Identification and characterisation of RNA targets of the RNA binding proteins Tra2α and Tra2β

Andrew Best

Institute of Genetic Medicine

A thesis submitted for the degree of

Doctor of Philosophy

September 2014

Declaration

I, Andrew Best, 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.

……..

Andrew Best

Acknowledgements

First and foremost, I would like to extend my thanks to a fantastic supervisor - Professor David Elliott – for his guidance, patience and support. David has been a kind and enthusiastic mentor. I would also like to thank my second supervisor – Dr Alison Tyson-Capper – for her constant encouragement and support throughout this project.

I would like to thank all members of the Elliott lab, both past and present. I would particularly like to thank Mrs Caroline Dalgliesh (Institute of Genetic Medicine, Newcastle University) for her valuable advice and for the countless 'quick questions' she has kindly answered over the years. Thanks to Dr Sushma Grellscheid (School of Biological and Biomedical Sciences, Durham University) for sharing her valuable expertise when I first started working in the lab. I would also like to thank Dr Jennifer Munkley, Dr Elaine Hong and Dr Ingrid Ehrmann for their help and advice. You have all made my PhD an enjoyable and rewarding experience. I extend a special thank you to my good friends and fellow PhD students Ms Marina Danilenko, Mr Morten Ritso and Dr Mahsa Kheirollahi-Kouhestani. I wish you all success and happiness in the future.

I am indebted to many collaborators for their time and generosity throughout this project. Without them this work would not have been possible. Thank you to the following collaborators: Dr Julian Venables (Institute of Genetic Medicine, Newcastle University), Dr Katherine James (School of Computing Science, Newcastle University), Professor Tomaz Curk (Faculty of Computer and Information Science, University of Ljubljana), Mr Yaobo Xu (Institute of Genetic Medicine, Newcastle University), Dr Mauro Santibanez-Koref (Institute of Genetic Medicine, Newcastle University), Mr Rafiq Hussain (Institute of Genetic Medicine, Newcastle University), Professor Bernard Keavney (Institute of Cardiovascular Sciences, University of Manchester), Dr Roscoe Klinck (Université de Sherbrooke, Québec, Canada), Professor Benoit Chabot (Université de Sherbrooke, Québec, Canada), Dr Ian Cowell (Institute for Cell and Molecular Biosciences, Newcastle University), Mr Ka Cheong Lee (Institute for Cell and Molecular Biosciences, Newcastle University), Professor Caroline Austin (Institute for Cell and Molecular Biosciences, Newcastle University), Dr Irina Neganova (Institute of Genetic Medicine, Newcastle University) and Mr Ian Dimmick (Institute of Genetic Medicine, Newcastle University). I also gratefully acknowledge the Breast Cancer Campaign for funding this project. Thanks also to my external examiner – Dr Emanuele Buratti – for an enjoyable and constructive discussion.

I would like to extend a big thank you to all my family and friends. Thank you to my mum – Janet – for the unconditional love, support and kindness you have shown to me throughout my life. Thank you to a wonderful sister – Helen – for your friendship and support. A special thanks to my girlfriend – Rebecca – for her continual love, care and support. I am truly grateful to you all.

This thesis is dedicated to the memory of my father – John – who I love and miss dearly. I will always remember your kindness and support. Thank you for always believing in me. This thesis is for you.

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations

ACE	Adenine- and Cytosine-rich Element
AGS	human gastric adenocarcinoma cell line
AR	androgen receptor
ASO	anti-sense oligonucleotide
BPS	Branch Point Sequence
BT20	human breast epithelial carcinoma cell line
cDNA	complementary DNA
CENPF	CENtromere Protein F
CLIP	UV Cross-Linking and ImmunoPrecipitation
CRPCa	Castrate-Resistant Prostate Cancer
DAPI	4',6-diamidino-2-phenylindole
DCIS	Ductal Carcinoma *in situ*
DEPC	diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPS	deoxynucleotide triphosphates
DSB	Double Strand Break
ECL	Enhanced ChemiLuminescence
EDTA	Ethylenediaminetetraacetic acid
EdU	5-ethynyl-2´-deoxyuridine
EJC	exon junction complex
EMSA	electrophoretic mobility shift assay
ER	oestrogen receptor
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
EST	Expressed Sequence Tag
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Serum
GFP	Green Fluorescent Protein
GO	Gene Ontology
HCT116	human colon carcinoma cell line
HEK-293	human embryonic kidney 293 cell line
HeLa	human cervical carcinoma cell line
HER-2	human epidermal growth factor receptor 2
HGP	Human Genome Project
HITS-CLIP	high-throughput sequencing of RNA isolated by CLIP
hnRNP	heterogeneous nuclear ribonucleoprotein
iCLIP	individual nucleotide resolution CLIP
IDC	Invasive Ductal Carcinoma
IP	Immunoprecipitation

ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
KG-1	acute myelogenous leukemia cell line
lncRNA	long non-coding RNA
MBNL1	Muscle-blind like protein
MCF-7	Michigan Cancer Foundation-7 human breast cancer cell line
MDA-MB-231	MD Anderson human invasive breast cancer cell line
miRNA	micro RNA
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger RNA
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
ncRNA	non-coding RNA
NLS	Nuclear Localisation Signal
NMD	Nonsense-Mediated Decay
NMR	Nuclear Magnetic Resonance
NTC	No Template Control
ORF	Open Reading Frame
PAR-CLIP	PhotoActivatable-Ribonucleoside-enhanced CLIP
PBS	Phophate Buffered Saline
PC-3	human prostate cancer cell line
PCa	Prostate Cancer
PCR	Polymerase Chain Reaction
PK	Proteinase K
PNK	Polynucleotide Kinase
PPT	Polypyrimidine Tract
PR	progesterone receptor
PSI	Percentage Splicing Inclusion
PTB	Polypyrimidine Tract Binding protein
PTC	Premature Termination Codon
PVDF	polyvinylidene difluoride
qPCR	quantitative real-time PCR
RBP	RNA-binding protein
RIN	RNA integrity number
RIP	RNA immunoprecipitation
RIP-seq	RNA immunoprecipitation coupled to sequencing
RNA	ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
RRM	RNA recognition motif
rRNA	ribosomal RNA
RS-domain	arginine/serine-rich domain
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SDS	sodium dodecyl sulfate

SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SELEX	Systematic Evolution of Ligands by EXponential enrichment
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
SROOGLE	Splicing RegulatiOn Online GraphicaL Engine
SSO	splice-switching oligonucleotides
TBE	Tris/Borate/EDTA
TBST	Tris-Buffered Saline and Tween-20
TE	Tris/EDTA
TIS	Translation Initiation Site
UCSC	University of California, Santa Cruz
UTR	Untranslated region

Abstract

Alternative splicing – the production of multiple messenger RNA isoforms from a single gene – is regulated in part by RNA binding proteins. The overall aim of this study was to identify and characterise novel targets of the RNA binding proteins Tra2α and Tra2β, in order to further understand their biological functions. Tra2β is implicated in male germ cell development and in the initial stages of this project, I utilised data from a previous Tra2β HITS-CLIP experiment to validate and characterise novel RNA targets from the mouse testis using minigenes. These included a large testis-enriched exon from *Nasp* and a posion exon from *Tra2a*. The identification of a Tra2β-responsive poison exon within the *Tra2a* gene suggested that Tra2β may directly regulate Tra2α protein expression. Subsequent experiments in a human breast cancer cell line revealed that following depletion of Tra2β, Tra2α is up-regulated, and could functionally compensate in splicing regulation. Tra2β is also up-regulated in several human cancers and we hypothesised that Tra2β may regulate alternative splicing programmes of functional importance in cancer. Therefore for the majority of this project, I investigated RNA targets of Tra2α and Tra2β in the human invasive breast cancer cell line MDA-MB-231. Two transcriptome-wide approaches were used to identify RNA targets in this study. Firstly, I used iCLIP to map the transcriptome-wide binding sites of Tra2β in MDA-MB-231 cells. Secondly, I used RNA-seq to investigate the functional effect of joint Tra2 protein depletion on the transcriptome. Combining the iCLIP and RNA-seq data facilitated the identification of target exons which were both directly bound by Tra2β and functionally responsive to Tra2 protein depletion. Unexpectedly, Tra2 protein dependent exons included both alternative and constitutively spliced exons. A Gene Ontology enrichment analysis of the experimentally validated exons revealed that Tra2 protein dependent exons were functionally enriched in genes associated with chromosome biology. These included a functionally important exon from *CHEK1*, which encodes a key DNA damage response protein. Joint depletion of Tra2α and Tra2β led to reduced expression of the full-length CHK1 protein, accumulation of the DNA damage marker γH2AX and decreased cell viability. Together, this data suggests that human Tra2 proteins jointly control constitutive and alternative splicing patterns via paralog compensation which are important for cell viability.

Chapter 1: Introduction

1.1 Pre-mRNA splicing

1.1.1 One gene, multiple mRNAs

The human genome contains approximately 21,000 protein-coding genes (Clamp et al., 2007), far fewer than many had predicted prior to the completion of the Human Genome Project (HGP) in 2004 (Nilsen and Graveley, 2010). Earlier estimates of human gene copy number varied considerably, as different predictions were based on a variety of factors including the average size of a gene, the quantity of unique expressed sequence tags (ESTs) identified and the perceived biological complexity of humans relative to other organisms (Pertea and Salzberg, 2010). Our modest number of protein-coding genes is even more remarkable when considering that the human genome is estimated to produce over one million different protein species (Harrison *et al.*, 2002; Jensen, 2004).

