choi, Y.D., Matunis, M.J. and Dreyfuss, G. (1988) 'Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins', Genes Dev, 2(2), pp. 215-27.

Pollard, V.W. and Malim, M.H. (1998) 'The HIV-1 Rev protein', Annu Rev Microbiol, 52, pp. 491-532.

Proudfoot, N. (1996) 'Ending the message is not so simple', Cell, 87(5), pp. 779-81.

Rajagopalan, L.E., Westmark, C.J., Jarzembowski, J.A. and Malter, J.S. (1998) 'hnRNP C increases amyloid precursor protein (APP) production by stabilizing APP mRNA', Nucleic Acids Res, 26(14), pp. 3418-23.

Rajan, P., Dalgliesh, C., Bourgeois, C.F., Heiner, M., Emami, K., Clark, E.L., Bindereif, A., Stevenin, J., Robson, C.N., Leung, H.Y. and Elliott, D.J. (2009) 'Proteomic identification of heterogeneous nuclear ribonucleoprotein L as a novel component of SLM/Sam68 Nuclear Bodies', BMC Cell Biol, 10, p. 82.

Rajan, P., Gaughan, L., Dalgliesh, C., El-Sherif, A., Robson, C.N., Leung, H.Y. and Elliott, D.J. (2008a) 'Regulation of gene expression by the RNA-binding protein Sam68 in cancer', Biochem Soc Trans, 36(Pt 3), pp. 505-7.

Rajan, P., Gaughan, L., Dalgliesh, C., El-Sherif, A., Robson, C.N., Leung, H.Y. and Elliott, D.J. (2008b) 'The RNA-binding and adaptor protein Sam68 modulates signal-dependent splicing and transcriptional activity of the androgen receptor', J Pathol, 215(1), pp. 67-77.

Rappsilber, J., Ryder, U., Lamond, A.I. and Mann, M. (2002) 'Large-scale proteomic analysis of the human spliceosome', Genome Res, 12(8), pp. 1231-45.

Reddy, T.R. (2000) 'A single point mutation in the nuclear localization domain of Sam68 blocks the Rev/RRE-mediated transactivation', Oncogene, 19(27), pp. 3110-4.

Reddy, T.R., Tang, H., Xu, W. and Wong-Staal, F. (2000) 'Sam68, RNA helicase A and Tap cooperate in the post-transcriptional regulation of human immunodeficiency virus and type D retroviral mRNA', Oncogene, 19(32), pp. 3570-5.

Reddy, T.R., Xu, W., Mau, J.K., Goodwin, C.D., Suhasini, M., Tang, H., Frimpong, K., Rose, D.W. and Wong-Staal, F. (1999) 'Inhibition of HIV replication by dominant negative mutants of Sam68, a functional homolog of HIV-1 Rev', Nat Med, 5(6), pp. 635-42.

Reichman, T.W., Parrott, A.M., Fierro-Monti, I., Caron, D.J., Kao, P.N., Lee, C.G., Li, H. and Mathews, M.B. (2003) 'Selective regulation of gene expression by nuclear factor 110, a member of the NF90 family of double-stranded RNA-binding proteins', J Mol Biol, 332(1), pp. 85-98.

Reitsema, T., Klokov, D., Banath, J.P. and Olive, P.L. (2005) 'DNA-PK is responsible for enhanced phosphorylation of histone H2AX under hypertonic conditions', DNA Repair (Amst), 4(10), pp. 1172-81.

Richard, S., Torabi, N., Franco, G.V., Tremblay, G.A., Chen, T., Vogel, G., Morel, M., Cleroux, P., Forget-Richard, A., Komarova, S., Tremblay, M.L., Li, W., Li, A., Gao, Y.J. and Henderson, J.E. (2005) 'Ablation of the Sam68 RNA

binding protein protects mice from age-related bone loss', PLoS Genet, 1(6), p. e74.

Richard, S., Vogel, G., Huot, M.E., Guo, T., Muller, W.J. and Lukong, K.E. (2008) 'Sam68 haploinsufficiency delays onset of mammary tumorigenesis and metastasis', Oncogene, 27(4), pp. 548-56.