The 'missing information' which generates proteome complexity is explained in part by a series of post-transcriptional modifications which occur throughout the course of gene expression. When information encoded by DNA is expressed as a precursor messenger RNA (pre-mRNA), processing of the primary transcript occurs almost immediately after transcription initiation. Pre-mRNA processing begins with 5' capping, and is closely followed by splicing, transcription termination and polyadenylation. These modifications to the primary transcript allow the creation of multiple messenger RNA (mRNA) isoforms from a single gene (Matlin et al., 2005). As a result, a single gene may encode multiple protein isoforms.

Post-transcriptional modifications such as alternative splicing, alternative polyadenylation and RNA editing expand the number of proteins encoded by the genome enormously (Figure 1.1) (Lander, 2011). As different protein isoforms may be structurally and functionally distinct from one another, post-transcriptional modifications such as alternative splicing dramatically increase the complexity of the proteome from a defined number of genes (Nilsen and Graveley, 2010). When post-transcriptional modifications are combined with subsequent post-translational modifications to proteins (such as phosphorylation and ubiquitylation), the protein coding potential of the genome is massive (Wilhelm et al., 2014).

Alternative splicing is one of the major sources of proteomic diversity in eukaryotes (Nilsen and Graveley, 2010); approximately 95% of human genes are alternatively spliced (Pan et al., 2008). However, it is worth noting that relatively few of these alternative splicing events are highly conserved and not all alternative splicing events are of functional importance. Alternative splicing can also create non-productive transcripts, and therefore alternative splicing provides a key mechanism for post-transcriptional regulation of gene expression (Lareau et al., 2007).

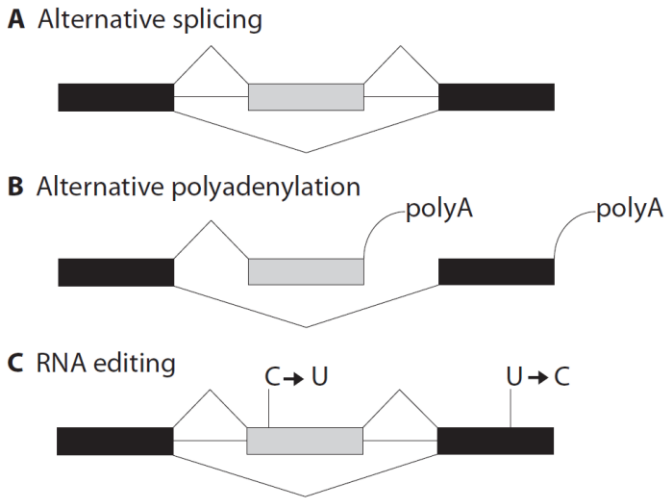


Figure 1.1 Post-transcriptional modifications create multiple mRNA isoforms from a single gene. Post-transcriptional modifications include (A) alternative splicing, (B) alternative polyadenylation and (C) RNA editing. This image is adapted from (Siomi and Dreyfuss, 1997).

1.1.2 General pre-mRNA processing

In the initial stages of eukaryotic gene expression, pre-mRNAs are transcribed from a DNA template by RNA polymerase II during transcription (Hurwitz, 2005). The emerging pre-mRNA transcripts are then subject to a series of processing events known as post-transcriptional modifications, resulting in production of a mature mRNA. Shortly after transcription initiation and 5' capping, components of the splicing machinery bind to the primary transcript, including small nuclear ribonucleoproteins (snRNPs), heterogeneous nuclear ribonucleoproteins (hnRNPs) and a large range of auxiliary RNA-binding proteins (Cramer et al., 2001). Originally, splicing was thought to occur entirely post-transcriptionally. However, it was later recognized that many of these “post”-transcriptional modifications are directly linked to transcription (Neugebauer, 2002) and frequently occur co-transcriptionally (including capping,

splicing and polyadenylation) (Proudfoot et al., 2002). Following transcriptional termination and polyadenylation, mature mRNAs are exported from the nucleus through nuclear pores to the cytoplasm, where they serve as templates for protein synthesis (translation) prior to their eventual degradation (Stewart, 2010).

1.1.3 Splicing and splice site recognition by the spliceosome

Pre-mRNA splicing is the process by which introns are removed from the pre-mRNA whilst exons are simultaneously joined together, forming a continuous protein-coding region (open reading frame, ORF) within the RNA sequence (Matlin et al., 2005). Splicing is performed by a macromolecular complex called the spliceosome. In eukaryotes, the majority of introns are processed by the major spliceosome. In fact, >99% of human introns are excised by the major spliceosome in a process termed canonical splicing (Wahl et al., 2009). However, a less abundant class of non-canonical introns are also processed by distinct splicing machinery termed the minor spliceosome, which recognises distinct splice site sequences from the major spliceosome (Patel and Steitz, 2003).

The splicesome consists of between 150 and 300 individual proteins (Rappsilber et al., 2002), together with a group of small nuclear ribonucleoproteins (snRNPs) which consist of proteins bound to an RNA component (snRNA). However, the number of protein components of the spliceosome varies depending of which biochemical conditions are used experimentally and on how the spliceosome is defined. The major spliceosome consists of five snRNPs; designated U1, U2, U4/U6, and U5, which are essential for spliceosome formation on the RNA transcript (Wahl et al., 2009). Assembly of the spliceosome is guided by two sets of degenerate sequences called the 5' and 3' splice sites (subsequently referred to as 5'ss and 3'ss respectively), which help define the intron/exon boundaries. In humans, the consensus 5'ss sequence recognised by the major spliceosome is CAG**GU**RAGU (where R = purine and GU is almost always invariant). The 3'ss sequences consist of three elements; the consensus 3'ss Y**AG**G (where Y = pyrimidine and AG is almost always invariant), an upstream polypyrimidine tract (PPT) and the branch point sequence (BPS) (see Figure 1.2).

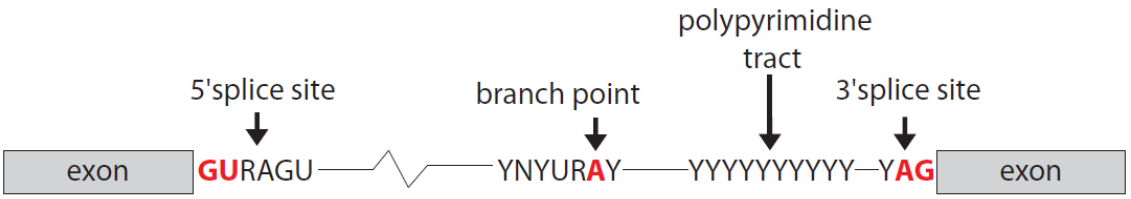


Figure 1.2 The four *cis*-elements of a typical metazoan intron which are recognised by the major spliceosome. Intron recognition is guided by four *cis*-elements; the 5' splice site, the branch point sequence (BPS), the polypyrimadine tract (PPT) and the 3' splice site. (R = purine, Y = pyrimidine, N = any nucleotide). This image is adapted from (Wahl et al., 2009).

During early formation of the major spliceosome, each of the four *cis*-elements are bound by specific spliceosomal components (Valadkhan, 2007). The 5'ss is recognised by U1 snRNP. The branch point sequence (BPS) is initially recognised by splicing factor 1/branch point bound protein (SF1/BBP) and is subsequently replaced by U2 snRNP, whilst the polypyrimidine tract (PPT) and 3'ss interact with two subunits of the U2 auxilliary factor heterodimer (U2AF35 and U2AF65 respectively) (see Figure 1.3) (Selenko et al., 2003). Most of the essential RNA-RNA interactions between the snRNA component of the snRNPs and the pre-mRNA are weak, and often require additional proteins to enhance their stability on the RNA transcript (Wahl et al., 2009).

Spliceosome assembly is also influenced by the size of an intron. In lower eukaryotes such as the fission yeast *Schizosaccharomyces pombe*, coding exons are typically separated by very small introns which are often <100 nucleotides long (Berget, 1995). For small introns such as these, splice site pairs are recognised across the intron; a process termed “intron definition”. However, the average vertebrate gene consists of multiple small exons, separated by significantly larger introns. For large introns, splice site recognition initially occurs across the exon; a process termed “exon definition” (see Figure 1.3). In the exon definition model, U1 snRNP binds to the 5'ss and promotes recognition of the upstream 3'ss across the exon (Schellenberg et al., 2008). Consequently, the splicing reaction may be considered as the process of defining two exons, rather than the process of defining one intron.

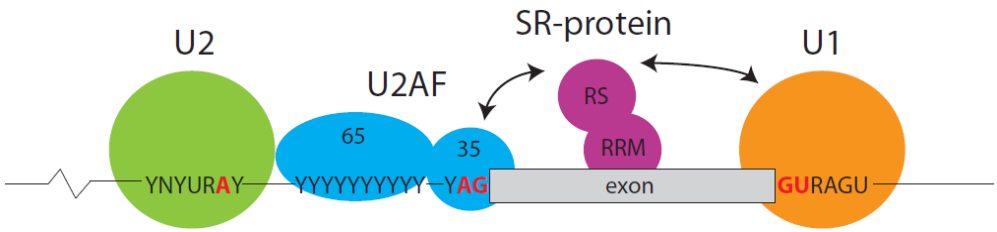


Figure 1.3 During early spliceosome formation, cross-exon splicing complexes form along the pre-mRNA during the process of exon definition. The U1 and U2 snRNPs, together with the two subunits of the U2 auxilliary factor heterodimer (U2AF65 and U2AF35) recognise specific components from the consensus splice sites. In addition to the essential splicing components, other auxiliary RNA-binding proteins (such as SR-proteins and hnRNPs) bind to splicing enhancers and silencers to either enhance or inhibit spliceosome formation. This image is adapted from (Wahl et al., 2009).