Richard, S., Yu, D., Blumer, K.J., Hausladen, D., Olszowy, M.W., Connelly, P.A. and Shaw, A.S. (1995) 'Association of p62, a multifunctional SH2- and SH3-domain-binding protein, with src family tyrosine kinases, Grb2, and phospholipase C gamma-1', Mol Cell Biol, 15(1), pp. 186-97.

Rocak, S. and Linder, P. (2004) 'DEAD-box proteins: the driving forces behind RNA metabolism', Nat Rev Mol Cell Biol, 5(3), pp. 232-41.

Rosen, E.M., Fan, S. and Ma, Y. (2006) 'BRCA1 regulation of transcription', Cancer Lett, 236(2), pp. 175-85.

Rowell, S., Newman, B., Boyd, J. and King, M.C. (1994) 'Inherited predisposition to breast and ovarian cancer', Am J Hum Genet, 55(5), pp. 861-5.

Saavedra, C., Felber, B. and Izaurralde, E. (1997) 'The simian retrovirus-1 constitutive transport element, unlike the HIV-1 RRE, uses factors required for cellular mRNA export', Curr Biol, 7(9), pp. 619-28.

Samson, M.L. (2008) 'Rapid functional diversification in the structurally conserved ELAV family of neuronal RNA binding proteins', BMC Genomics, 9, p. 392.

Sanchez-Jimenez, F., Perez-Perez, A., Gonzalez-Yanes, C., Varone, C.L. and Sanchez-Margalet, V. (2011) 'Sam68 mediates leptin-stimulated growth by modulating leptin receptor signaling in human trophoblastic JEG-3 cells', Hum Reprod, 26(9), pp. 2306-15.

Sanchez-Margalet, V. and Najib, S. (2001) 'Sam68 is a docking protein linking GAP and PI3K in insulin receptor signaling', Mol Cell Endocrinol, 183(1-2), pp. 113-21.

Scharfe, C., Lu, H.H., Neuenburg, J.K., Allen, E.A., Li, G.C., Klopstock, T., Cowan, T.M., Enns, G.M. and Davis, R.W. (2009) 'Mapping gene associations in human mitochondria using clinical disease phenotypes', PLoS Comput Biol, 5(4), p. e1000374.

Schlegel, B.P., Starita, L.M. and Parvin, J.D. (2003) 'Overexpression of a protein fragment of RNA helicase A causes inhibition of endogenous BRCA1 function and defects in ploidy and cytokinesis in mammary epithelial cells', Oncogene, 22(7), pp. 983-91.

Schmid, R., Grellscheid, S.N., Ehrmann, I., Dalgliesh, C., Danilenko, M., Paronetto, M.P., Pedrotti, S., Grellscheid, D., Dixon, R.J., Sette, C., Eperon, I.C. and Elliott, D.J. (2013) 'The splicing landscape is globally reprogrammed during male meiosis', Nucleic Acids Res.

Schumacher, B., Hanazawa, M., Lee, M.H., Nayak, S., Volkmann, K., Hofmann, E.R., Hengartner, M., Schedl, T. and Gartner, A. (2005) 'Translational repression of C. elegans p53 by GLD-1 regulates DNA damage-induced apoptosis', Cell, 120(3), pp. 357-68.

Sciabica, K.S. and Hertel, K.J. (2006) 'The splicing regulators Tra and Tra2 are unusually potent activators of pre-mRNA splicing', Nucleic Acids Res, 34(22), pp. 6612-20.

Sergeant, K.A., Bourgeois, C.F., Dalgliesh, C., Venables, J.P., Stevenin, J. and Elliott, D.J. (2007) 'Alternative RNA splicing complexes containing the scaffold attachment factor SAFB2', J Cell Sci, 120(Pt 2), pp. 309-19.

Sette, C. (2010) 'Post-translational regulation of star proteins and effects on their biological functions', Adv Exp Med Biol, 693, pp. 54-66.

Sette, C., Messina, V. and Paronetto, M.P. (2010) 'Sam68: a new STAR in the male fertility firmament', J Androl, 31(1), pp. 66-74.

Sherman, L., Sleeman, J., Dall, P., Hekele, A., Moll, J., Ponta, H. and Herrlich, P. (1996) 'The CD44 proteins in embryonic development and in cancer', Curr Top Microbiol Immunol, 213 (Pt 1), pp. 249-69.