1.1.4 The splicing reaction

Mechanistically, splicing occurs as a two-step process, centered around two S_N2-type transesterification reactions (Valadkhan and Jaladat, 2010). Briefly, the first step of splicing involves the 2'OH group of the branchpoint adenosine nucleotide attacking the phosphodiester bond of the 5'ss. This results in cleavage of the 5' exon and generation of a lariat intron, which is temporarily attached to the 3' exon. Subsequently, the 3'OH group of the cleaved 5' exon attacks the 3'ss, resulting in ligation of both exons through another transesterification reaction, whilst the intron is released as a lariat (see Figure 1.4) (Moore and Sharp, 1993; Sharp, 1994; Wahl *et al.*, 2009).

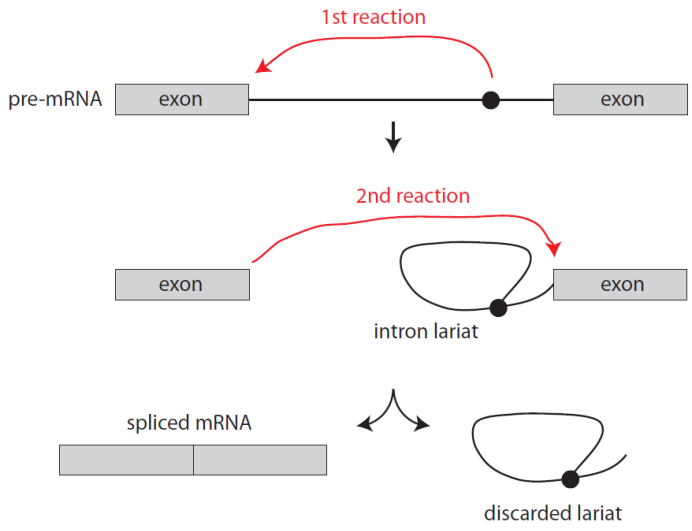


Figure 1.4 The two S_N2-type transesterification reactions of splicing. Reaction 1: the 2'OH group of the branchpoint adenosine nucleotide (black circle) attacks the

phosphodiester bond of the 5'ss, forming the lariat intron. Reaction 2: the 3'OH group of the cleaved 5' exon attacks the 3'ss, resulting in ligation of both exons, whilst the lariat intron is released and subsequently degraded. This image is adapted from (Black, 2003).

Throughout the splicing reaction, the spliceosomal components assemble on the RNA substrate in a series of complexes; termed the H, E, A, B and C splicing complexes in order of formation (see Figure 1.5) (Wahl *et al.*, 2009; Valadkhan and Jaladat, 2010). Complex H is effectively pre-spliceosome assembly, recognising the non-specific association of a group of RNA-binding proteins called the heterogeneous nuclear ribonucleoproteins (hnRNPs) to the primary transcript immediately following transcription (Wahl *et al.*, 2009).

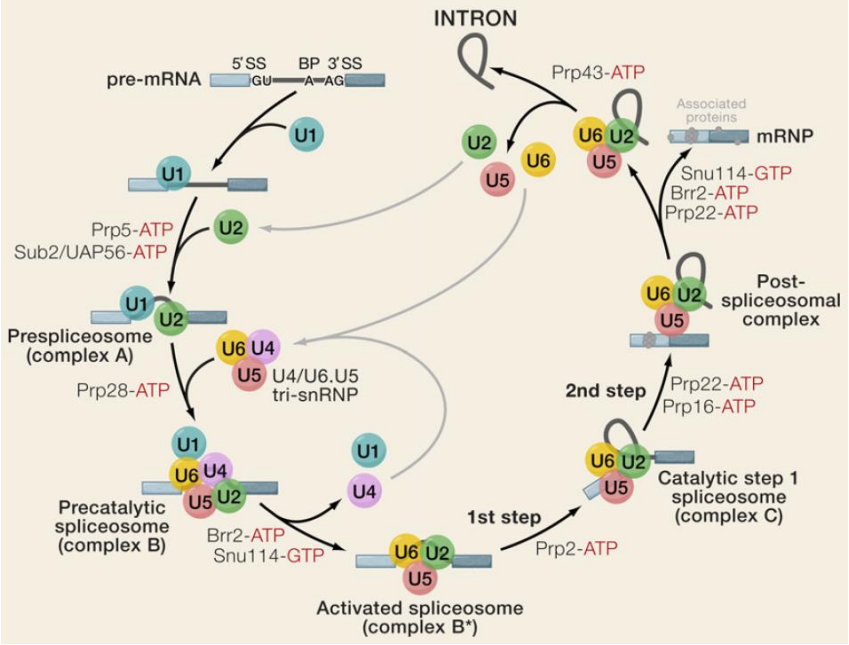


Figure 1.5 Assembly of the major splicesome leading to intron removal from the pre-mRNA. This image is taken from a recent review by (Wahl et al., 2009).

Spliceosome assembly begins with the ATP-independent interaction of U1 snRNP with the 5'ss. This interaction can be stabilised by other auxiliary RNA-binding proteins (including members of the SR-protein family) which interact with the protein component of U1 snRNP. The splicing factor 1/branch point bound protein (SF1/BBP) recognises the branch point sequence (BPS) and interacts with the U2 auxiliary factor (U2AF) heterodimer. U2AF consists of a 65kDa subunit and a 35kDa subunit, which recognise the polypyrimidine tract (PPT) and 3'ss respectively. This formation is referred to as the E complex. Subsequently, SF1/BBP is replaced by U2 snRNP, which

interacts with the BPS in an ATP-dependent manner. U2 snRNP interacts with U1 snRNP, forming complex A. Following this, a tri-snRNP complex composed of U4/U6 and U5 attaches to the spliceosome, forming complex B. Spliceosomal remodelling leads to the dissociation of U1 and U4, allowing the first transesterification reaction to occur. The first transesterification reaction is catalysed by the RNA-dependent ATPase Prp2, forming the lariat intron structure, termed complex C. In the second transesterification reaction, the upstream 5'ss ligates to the 3'ss and the intron lariat is subsequently released. The overall splicing reaction is catalysed by eight, evolutionarily-conserved DExD/H-type RNA-dependent ATPases/helicases which catalyse RNA-RNA rearrangements and spliceosome remodelling events (see Figure 1.5). Following completion of the splicing reaction, the lariat intermediate is degraded and the snRNPs are recycled in successive splicing reactions. Formation of the spliceosome is reviewed in further detail by (Black, 2003; Wahl *et al.*, 2009).

1.1.5 The minor spliceosome

Whilst the vast majority of human introns are removed by the major spliceosome, a less-abundant class of non-canonical introns are removed by the minor spliceosome. The minor spliceosome contains four specific snRNPs; U11, U12, $U4_{atac}$ and $U6_{atac}$, which are functionally analogous to the U1, U2, U4 and U6 components of the major spliceosome (Turunen et al., 2013). Consequently by convention, introns recognised by the major spliceosome are referred to as “U2-type” introns, whilst introns recognised by the minor spliceosome are referred to as “U12-type” introns (Patel and Steitz, 2003). Both spliceosomes share the same U5 snRNP component. U12-type introns were originally defined by distinct 5' and 3' splice site dinucleotide sequences (AT-AC termini rather than GT-AG) and consequently were originally referred to as “atac introns” (Wu and Krainer, 1997). However, it was later discovered that the minor spliceosome recognises both sets of terminal dinucleotide sequences (in fact GT-AG termini are more frequently recognised by the minor spliceosome than AT-AC), and that a highly conserved 5'ss and branch point sequence (BPS) are the most defining features of a U12-type intron (Burge et al., 1998). The minor spliceosome is reviewed in detail by (Patel and Steitz, 2003; Turunen *et al.*, 2013). However for the remainder of this thesis, I will focus on U2-type introns recognised by the major spliceosome.

1.1.6 Alternative splicing

Whilst some exons are always spliced into the mRNA transcript (constitutively spliced), other exons are sometimes included and sometimes skipped (alternatively spliced). When splicing was first discovered in 1977 (Chow *et al.*, 1977; Gelinas and Roberts, 1977), alternative splicing was estimated to occur in less than 1% of human genes. Subsequently, through bioinformatic analyses of expressed sequence tags (ESTs) and the more recent development of RNA sequencing (RNA-seq), that number has continually risen to approximately 95%, indicating alternative splicing events occur in nearly all human multi-exon genes (Martin Dutertre, 2010).

On average, human protein-coding genes produce three different mRNA transcripts via alternative splicing (Djebali et al., 2012). In addition to generating proteome diversity, alternative splicing also plays a significant role in quantitative gene control. Alteration to the open reading frame (ORF) of a gene is estimated to occur in around one third of alternative splicing events, potentially leading to the generation of premature stop codons (PTCs) within the RNA sequence and subsequently targeting the transcript for nonsense-mediated decay (NMD) (Lewis et al., 2003). Alternative splicing of the untranslated regions (UTRs) within a gene can also regulate gene expression through changes to mRNA stability (Matlin et al., 2005). Accurate splicing of the pre-mRNA is therefore crucial to maintain normal cellular physiology. This is exemplified by the estimation that between 15% to 50% of human genetic diseases may arise from mutations to splice sites or splicing regulatory sequences (Cartegni *et al.*, 2002; Faustino and Cooper, 2003).