Sherman, L., Wainwright, D., Ponta, H. and Herrlich, P. (1998) 'A splice variant of CD44 expressed in the apical ectodermal ridge presents

fibroblast growth factors to limb mesenchyme and is required for limb outgrowth', Genes Dev, 12(7), pp. 1058-71.

Shetty, S. (2005) 'Regulation of urokinase receptor mRNA stability by hnRNP C in lung epithelial cells', Mol Cell Biochem, 272(1-2), pp. 107-18.

Shin, K.H., Kim, R.H., Kang, M.K., Kim, S.G., Lim, P.K., Yochim, J.M., Baluda, M.A. and Park, N.H. (2007) 'p53 promotes the fidelity of DNA end-joining activity by, in part, enhancing the expression of heterogeneous nuclear ribonucleoprotein G', DNA Repair (Amst), 6(6), pp. 830-40.

Singh, R. and Valcarcel, J. (2005) 'Building specificity with nonspecific RNA-binding proteins', Nat Struct Mol Biol, 12(8), pp. 645-53.

Siomi, H. and Dreyfuss, G. (1997) 'RNA-binding proteins as regulators of gene expression', Curr Opin Genet Dev, 7(3), pp. 345-53.

Siomi, H., Matunis, M.J., Michael, W.M. and Dreyfuss, G. (1993) 'The pre-mRNA binding K protein contains a novel evolutionarily conserved motif', Nucleic Acids Res, 21(5), pp. 1193-8.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.P., Waterston, R.H., Wilson, R.K., Rozen, S. and Page, D.C. (2003) 'The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes', Nature, 423(6942), pp. 825-37.

Skrisovska, L., Bourgeois, C.F., Stefl, R., Grellscheid, S.N., Kister, L., Wenter, P., Elliott, D.J., Stevenin, J. and Allain, F.H. (2007) 'The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction', EMBO Rep, 8(4), pp. 372-9.

Smith, H.C. and Sowden, M.P. (1996) 'Base-modification mRNA editing through deamination--the good, the bad and the unregulated', Trends Genet, 12(10), pp. 418-24.

Soulard, M., Barque, J.P., Della Valle, V., Hernandez-Verdun, D., Masson, C., Danon, F. and Larsen, C.J. (1991) 'A novel 43-kDa glycoprotein is detected in the nucleus of mammalian cells by autoantibodies from dogs with autoimmune disorders', Exp Cell Res, 193(1), pp. 59-71.

Soulard, M., Della Valle, V., Siomi, M.C., Pinol-Roma, S., Codogno, P., Bauvy, C., Bellini, M., Lacroix, J.C., Monod, G., Dreyfuss, G. and et al. (1993) 'hnRNP G: sequence and characterization of a glycosylated RNA-binding protein', Nucleic Acids Res, 21(18), pp. 4210-7.

Spector, D.L. (2001) 'Nuclear domains', J Cell Sci, 114(Pt 16), pp. 2891-3.
Stallcup, M.R., Kim, J.H., Teyssier, C., Lee, Y.H., Ma, H. and Chen, D. (2003) 'The roles of protein-protein interactions and protein methylation in transcriptional activation by nuclear receptors and their coactivators', J Steroid Biochem Mol Biol, 85(2-5), pp. 139-45.

Stamm, S. (2002) 'Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome', Hum Mol Genet, 11(20), pp. 2409-16.

Starita, L.M. and Parvin, J.D. (2003) 'The multiple nuclear functions of BRCA1: transcription, ubiquitination and DNA repair', Curr Opin Cell Biol, 15(3), pp. 345-50.

Stoilov, P., Daoud, R., Nayler, O. and Stamm, S. (2004) 'Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA', Hum Mol Genet, 13(5), pp. 509-24.

Stojdl, D.F. and Bell, J.C. (1999) 'SR protein kinases: the splice of life', Biochem Cell Biol, 77(4), pp. 293-8.

Stoss, O., Novoyatleva, T., Gencheva, M., Olbrich, M., Benderska, N. and Stamm, S. (2004) 'p59(fyn)-mediated phosphorylation regulates the activity of the tissue-specific splicing factor rSLM-1', Mol Cell Neurosci, 27(1), pp. 8-21.