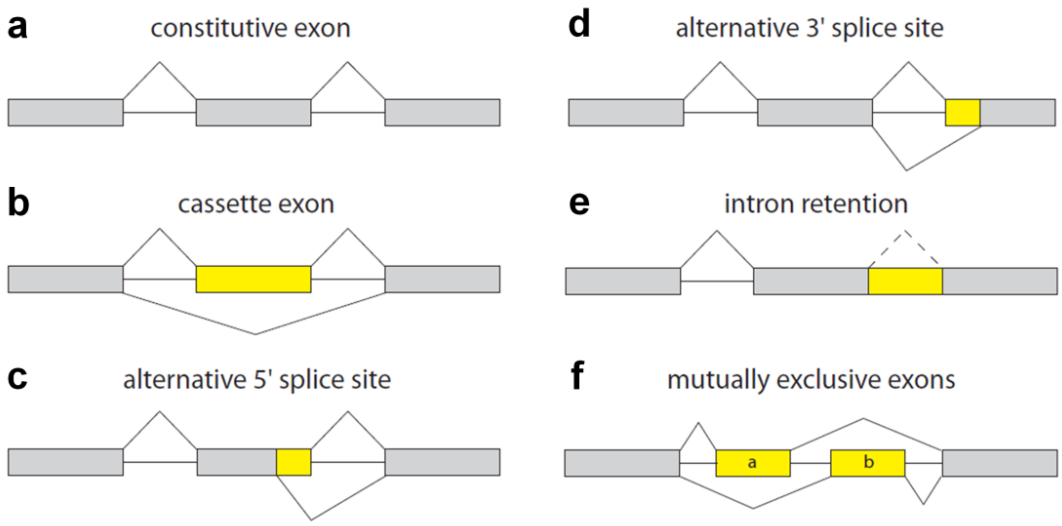


Figure 1.5 Classification of different alternative splicing events. (a) Constitutive exons are spliced into all mRNA transcripts from a gene. (b) Cassette exons (or alternative exons) are sometimes included and sometimes skipped from the mRNA transcript in an independent manner. (c-d) Alternative splice sites compete for recognition with primary splice sites to modify the length of an exon. (e) Intron retentions occur when an intron is not removed from the primary transcript. (f) Mutually exclusive exons are spliced in a coordinated manner, where inclusion of each exon is not independent from the other. This image is adapted from (Srebrow and Kornblihtt, 2006).

1.1.7 Regulation of alternative splicing

Alternative splicing is a highly regulated process. Splice site selection is influenced by both *cis*-acting elements (within the RNA sequence) and *trans*-acting factors which determine splice site selection in a combinational manner (Smith and Valcárcel, 2000). The expression of RNA-binding proteins can vary in a tissue-specific (Venables et al., 2013a) and developmental-stage-specific manner (Matsui et al., 2000), generating different splicing patterns depending on the cellular context. Furthermore, several other factors have been demonstrated to influence splice site selection, including the formation of secondary structures within the pre-mRNA (Warf and Berglund, 2010), the rate of transcription (Cáceres and Kornblihtt, 2002) and epigenetic factors (Luco et al., 2011). Consequently, the regulation of alternative splicing is complex. However, with advances in experimental technology, such as high-throughput sequencing, as well as advanced bioinformatic analyses of splicing regulatory motifs, a “splicing code” is beginning to be deciphered (Barash et al., 2010).

1.1.7.1 *Cis*-regulatory sequences

1.1.7.1.1 Splice site strength

In the initial stage of spliceosome formation, snRNPs are recruited to the pre-mRNA by the presence of the 5' and 3' splice sites. Recognition of the 5' and 3' splice site sequences depends on complementary base-pairing between the pre-mRNA and the snRNA component of the snRNP. Consequently, the extent of the complementarity between the splice site sequence and the snRNA affects the efficiency of splice site recognition by the spliceosome. This degree of complementarity is often quantified as the "splice site strength". Generally speaking, constitutively spliced exons are often flanked by strong splice sites (they are efficiently recognised by the spliceosome), whereas alternative exons are more likely to be flanked by at least one weak splice site (they are less efficiently recognised by the spliceosome), although there are exceptions. A variety of computer programs have been developed to predict the relative strength of splice sites, including the Splicing RegulatiOn Online Graphical Engine (SROOGLE) (Schwartz et al., 2009) which I use in this study.

1.1.7.1.2 Splicing enhancer and silencer sequences

In combination with splice site strength, splice site selection is influenced by other *cis*-acting elements, including splicing enhancer sequences (termed exonic splicing enhancers (ESE) and intronic splicing enhancers (ISE)) and splicing silencer sequences (termed exonic splicing silencers (ESS) and intronic splicing silencers (ISS)). Splicing enhancer and silencer sequences play an essential role in the recognition of both constitutive and alternative exons (Matlin et al., 2005). The splicing enhancer and silencer sequences are recognised by a large number of auxiliary RNA-binding proteins (RBPs) which compete to either enhance or repress splice site recognition (see Figure 1.6). Analyses of sequence conservation indicate that *cis*-regulatory elements are most highly enriched within the regulated exons (exonic) or within close proximity to the splice sites (within 103 nucleotides upstream or 94 nucleotides downstream of regulated exons) (Sorek and Ast, 2003). However, there are also many examples of more distal intronic regulatory elements which can influence splice site selection, such as distal intronic regulatory sequences bound by TIA-1 (Witten and Ule, 2011) and NOVA (Ule et al., 2006).

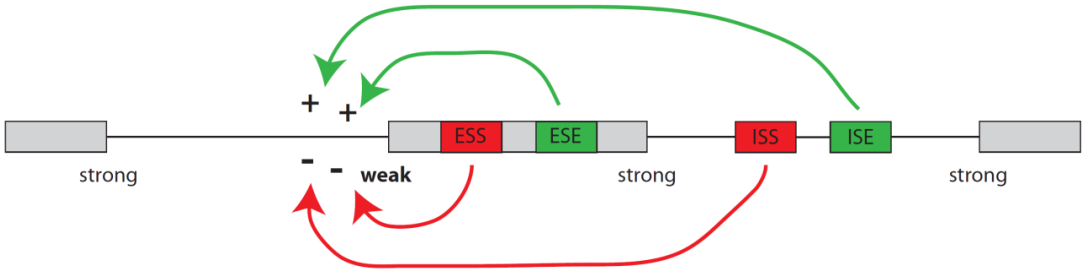


Figure 1.6 *Cis*-elements consisting of splicing enhancers (ESE/ISE) and splicing silencers (ESS/ISS) can have antagonistic effects on splice site selection by the splicesome. RNA-binding proteins which recognise the enhancer and silencer sequences compete to either stabilise or disrupt spliceosome formation respectively. This image is adapted from (Srebrow and Kornblihtt, 2006).

1.1.7.2 *Trans*-regulatory proteins

Trans-acting influences on splicing include the relative concentration, localisation and activity of RNA-binding proteins, such as members of the SR-protein family and the heterogenous nuclear ribonucleoproteins (hnRNPs). SR-proteins are generally considered enhancers of splicing, whilst hnRNPs are often considered splicing repressors. However, the activating or repressive activity of SR-proteins and hnRNPs is largely position-dependent. For example, Erkelenz et al. have recently demonstrated that SR-proteins can activate splicing when recruited to an exon, yet repress splicing inclusion of the same exon when recruited to the downstream intron (Erkelenz et al., 2013). Similarly, hnRNPs were found to exert both activating and repressive activity over a single exon, but in the opposite positions to SR-proteins (Erkelenz et al., 2013).

The number of proteins identified with RNA-binding capacity continues to grow. Remarkably, a recent study of the mRNA-bound proteome in the human embryonic kidney cell line HEK-293 identified nearly 800 mRNA-bound proteins (Baltz et al., 2012). However this thesis will specifically focus on two paralogous RNA-binding proteins; Tra2α and Tra2β.

1.1.8 The SR-protein family

The SR-protein family consist of a group of RNA-binding proteins which share two main structural features; an arginine/serine-rich domain (RS-domain) and one or more RNA-recognition motifs (RRMs) (Shepard and Hertel, 2009a). RS-domains can facilitate protein-protein interactions with other RS-domain-containing proteins, including other SR-proteins, SR-like proteins and components of the spliceosome including the U1 snRNP-associated protein U1-70K and U2AF65 (Wu and Maniatis, 1993; Kohtz *et al.*, 1994; Graveley *et al.*, 2001). RRMs facilitate direct interaction between SR-proteins and short motifs within the RNA sequence. The term "SR-protein" was originally given to a group of RS-domain-containing proteins which were recognised by the monoclonal antibody mAb104 (Roth et al., 1990). These include SF2/ASF (SFRS1), SC35 (SFRS2), SRp20 (SFRS3), SRp75 (SFRS4), SRp40 (SFRS5) and SRp55 (SFRS6). Classical SR-protein family members are defined by four criteria; (1) structural similarity, (2) dual function in both alternative and constitutive splicing, (3) recognition by the monoclonal antibody mAb104 and (4) purification using magnesium chloride (Long and Caceres, 2009).

In addition to the classical group of SR-proteins, a structurally-related group of "SR-like" proteins are also involved in alternative splicing regulation. SR-like proteins share some degree of structural similarity to classical SR-proteins (both groups of proteins contain at least one RS-domain and an RRM) and both have roles in pre-mRNA splicing, yet SR-like proteins do not meet all of the criteria which defines classical SR-proteins. Examples from the SR-like group of proteins include both subunits of the U2AF heterodimer (U2AF65 and U2AF35), as well as the human homologues of the *Drosophila* Tra2 splicing factor; Tra2α and Tra2β (Long and Caceres, 2009).