Stoss, O., Olbrich, M., Hartmann, A.M., Konig, H., Memmott, J., Andreadis, A. and Stamm, S. (2001) 'The STAR/GSG family protein rSLM-2 regulates the selection of alternative splice sites', J Biol Chem, 276(12), pp. 8665-73.

Stoss, O., Stoilov, P., Hartmann, A.M., Nayler, O. and Stamm, S. (1999) 'The in vivo minigene approach to analyze tissue-specific splicing', Brain Res Brain Res Protoc, 4(3), pp. 383-94.

Sun, W., Xing, B., Sun, Y., Du, X., Lu, M., Hao, C., Lu, Z., Mi, W., Wu, S., Wei, H., Gao, X., Zhu, Y., Jiang, Y., Qian, X. and He, F. (2007) 'Proteome analysis of hepatocellular carcinoma by two-dimensional difference gel electrophoresis: novel protein markers in hepatocellular carcinoma tissues', Mol Cell Proteomics, 6(10), pp. 1798-808.

Sutherland, H.G., Mumford, G.K., Newton, K., Ford, L.V., Farrall, R., Dellaire, G., Caceres, J.F. and Bickmore, W.A. (2001) 'Large-scale identification of mammalian proteins localized to nuclear sub-compartments', Hum Mol Genet, 10(18), pp. 1995-2011.

Tabernero, C., Zolotukhin, A.S., Valentin, A., Pavlakis, G.N. and Felber, B.K. (1996) 'The posttranscriptional control element of the simian retrovirus type 1 forms an extensive RNA secondary structure necessary for its function', J Virol, 70(9), pp. 5998-6011.

Tacke, R., Tohyama, M., Ogawa, S. and Manley, J.L. (1998) 'Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing', Cell, 93(1), pp. 139-48.

Tang, H., Gaietta, G.M., Fischer, W.H., Ellisman, M.H. and Wong-Staal, F. (1997) 'A cellular cofactor for the constitutive transport element of type D retrovirus', Science, 276(5317), pp. 1412-5.

Tang, H., McDonald, D., Middlesworth, T., Hope, T.J. and Wong-Staal, F. (1999) 'The carboxyl terminus of RNA helicase A contains a bidirectional nuclear transport domain', Mol Cell Biol, 19(5), pp. 3540-50.

Tang, H. and Wong-Staal, F. (2000) 'Specific interaction between RNA helicase A and Tap, two cellular proteins that bind to the constitutive transport element of type D retrovirus', J Biol Chem, 275(42), pp. 32694-700.

Tange, T.O., Damgaard, C.K., Guth, S., Valcarcel, J. and Kjems, J. (2001) 'The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element', Embo J, 20(20), pp. 5748-58.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D. and Patel, D.J. (2007) 'How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers', Nat Struct Mol Biol, 14(11), pp. 1025-40.

Taylor, S.J., Resnick, R.J. and Shalloway, D. (2004) 'Sam68 exerts separable effects on cell cycle progression and apoptosis', BMC Cell Biol, 5, p. 5.

Taylor, S.J. and Shalloway, D. (1994) 'An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis', Nature, 368(6474), pp. 867-71.

Terns, M.P. and Terns, R.M. (2001) 'Macromolecular complexes: SMN--the master assembler', Curr Biol, 11(21), pp. R862-4.

Thiery, J.P. (2002) 'Epithelial-mesenchymal transitions in tumour progression', Nat Rev Cancer, 2(6), pp. 442-54.

Tsai, M.J. and O'Malley, B.W. (1994) 'Molecular mechanisms of action of steroid/thyroid receptor superfamily members', Annu Rev Biochem, 63, pp. 451-86.

Tsend-Ayush, E., O'Sullivan, L.A., Grutzner, F.S., Onnebo, S.M., Lewis, R.S., Delbridge, M.L., Marshall Graves, J.A. and Ward, A.C. (2005) 'RBMX gene is essential for brain development in zebrafish', Dev Dyn, 234(3), pp. 682-8.