1.1.9 Regulation of SR-protein activity

SR-proteins regulate alternative splicing in a concentration-dependent and phosphorylation-state-dependent manner. SR-proteins are directly phosphorylated on serine residues located within their RS-domains by a group of SR-protein kinases, including SR-protein kinase 1 (SRPK1) (Zhong et al., 2009) and CDC-like kinase 1 (CLK1) (Prasad and Manley, 2003), which can modulate SR-protein activity. In addition, recent evidence has also suggested a role for long non-coding RNAs (lncRNAs) in the

regulation of SR-protein activity. The nuclear-retained lncRNA MALAT1 directly interacts with SR-proteins and can regulate their localisation within nuclear speckles (Tripathi *et al.*, 2010a). Furthermore, lncRNAs may also modulate the phosphorylation state of SR-proteins, suggesting that lncRNAs may have a significant role in alternative splicing regulation via modulation of SR-protein activity (Tripathi *et al.*, 2010a).

1.1.10 Models of splicing regulation by SR-proteins

To date, a number of different models have been proposed to explain the mechanisms of SR-protein-mediated splicing activation and repression (see Figure 1.7, A-D). The "U2AF recruitment model" is the classical model of SR-protein splicing enhancer function. This model describes the ability of SR-proteins to directly bind to exonic splicing enhancer (ESE) sequences and stabilise interactions between the spliceosome and the consensus splice sites, specifically the interactions of U1 snRNP and U2AF65 with the 5' and 3' splice sites respectively (see Figure 1.7 A) (Robberson *et al.*, 1990; Graveley *et al.*, 2001). The U2AF recruitment model is therefore closely linked to the process of exon definition.

The "inhibitor model" hypothesises that SR-proteins may also act through antagonism of splicing repressor proteins, such as the hnRNPs. Recruitment of SR-proteins to splicing enhancer sequences may block splicing repressor proteins from binding to nearby splicing silencer sequences, preventing them from disrupting spliceosome formation (see Figure 1.7 B). Therefore in the inhibitor model, splicing activators and splicing repressors act in a competitive, antagonistic manner (Zhu *et al.*, 2001; Martinez-Contreras, 2007).

A "splicing repressor model" has also been proposed to explain the splicing repressor activity of SR-proteins. When SR-proteins bind to intronic regions close to the splice sites of regulated exons, recruitment of the spliceosomal components (such as the U2AF heterodimer) may be inhibited by steric hinderance, leading to non-productive spliceosome assembly (see Figure 1.7 C) (Shepard and Hertel, 2009b; Erkelenz *et al.*, 2013).

Finally, in the "co-activator model", SR-proteins have also been proposed to enhance splicing activation independent of direct RNA-binding, through multiple potential protein-protein interactions (see Figure 1.7 D). For example, SR-proteins may augment

splice site pairing by simultaneously interacting with both U1 snRNP and U2AF65 across the intron. Alternatively, SR-proteins may recruit the U4/U5+U6 tri-snRNP complex to the splicesome. Similarly, SR-proteins may also interact with the splicing co-activator SRm160, promoting spliceosome formation without direct interaction with U2AF65 (Blencowe, 2000).

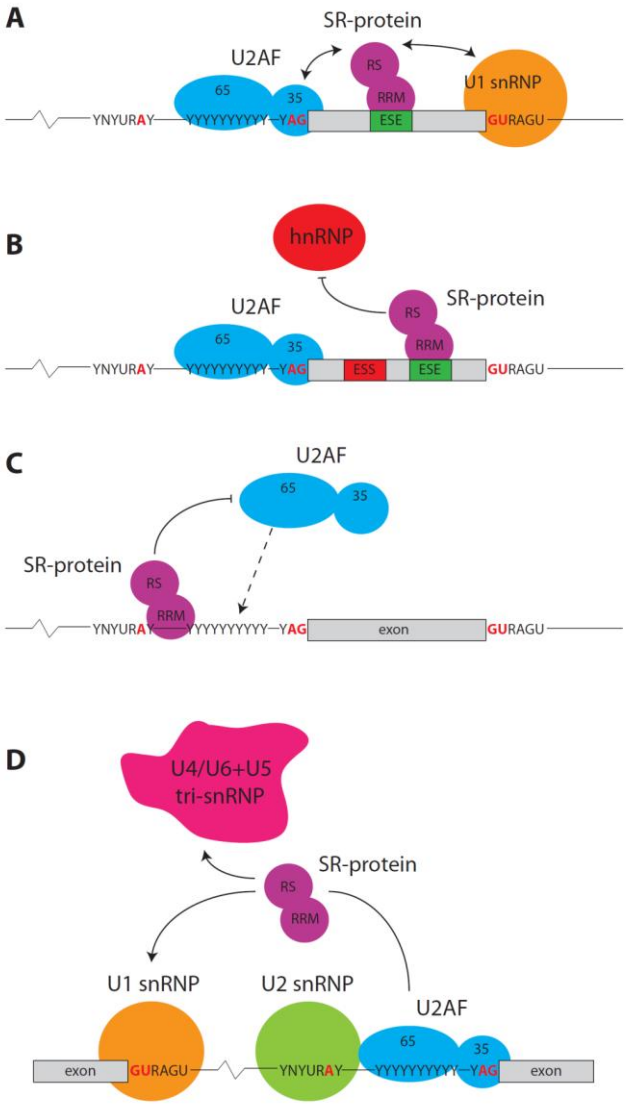


Figure 1.7 Four proposed models of splicing regulation by SR-proteins. (A) In the classical U2AF recruitment model, SR-proteins stabilise the interactions between U2AF heterodimer and the 3' splice site by binding to ESE's. SR-proteins may also stablise the interaction of U1 snRNP with the 5' splice site by binding to ESE's. (B) In the inhibitor model, SR-proteins block recruitment of splicing repressor proteins, preventing them from inhibiting spliceosome assembly. (C) SR-proteins can also function as splicing repressors by binding to intronic regions in close proximity to the splice sites of regulated exons (D) In the splicing co-activator model, SR-proteins may recruit spliceosomal components or stablise spliceosome interactions independent from direct RNA-binding. The images in Figure 1.7 are adapted from the following reviews (Matlin *et al.*, 2005; Long and Caceres, 2009; Shepard and Hertel, 2009a).

It is possible that the recruitment of spliceosomal components and silencer antagonism occur simultaneously to contribute to splicing activation by SR-proteins. In fact, each of the aforementioned models (Figure 1.7) may contribute to splicing regulation by SR-proteins in different contexts.

1.1.11 Other factors which influence splicing

Splicing decisions are made in a dynamic environment and RNA-binding proteins do not determine splicing activation in isolation. Formation of secondary structures within the pre-mRNA can affect splicing activation in two ways; either by preventing RNA-binding proteins from recognising a specific sequence, or by altering the relative distance between splice sites and the auxiliary elements (Warf and Berglund, 2010). Likewise, as splicing occurs predominantly co-transcriptionally, the rate of transcription (or rate of RNA elongation) by RNA polymerase II has also been shown to influence splicing through changes to the relative speed in which competitive splice sites and auxiliary elements are synthesised (Cáceres and Kornblihtt, 2002). Finally, independent of elongation rate, epigenetic factors such as histone modifications have been proposed to regulate alternative splicing (Luco et al., 2011). For example, the histone arginine methyltransferase CARM1 can directly interact with U1 snRNP (Ohkura *et al.*, 2005; Cheng *et al.*, 2007), suggesting that chromatin complexes may have a role in facilitating the correct assembly of pre-spliceosome components onto the pre-mRNA (Luco et al., 2011).

1.2 The SR-like splicing factors Tra2α and Tra2β

1.2.1 Tra2 proteins

The *transformer-2* gene was originally identified in the fruit-fly *Drosophila melanogaster*, where the Tra2 protein is a key component of an alternative splicing complex involved in sex-determination (Tian and Maniatis, 1993). In the *Drosophila* sex-determination pathway, female-specific expression of the Tra protein is established through alternative splicing of the *transformer* (*tra*) gene, regulated by female-specific expression of the Sex-lethal protein (Valcarcel et al., 1993). In turn, the Tra protein cooperates with the non-sex-specific Tra2 protein, to promote splicing inclusion of exon 4 of the *doublesex* (*dsx*) pre-mRNA, leading to female fruit-fly development (Lopez, 1998).

Tra2 proteins are well conserved across the animal kingdom. However, whilst invertebrate genomes contain a single *tra2* gene, vertebrate genomes contain two distinct *tra2* genes. The human *TRA2* genes are designated *TRA2A* and *TRA2B*, which encode the proteins Tra2α and Tra2β respectively. It is likely that early in vertebrate evolution, the *Tra2* gene was duplicated, resulting in vertebrates obtaining two copies of the *Tra2* gene and invertebrates retaining a single copy. In evolutionary terms, there is substantial functional conservation between the invertebrate and vertebrate Tra2 homologues. For example, the introduction of human Tra2α protein in transgenic Tra2-deficient fruit-flies is able to partially rescue Tra-dependent splicing and female sex-determination (Dauwalder et al., 1996).