Tsuda, K., Someya, T., Kuwasako, K., Takahashi, M., He, F., Unzai, S., Inoue, M., Harada, T., Watanabe, S., Terada, T., Kobayashi, N., Shirouzu, M., Kigawa, T., Tanaka, A., Sugano, S., Guntert, P., Yokoyama, S. and Muto, Y. (2011) 'Structural basis for the dual RNA-recognition modes of human Tra2-beta RRM', Nucleic Acids Res, 39(4), pp. 1538-53.

Turner, J.M., Mahadevaiah, S.K., Ellis, P.J., Mitchell, M.J. and Burgoyne, P.S. (2006) 'Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids', Dev Cell, 10(4), pp. 521-9.

Turner, J.M., Mahadevaiah, S.K., Fernandez-Capetillo, O., Nussenzweig, A., Xu, X., Deng, C.X. and Burgoyne, P.S. (2005) 'Silencing of unsynapsed meiotic chromosomes in the mouse', Nat Genet, 37(1), pp. 41-7.

Urbanelli, L., Massini, C., Emiliani, C., Orlacchio, A. and Bernardi, G. (2006) 'Characterization of human Enah gene', Biochim Biophys Acta, 1759(1-2), pp. 99-107.

Valcarcel, J. and Green, M.R. (1996) 'The SR protein family: pleiotropic functions in pre-mRNA splicing', Trends Biochem Sci, 21(8), pp. 296-301.

Valdiglesias, V., Giunta, S., Fenech, M., Neri, M. and Bonassi, S. (2013) 'gammaH2AX as a marker of DNA double strand breaks and genomic instability in human population studies', Mutat Res.

van Eekelen, C.A., Riemen, T. and van Venrooij, W.J. (1981) 'Specificity in the interaction of hnRNA and mRNA with proteins as revealed by in vivo cross linking', FEBS Lett, 130(2), pp. 223-6.

Velazquez-Dones, A., Hagopian, J.C., Ma, C.T., Zhong, X.Y., Zhou, H., Ghosh, G., Fu, X.D. and Adams, J.A. (2005) 'Mass spectrometric and kinetic analysis of ASF/SF2 phosphorylation by SRPK1 and Clk/Sty', J Biol Chem, 280(50), pp. 41761-8.

Venables, J.P. (2002) 'Alternative splicing in the testes', Curr Opin Genet Dev, 12(5), pp. 615-19.

Venables, J.P., Bourgeois, C.F., Dalgliesh, C., Kister, L., Stevenin, J. and Elliott, D.J. (2005) 'Up-regulation of the ubiquitous alternative splicing factor Tra2beta causes inclusion of a germ cell-specific exon', Hum Mol Genet, 14(16), pp. 2289-303.

Venables, J.P., Dalgliesh, C., Paronetto, M.P., Skitt, L., Thornton, J.K., Saunders, P.T., Sette, C., Jones, K.T. and Elliott, D.J. (2004) 'SIAH1 targets the alternative splicing factor T-STAR for degradation by the proteasome', Hum Mol Genet, 13(14), pp. 1525-34.

Venables, J.P., Elliott, D.J., Makarova, O.V., Makarov, E.M., Cooke, H.J. and Eperon, I.C. (2000) 'RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing', Hum Mol Genet, 9(5), pp. 685-94.

Venables, J.P., Klinck, R., Koh, C., Gervais-Bird, J., Bramard, A., Inkel, L., Durand, M., Couture, S., Froehlich, U., Lapointe, E., Lucier, J.F., Thibault, P.,

Rancourt, C., Tremblay, K., Prinos, P., Chabot, B. and Elela, S.A. (2009) 'Cancer-associated regulation of alternative splicing', Nat Struct Mol Biol, 16(6), pp. 670-6.

Venables, J.P., Vernet, C., Chew, S.L., Elliott, D.J., Cowmeadow, R.B., Wu, J., Cooke, H.J., Artzt, K. and Eperon, I.C. (1999) 'T-STAR/ETOILE: a novel relative of SAM68 that interacts with an RNA-binding protein implicated in spermatogenesis', Hum Mol Genet, 8(6), pp. 959-69.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al. (2001) 'The sequence of the human genome', Science, 291(5507), pp. 1304-51.

Vernet, C. and Artzt, K. (1997) 'STAR, a gene family involved in signal transduction and activation of RNA', Trends Genet, 13(12), pp. 479-84.

Volk, T., Israeli, D., Nir, R. and Toledano-Katchalski, H. (2008) 'Tissue development and RNA control: "HOW" is it coordinated?', Trends Genet, 24(2), pp. 94-101.

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

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) 'Alternative isoform regulation in human tissue transcriptomes', Nature, 456(7221), pp. 470-6.

Wang, P.J., Page, D.C. and McCarrey, J.R. (2005) 'Differential expression of sex-linked and autosomal germ-cell-specific genes during spermatogenesis in the mouse', Hum Mol Genet, 14(19), pp. 2911-8.

Wang, Y., Wang, J., Gao, L., Stamm, S. and Andreadis, A. (2011) 'An SRp75/hnRNPG complex interacting with hnRNPE2 regulates the 5' splice site of tau exon 10, whose misregulation causes frontotemporal dementia', Gene, 485(2), pp. 130-8.

Ward, I.M. and Chen, J. (2001) 'Histone H2AX is phosphorylated in an ATR-dependent manner in response to replicational stress', J Biol Chem, 276(51), pp. 47759-62.

Ward, I.M., Minn, K. and Chen, J. (2004) 'UV-induced ataxia-telangiectasia-mutated and Rad3-related (ATR) activation requires replication stress', J Biol Chem, 279(11), pp. 9677-80.

Warf, M.B. and Berglund, J.A. (2010) 'Role of RNA structure in regulating pre-mRNA splicing', Trends Biochem Sci, 35(3), pp. 169-78.

Warzecha, C.C., Sato, T.K., Nabet, B., Hogenesch, J.B. and Carstens, R.P. (2009) 'ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing', Mol Cell, 33(5), pp. 591-601.

Watermann, D.O., Tang, Y., Zur Hausen, A., Jager, M., Stamm, S. and Stickeler, E. (2006) 'Splicing factor Tra2-beta1 is specifically induced in breast cancer and regulates alternative splicing of the CD44 gene', Cancer Res, 66(9), pp. 4774-80.

Wei, X., Pacyna-Gengelbach, M., Schluns, K., An, Q., Gao, Y., Cheng, S. and Petersen, I. (2004) 'Analysis of the RNA helicase A gene in human lung cancer', Oncol Rep, 11(1), pp. 253-8.

Weighardt, F., Biamonti, G. and Riva, S. (1996) 'The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism', Bioessays, 18(9), pp. 747-56.

Westerveld, G.H., Gianotten, J., Leschot, N.J., van der Veen, F., Repping, S. and Lombardi, M.P. (2004) 'Heterogeneous nuclear ribonucleoprotein G-T (HNRNP G-T) mutations in men with impaired spermatogenesis', Mol Hum Reprod, 10(4), pp. 265-9.

Wu, J.I., Reed, R.B., Grabowski, P.J. and Artzt, K. (2002) 'Function of quaking in myelination: regulation of alternative splicing', Proc Natl Acad Sci U S A, 99(7), pp. 4233-8.

Wu, J.Y. and Maniatis, T. (1993) 'Specific interactions between proteins implicated in splice site selection and regulated alternative splicing', Cell, 75(6), pp. 1061-70.

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

Xiong, Y., Zhang, M., Hong, Y., Wei, M., Ai, D., Meng, P., Han, Y., Fu, Z., Shi, Y., Yang, J. and Lin, J. (2013) 'Characterization Analysis of Plasma Membrane Repair Relative Gene Myoferlin', PLoS One, 8(6), p. e66396.

Xu, X.L., Chen, S.Z., Chen, W., Zheng, W.H., Xia, X.H., Yang, H.J., Li, B. and Mao, W.M. (2013) 'The impact of cyclin D1 overexpression on the prognosis of ER-positive breast cancers: a meta-analysis', Breast Cancer Res Treat.

Yang, J., Mani, S.A., Donaher, J.L., Ramaswamy, S., Itzykson, R.A., Come, C., Savagner, P., Gitelman, I., Richardson, A. and Weinberg, R.A. (2004) 'Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis', Cell, 117(7), pp. 927-39.