1.2.2 Human Tra2α and Tra2β

The human Tra2α and Tra2β proteins share both structural and functional similarities. Tra2α and Tra2β have the same modular protein structure, consisting of a single, central RNA recognition motif (RRM) flanked by N-terminal and C-terminal RS-domains. Tra2α and Tra2β share a 63% amino acid homology, with particularly high conservation throughout their RRMs and both proteins were determined to have indistinguishable RNA sequence specificities by SELEX (Systematic Evolution of Ligands by EXponential enrichment) (Tacke *et al.*, 1998a). Both RBPs are known to be sequence-specific activators of alternative splicing (Tacke *et al.*, 1998a), however the relative contributions of endogenous Tra2α and Tra2β to this process and their functional relationship is not well understood (Best *et al.*, 2014b). For example, it is not clear whether endogenous Tra2α and Tra2β share the same functions and regulate the same target exons. Both Tra2α and Tra2β can regulate splicing of the same minigene model exons when over-expressed in HEK-293 cells (Grellscheid et al., 2011a), demonstrating some degree of functional redundancy. However, the *Tra2a* gene alone is not sufficient to maintain the phenotype of *Tra2b* knockout mice (Mende *et al.*, 2010b; Roberts *et al.*, 2014; Storbeck *et al.*, 2014). Ubiquitous deletion of *Tra2b* is embryonic lethal, resulting in highly disorganised embryos at day E7.5 and death during early embryonic development (Mende et al., 2010b). Similarly, conditional knockout of *Tra2b* in the nervous system severely disrupts brain development. A cortex-specific *Tra2b* knockout mouse model led to loss of neural progenitor cell survival caused by apoptosis (Roberts et al., 2014). Likewise, a broader neuronal-specific knockout of

Tra2b caused severe abnormalities in the cortex and thalamus, leading to death shortly after birth (Storbeck et al., 2014). Hence despite sharing similar splicing targets *in cellulo*, the *Tra2a* and *Tra2b* genes are not entirely redundant *in vivo*.

Interestingly, Tra2β was previously found to auto-regulate its own protein expression by promoting inclusion of a poison exon within the *TRA2B* pre-mRNA (Stoilov et al., 2004). However, it is not known whether Tra2β may similarly regulate expression of Tra2α, or vice versa. Despite their similarities, generally speaking there have been far more studies of Tra2β-mediated splicing regulation than of its paralogous protein Tra2α. Hence for the remainder of this section, I largely focus on studies involving Tra2β, although some observations may apply to both Tra2 proteins.

1.2.3 The *TRA2B* gene generates multiple mRNA isoforms

The human *TRA2B* gene contains ten exons, generating at least five known mRNA isoforms and two known protein isoforms via a combination of alternative splicing, alternative polyadenylation and the use of alternative promoters (Stoilov et al., 2004). *TRA2B* mRNA isoform 1 (termed *TRA2B*-1) is ubiquitously expressed (Daoud et al., 1999) and encodes the full-length Tra2β protein (Tra2β-1) which contains both RS-domains together with a central RRM (see Figure 1.8 A). Interestingly, the Tra2β-1 protein isoform has been shown to auto-regulate its own protein concentration through a negative feedback loop (Stoilov et al., 2004). Tra2β-1 was found to directly bind to a posion exon (exon 2) within the *TRA2B* pre-mRNA, promoting inclusion of this frequently skipped exon. Inclusion of exon 2 produces of a non-productive mRNA species (termed *TRA2B*-2), which is not translated into protein (see Figure 1.8 B). Stoilov et al. suggest that this molecular sensor may contribute to the maintenance of normal cell physiology, given deregulation of Tra2β has been observed in numerous disease states (2004), such as nerve injury (Kiryu-Seo et al., 1998), hypoxia (Matsuo et al., 1995), and silicosis (Segade et al., 1995).

A third *TRA2B* mRNA isoform (termed *TRA2B*-3) encodes a shorter protein isoform which lacks the functional RS1 domain (see Figure 1.8 C) and is expressed in a tissue-specific manner (predominantly in the brain) (Nayler et al., 1998a). Whether full-length Tra2β-1 and the shorter Tra2β-3 protein isoform share similar functions in splicing regulation or are functionally distinct is currently unclear. It was previously reported

that unlike full-length Tra2β-1, the shorter Tra2β-3 protein isoform did not influence splice site selection (Stoilov et al., 2004). However, subsequent data has indicated that Tra2β-3 can activate inclusion of *SMN2* exon 7 (Cléry et al., 2011). Consistent with an important function however, expression of the *TRA2B*-3 mRNA isoform is conserved in invertebrates (Nayler et al., 1998a). Interestingly, it was previously demonstrated that the dopamine and cAMP regulated phosphoprotein 32kDa (DARPP-32) directly interacts with Tra2β-1, and over-expression of DARPP-32 was associated with a reduction in splicing inclusion of Tra2β-regulated exons (Benderska et al., 2010). As DARPP-32 also directly interacts with the Tra2β-3 protein isoform, Benderska et al. have postulated that one possible function of the Tra2β-3 protein isoform may be to sequester DARPP-32/Tra2β interactions, without influencing splice site selection directly (Benderska *et al.*, 2010). A fourth *TRA2B* mRNA isoform (*TRA2B*-4) is generated when exons 2 and 3 are skipped, although this isoform is not translated. A fifth *TRA2B* mRNA isoform (*TRA2B*-5) is also generated through use of an alternative promoter within the second intron (Stoilov et al., 2004).

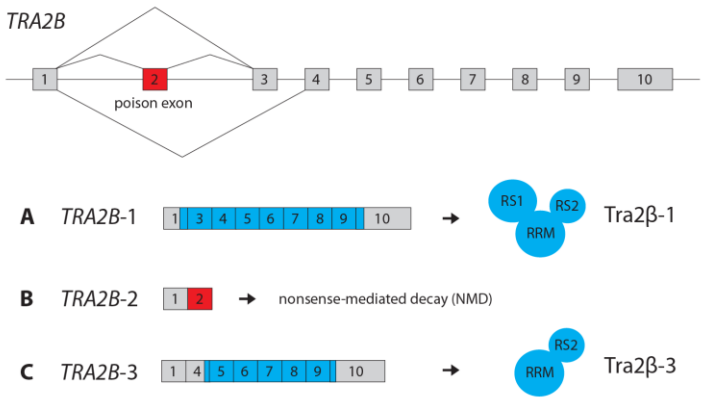


Figure 1.8 The human *TRA2B* gene generates two known protein isoforms via alternative splicing. The open reading frames (ORFs) are highlighted in blue. The *TRA2B* poison exon (exon 2) is highlighted in red. (A) Skipping of exon 2 generates the *TRA2B*-1 mRNA isoform which encodes full-length Tra2β. (B) Inclusion of exon 2 generates a downstream frameshift, resulting in a non-producive transcript. (C) Skipping of exons 2 and 3 generates the *TRA2B*-3 mRNA isoform, which encodes a shorter Tra2β-3 protein isofrom which lacks the RS1 domain. This image is adapted from (Stoilov et al., 2004).

1.2.4 Tra2β protein structure and sequence specificity

Classical SR-proteins contain one or two N-terminal RNA recognition motifs (RRMs) and a single C-terminal RS-domain (Wu and Maniatis, 1993; Kohtz *et al.*, 1994).

However unlike the classical SR-proteins, the full-length Tra2α and Tra2β proteins share an unusual modular structure, which consists of a single, central RRM flanked by N-terminal and C-terminal RS-domains (see Figure 1.9). The larger, N-terminal RS-domain (termed RS1) of the Tra2β protein contains a region of 19 arginine/serine dipeptides, whilst the smaller C-terminal RS-domain (termed RS2) is located within close proximity of a hinge region and contains just 7 arginine/serine dipeptides (Beil, 1997). Similarly to classical SR-proteins, the RS-domains of Tra2β facilitate a number of direct protein-protein interactions. Tra2β has been shown to directly interact with a number of other RNA-binding proteins, including SRSF1 (Ge et al., 1991), hnRNPG and hnRNP-GT (Soulard *et al.*, 1993; Venables *et al.*, 2000), RMBY (Venables et al., 2000), SRp30c and SAF-B (Nayler et al., 1998b). In addition to facilitating direct protein-protein interactions, it was recently identified that both the RS1 and RS2 domains contain nuclear localisation signals (NLS) important for the nuclear import of Tra2β. Using mutational analysis, Li et al. found that the presence of either the RS1-domain or the RS2-domain was sufficient for nuclear import of Tra2β (Li *et al.*, 2013). What's more, two sub-nuclear localisation signals were also identified within the RS1-domain (but not the RS2-domain) which are required for the sub-nuclear localisation of Tra2β to nuclear speckles (Li et al., 2013).

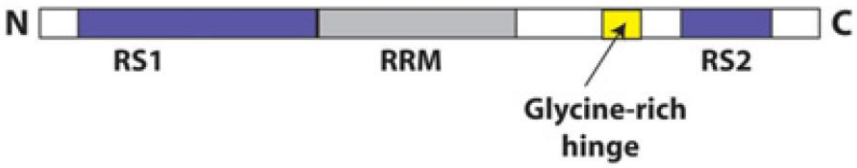


Figure 1.9 Modular structure of the full-length Tra2β protein. The full-length Tra2β protein contains a single, central RNA-recognition motif (RRM) flanked by two arginine/serine-rich (RS)-domains. This image is taken from a recent review by (Best et al., 2014a).

Both Tra2 proteins contain highly-conserved RNA-recognition motifs (RRMs) required for direct RNA binding. The first description of human Tra2 protein sequence specificity was reported by Tacke et al., who identified preferential binding of human Tra2 proteins to 'GAA' repeats using a series of SELEX experiments (1998a). Subsequently, Tra2β was also found to directly bind to another RNA sequence, specifically to C/A-rich elements (ACE) such as 'ACUUCAACAAGUU'; a thirteen-nucleotide sequence located within exon 4 of the *CGRP* gene (Tran et al., 2003). This suggested that Tra2β may

recognise multiple, distinct RNA sequences. Tra2β protein/RNA interactions have now been resolved at atomic resolution using nuclear magnetic resonance (NMR) (Cléry *et al.*, 2011; Kengo Tsuda, 2011). It was discovered that Tra2β recognises two distinct sequences; a longer, degenerate sequence containing 'CAA' and the shorter tetra-nucleotide sequence 'AGAA' (Kengo Tsuda, 2011). The tetra-nucleotide sequence 'AGAA' is recognised by specific residues within the β-sheets of the Tra2β RRM (Cléry et al., 2011). The longer degenerate sequence containing 'CAA' is only recognised when integrated into a stem-loop structure (RNA secondary structure) (Kengo Tsuda, 2011). Hence Tra2β has two distinct modes of RNA sequence recognition (Kengo Tsuda, 2011).