Yeakley, J.M., Tronchere, H., Olesen, J., Dyck, J.A., Wang, H.Y. and Fu, X.D. (1999) 'Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors', J Cell Biol, 145(3), pp. 447-55.

Yeo, G., Holste, D., Kreiman, G. and Burge, C.B. (2004) 'Variation in alternative splicing across human tissues', Genome Biol, 5(10), p. R74.

Yeo, G.W., Van Nostrand, E., Holste, D., Poggio, T. and Burge, C.B. (2005) 'Identification and analysis of alternative splicing events conserved in human and mouse', Proc Natl Acad Sci U S A, 102(8), pp. 2850-5.

Yoon, D.W., Lee, H., Seol, W., DeMaria, M., Rosenzweig, M. and Jung, J.U. (1997) 'Tap: a novel cellular protein that interacts with tip of herpesvirus saimiri and induces lymphocyte aggregation', Immunity, 6(5), pp. 571-82.

Yoshida, K. and Miki, Y. (2004) 'Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage', Cancer Sci, 95(11), pp. 866-71.

Zaffran, S., Astier, M., Gratecos, D. and Semeriva, M. (1997) 'The held out wings (how) Drosophila gene encodes a putative RNA-binding protein involved in the control of muscular and cardiac activity', Development, 124(10), pp. 2087-98.

Zarnack, K., Konig, J., Tajnik, M., Martincorena, I., Eustermann, S., Stevant, I., Reyes, A., Anders, S., Luscombe, N.M. and Ule, J. (2013) 'Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements', Cell, 152(3), pp. 453-66.

Zhang, J., Cho, S.J., Shu, L., Yan, W., Guerrero, T., Kent, M., Skorupski, K., Chen, H. and Chen, X. (2011) 'Translational repression of p53 by RNPC1, a p53 target overexpressed in lymphomas', Genes Dev, 25(14), pp. 1528-43.
Zhang, S. and Grosse, F. (1997) 'Domain structure of human nuclear DNA helicase II (RNA helicase A)', J Biol Chem, 272(17), pp. 11487-94.

Zhang, S. and Grosse, F. (2004) 'Multiple functions of nuclear DNA helicase II (RNA helicase A) in nucleic acid metabolism', Acta Biochim Biophys Sin (Shanghai), 36(3), pp. 177-83.

Zhang, S., Herrmann, C. and Grosse, F. (1999a) 'Nucleolar localization of murine nuclear DNA helicase II (RNA helicase A)', J Cell Sci, 112 (Pt 16), pp. 2693-703.

Zhang, S., Herrmann, C. and Grosse, F. (1999b) 'Pre-mRNA and mRNA binding of human nuclear DNA helicase II (RNA helicase A)', J Cell Sci, 112 (Pt 7), pp. 1055-64.

Zhang, S., Maacke, H. and Grosse, F. (1995) 'Molecular cloning of the gene encoding nuclear DNA helicase II. A bovine homologue of human RNA helicase A and Drosophila Mle protein', J Biol Chem, 270(27), pp. 16422-7.

Zhang, Y., Lu, Z., Ku, L., Chen, Y., Wang, H. and Feng, Y. (2003) 'Tyrosine phosphorylation of QKI mediates developmental signals to regulate mRNA metabolism', Embo J, 22(8), pp. 1801-10.

Zheng, Z.M. (2004) 'Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression', J Biomed Sci, 11(3), pp. 278-94.

Zhong, X. and Safa, A.R. (2004) 'RNA helicase A in the MEF1 transcription factor complex up-regulates the MDR1 gene in multidrug-resistant cancer cells', J Biol Chem, 279(17), pp. 17134-41.

Zhou, K., Choe, K.T., Zaidi, Z., Wang, Q., Mathews, M.B. and Lee, C.G. (2003) 'RNA helicase A interacts with dsDNA and topoisomerase IIalpha', Nucleic Acids Res, 31(9), pp. 2253-60.

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

Zolotukhin, A.S., Valentin, A., Pavlakis, G.N. and Felber, B.K. (1994) 'Continuous propagation of RRE(-) and Rev(-)RRE(-) human immunodeficiency virus type 1 molecular clones containing a cis-acting element of simian retrovirus type 1 in human peripheral blood lymphocytes', J Virol, 68(12), pp. 7944-52.