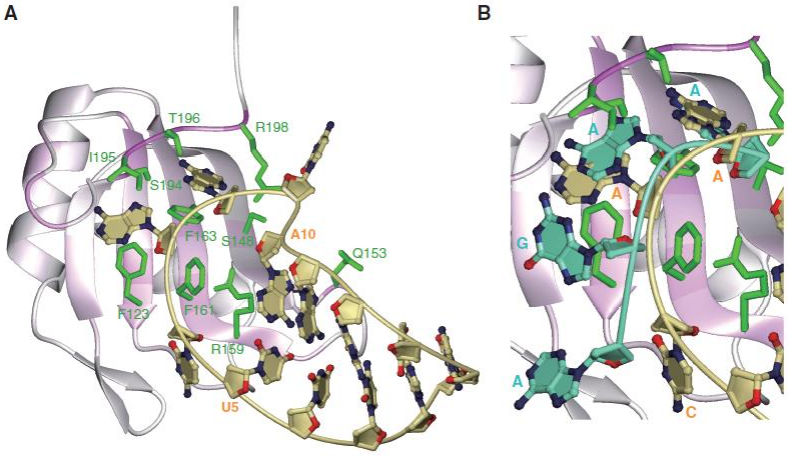


Figure 1.10 Tra2β has two distinct modes of RNA sequence recogntion. (A) The Tra2β RRM in complex with the sequence 'GACUU**CAACAA**GUC' when integrated into a stem loop structure. (B) The Tra2β RRM also specifically recognises the tetra-nucleotide sequence '**AGAA**'. This image is taken from (Kengo Tsuda, 2011).

1.2.5 Phosphorylation of Tra2β

Similarly to classical SR-proteins, phosphorylation of Tra2β can influence its activity in splicing regulation. Interestingly, phosphorylation of serine residues within different functional domains of the protein may alter splicing activity through distinct mechanisms. For example, phosphorylation of serine residues within the RS-domains of SR-proteins can regulate sub-cellular localisation (Lin and Fu, 2007), whilst phosphorylation of the RRM domain of SR-proteins has been shown to directly influence the affinity of protein-RNA interactions (Benderska et al., 2010).

Tra2β is actively dephosphorylated by protein phosphatase 1 (PP1) on residues located within its RRM and PP1 activity has been shown to influence splicing inclusion of Tra2β-regulated exons, including *SMN2* exon 7 (Novoyatleva et al., 2008). The dopamine and cAMP regulated phosphoprotein 32kDa (DARPP-32) is an inhibitor of PP1, which directly interacts with Tra2β to prevent PP1-mediated dephosphorylation. Over-expression of DARPP-32 was found to reduce splicing inclusion of the Tra2β targets using minigene assays, whilst depletion of DARPP-32 via RNAi increased splicing activation of the same target exons (Benderska et al., 2010). These experiments suggest that variable phosphorylation of the Tra2β RRM may significantly influence target exon inclusion by affecting the RNA-binding capacity of Tra2β (Stoilov *et al.*, 2004; Benderska *et al.*, 2010).

1.2.6 Well characterised RNA targets of vertebrate Tra2 proteins

Well characterised targets of vertebrate Tra2 proteins include *SMN2* (*survival motor neuron 2*) exon 7. Regulation of *SMN2* exon 7 is of significant clinical interest in the motor neuron disease Spinal Muscular Atrophy (SMA), which is caused by loss-of-function mutations to the *survival motor neuron (SMN1)* gene (Hofmann and Wirth, 2002). The highly-conserved paralogous gene *SMN2* fails to provide a functional protein replacement due to a translationally silent mutation within a splicing enhancer sequence. As a result, *SMN2* exon 7 is skipped, producing a truncated SMN2 protein which is unable to functionally compensate for the loss of *SMN1*. *SMN2* exon 7 is a potential therapeutic target for SMA, as restoring efficient inclusion of *SMN2* exon 7 could effectively restore full-length SMN expression. Tra2β directly enhances splicing inclusion of *SMN2* exon 7 through a GA-rich exonic splicing enhancer sequence (Hofmann et al., 2000).

Tra2β is also known to directly regulate a testis-specific exon from the *homeodomain interacting protein kinase 3* (*HIPK3*) gene, which is referred to as the *HIPK3-T* exon (Venables *et al.*, 2005; Grellscheid *et al.*, 2011b). The *HIPK3-T* exon is exclusively spliced in the testis and inclusion of this exon introduces a premature termination codon (PTC) within the *HIPK3* mRNA, targeting the *HIPK3* mRNA for degradation via the nonsense-mediated decay (NMD) pathway (Venables et al., 2005).

Other well established target exons include exon 10 from the *Tau* gene. Mutations within *Tau* exon 10 are associated with the neurological disease frontotemporal dementia with Parkinsonism linked to chromosome 17 (FTDP-17) (Fu et al., 2012). Tra2β also regulates exon 3 from the *cysteine rich 61* (*Cyr61*) gene, which encodes a matricellular protein linked with tumour progression and metastasis (Hirschfeld et al., 2011). Interestingly, Hirschfeld et al. reported that under acidic conditions, splicing inclusion of *Cyr61* exon 3 was significantly reduced and this was associated with a substantial switch in localisation of Tra2β from the nucleus to the cytoplasm (Hirschfeld *et al.*, 2011). Other examples of known target exons include exon 23 from the smooth muscle myosin phosphatase targeting subunit (*Mypt1*) gene (Fu et al., 2012), variable exons v4 and v5 from *CD44* (Watermann et al., 2006) and most recently exon 11 from *BRCA1* (Raponi et al., 2014).

The aforementioned studies largely focus on detailed characterisation of single exons. In addition, a number of recent studies have now investigated endogenous Tra2β target exons on a more global scale, using RIP-seq (Uren et al., 2012) and splicing-specific microarrays (Anderson *et al.*, 2012; Storbeck *et al.*, 2014), as well as studies published from the Elliott lab using HITS-CLIP (Grellscheid et al., 2011a), iCLIP and RNA-seq (Best et al., 2014b). Anderson et al. identified that the cardiotonic steroid digitoxin induces substantial changes in alternative splicing through depletion of the RNA-binding proteins SRp20 and Tra2β (Anderson *et al.*, 2012). Subsequently, they used shRNA-mediated depletion of Tra2β in HEK-293 cells to identify Tra2β-responsive exons using microarray analyses. Interestingly, depletion of Tra2β was found to induce skipping of some alternative exons whilst enhancing inclusion of others (Anderson et al., 2012). Similarly, Storbeck et al. analysed RNA from a neuronal-specific *tra2b* knockout mice using exon arrays and identified *Tubd1* exon 4 and *Sgol* exon 2 as *in vivo* targets of Tra2β (Storbeck *et al.*, 2014).

1.2.7 The physiological functions of Tra2β

Tra2β is associated with numerous physiological functions. These include a potential role in mammalian spermatogenesis, as Tra2β protein expression is up-regulated during meiosis in male germ cells (Grellscheid et al., 2011a) and Tra2β directly interacts with the germ-cell specific protein RBMY (Venables et al., 2000). Tra2β is also involved in generating tissue-specific splicing patterns of *Mypt1* in smooth muscle cells

(Fu et al., 2012), and the cardiotonic steroid digitoxin (prescribed for the treatment of heart failure) affects alternative splicing patterns through depletion of Tra2β (Anderson et al., 2012). Tra2β protein expression is also thought to be important to the maintenance of normal cell physiology, given the number of pathophysiological states now associated with deregulation of Tra2β expression (Best *et al.*, 2013). *TRA2B* mRNA expression changes throughout the ageing process (Holly et al., 2013) and age-related macular degeneration is associated with an increase in Tra2β expression and aberrant nuclear localisation of Tra2β in diseased retinal cells (Karunakaran et al., 2013). Changes in Tra2β protein expression is also associated with the disease pathology of Alzheimer's disease (Glatz et al., 2006), Frontotemporal Dementia and Parkinsonism linked to chromosome 17 (FTDP-17) (Jiang et al., 2003) and several human cancers (reviewed by Best et al. 2013).

1.3 Alternative splicing in cancer

1.3.1 Introduction

The importance of alternative splicing regulation is demonstrated by a growing number of diseases now associated with altered splicing activity, including aberrant splicing events in cancer (Srebrow and Kornblihtt, 2006). Alteration to splicing activity can result in potentially pathogenic consequences through multiple mechanisms. These include mutation of *cis*-elements, which alter the RNA sequence and can lead to aberrant splicing. For example, the introduction of a novel splice site through mutation and its subsequent recognition may produce disease-associated mRNA isoforms which are not typically produced under normal physiological conditions. Consequently, aberrant splicing may result in the production of novel protein isoforms that are structurally and functionally distinct from those typically produced, with the potential to exert profound effects on cellular physiology and phenotype (Pajares et al., 2007). Secondly, changes to *trans*-acting factors which regulate splicing may result in a deviation from the typical ratio of pre-existing alternatively spliced isoforms, or the recognition of cryptic splice sites (Pajares et al., 2007). Changes to the ratio of normal physiological isoforms may consequently drive cellular phenotype in a particular direction. Hence deregulation of splicing in cancer can result from either mutation to *cis*-elements, including splice sites and auxiliary elements, or alternatively through the

modification of *trans*-acting proteins, including changes to the concentration, activity or localisation of splicing factors (Peter Stoilov, 2002; Faustino and Cooper, 2003). Alternative splicing has now been demonstrated to produce different protein isoforms which promote many of the hallmarks of cancer. However, distinguishing between splice variants which genuinely contribute to cell transformation and those which are simply the by-product of lost splicing regulation in the transformed cell remains a considerable challenge.

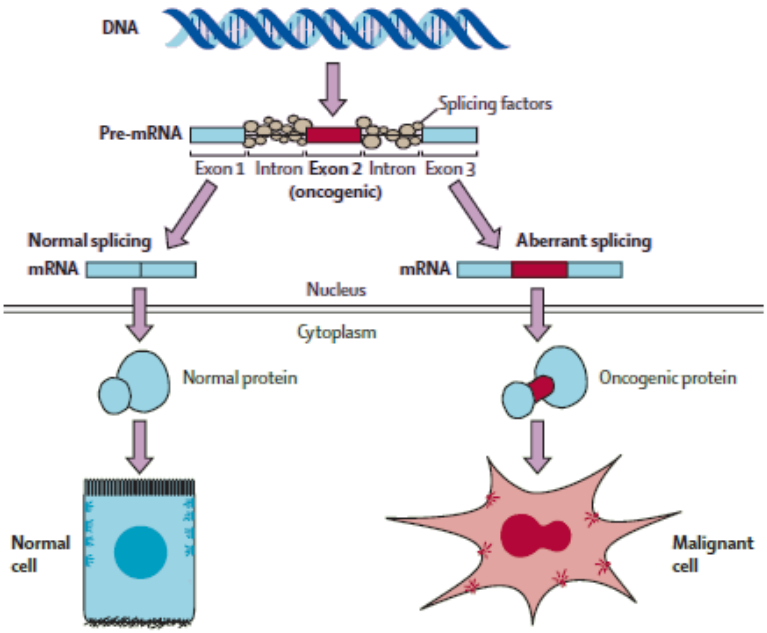


Figure 1.11 Abnormal pre-mRNA splicing can contribute to tumour progression through the production of protein isoforms with oncogenic properties. Mutation of *cis*-elements may result in the production of novel protein isoforms with potentially oncogenic function. Alternatively, changes to *trans*-acting factors may change the ratio of pre-existing mRNA isoforms, or promote recognition of cryptic splice sites, contributing to changes in cell phenotype. This image is adapted from (Pajares et al., 2007).

1.3.2 Mutations within *cis*-elements

Cells may require as few as 2-10 driving mutations to initiate oncogenic transformation (Hahn and Weinberg, 2002; Stratton *et al.*, 2009), a modest number considering many advanced tumours are estimated to contain up to 100,000 genetic alterations (Martin Dutertre, 2010). The cellular mechanisms which maintain genomic stability are frequently lost in tumour cells, including the loss or impairment of DNA repair pathways. Consequently, the accumulation of mutations allow tumour cells to acquire

many of the hallmarks of cancer, which include an increased proliferative capacity, resistance to apoptosis, unlimited replication potential, self-sufficiency in growth signals, insensitivity to growth inhibitory signals, angiogenesis and metastasis (Hanahan and Weinberg, 2011). Similarly, loss of normal splicing regulation has recently emerged as another possible mechanism which could contribute to cellular transformation.

Originally, analyses of cancer-associated mutations within genomic DNA typically focused on their downstream effect on primary protein structure (the amino acid sequence) (Pajares et al., 2007). For instance, point mutations may be categorised as nonsense (introduction of a premature stop codon), missense (change in the amino acid) or synonymous (does not change the amino acid sequence). However, whilst synonymous mutations may appear to be translationally silent (they do not appear to alter the amino acid sequence), such mutations can exert dramatic effects on splicing regulation (Blencowe, 2006). Synonymous mutations can introduce novel splice sites or regulatory elements, or disrupt pre-existing ones, potentially altering the mRNA isoform and ultimately the protein that is produced.

Mutations which cause changes to splice site selection have been reported in numerous human cancers (Liu *et al.*, 2001; Lukas *et al.*, 2001; Oltean *et al.*, 2006). Key examples include the inherited nonsense mutation within exon 18 from the *BRCA1* gene, which encodes an important protein involved in repair of DNA double-strand breaks (DSBs) (Friedenson, 2007). This inherited mutation disrupts the binding site of the splicing factor SRSF1 (ASF/SF2), resulting in the aberrant skipping of exon 18 from the *BRCA1* mRNA (Liu et al., 2001). Similar mutations which affect splice site selection have been reported in *MDM2*, which encodes a nuclear phosphoprotein and key regulator of the tumour suppressor P53 (Lukas et al., 2001). Mutations which disrupt pre-mRNA splicing of the DNA mismatch repair genes *MSH2* and *MLH1* are also common in hereditary non-polyposis colorectal cancer (HNPCC) (Stella et al., 2001).

1.3.3 *Trans*-acting factors in cancer

Changes to *trans*-acting splice factors can also exert significant effects on pre-mRNA splicing in tumour cells. The alteration in expression, stability, localisation and/or activity of splicing factors can have a significant effect on splicing decisions in cancer

cells (Pajares et al., 2007). There is some evidence that the general splicing machinery of a cell may be fundamentally changed in cancer (Mee Young Kim, 2009; Pedro A. F. Galante, 2009) and it has been hypothesized that there may even be a "splicing switch" in cancer cells, allowing cells to generate a vast array of splice variants, some of which may contribute to oncogenic transformation (Skotheim and Nees, 2007). The expression of several splicing factors has been found to be significantly altered in breast cancer tumours compared to corresponding normal tissue, including FOX2 (Venables et al., 2009b), SRSF1 (Karni et al., 2007) and Tra2β (Watermann et al., 2006). Interestingly, the splicing factor oncoprotein SRSF1 (ASF/SF2) has been found to promote malignant transformation and is over-expressed in multiple human tumours (Chang et al., 2007). It was subsequently identified that SRSF1 regulates alternative splicing of the *BIM* and *BIN1* genes, which encode proteins involved in the regulation of apoptosis (Anczuków *et al.*, 2012). Over-expression of SRSF1 promotes expression of BIM and BIN1 protein isoforms which lack pro-apoptotic functions, possibly contributing to epithelial cell transformation (Anczuków et al., 2012).

1.3.4 Tra2β expression in cancer

TRA2B gene expression is amplified in several human cancers, including cancers of the lung, cervix, head and neck, ovary, stomach, and uterus (reviewed by Best et al., 2013). Tra2β protein expression was also found to be up-regulated in a subset of breast (Watermann et al., 2006), cervical (Gabriel et al., 2009), ovarian (Fischer et al., 2004) and colon (Kajita et al., 2013) cancers. High Tra2β protein expression is associated with a poorer prognosis for patients with cervical cancer (Gabriel et al., 2009). Tra2β protein expression has also been shown to be important for cell proliferation in several cancer cell lines. Knockdown of Tra2β in the gastric cancer cell line AGS was found to suppress cell growth (Takeo et al., 2009), whilst knockdown of Tra2β in the colon cancer cell line HCT116 reduced cell viability and led to increased apoptosis (Kajita et al., 2013). Watermann et al. found that both *TRA2B* mRNA and Tra2β protein expression was significantly up-regulated in a panel of invasive breast cancer tumour samples when compared to normal breast tissue from the same patient (Watermann *et al.*, 2006). Conversely, their analysis of SRp40 expression revealed no change between normal and tumour samples, suggesting a specific up-regulation of Tra2β, rather than a broader change to the splicing machinery. Hence it has been speculated that Tra2β

may have an important role in breast cancer, through regulation of alternatively spliced isoforms associated with tumour progression and metastasis (Watermann et al., 2006).

1.3.5 Breast cancer and the MDA-MB-231 cell line

Breast cancer is a complex and heterogeneous disease, consisting of multiple disease sub-types which have distinct clinical implications (Holliday and Speirs, 2011). Most breast cancers originate from epithelial cells and initially develop following an increase in the epithelial cell mass (hyperplasia), followed by the expansion of abnormal cells (atypical hyperplasia), carcinoma *in situ* (non-invasive breast cancer) and ultimately infiltrating carcinoma (invasive breast cancer). Traditionally, breast cancer classification was based on the histological type, tumour grade and lymph node status, as well as a limited number of predictive markers such as expression of the oestrogen receptor (ER) (Holliday and Speirs, 2011). Advances in molecular profiling such as the use of DNA microarrays and immunohistochemical analysis of hormone receptors including the oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER-2) subsequently led to the recognition that breast cancer could be classified into at least five distinct subtypes based on expression profiles and hormone receptor status (Perou *et al.*, 1999; Perou *et al.*, 2000). Most recently, genomic and transcriptomic profiling has revealed that breast cancer consists of at least ten different subtypes with distinct “molecular signatures”, each with distinct clinical outcomes (Curtis et al., 2012).

Breast cancer is often modelled in the laboratory using established breast cancer cell lines. The very first breast cancer cell line (BT20) was established in 1958 and originated from an invasive ductal carcinoma (Lasfargues, 1958). Since then, a number of different breast cancer cell lines have become widely used including cell lines from the MD Anderson series (such as the MDA-MB-231 cell line used extensively in my study) (Cailleau R, 1978a), as well as cell lines from the Michigan Cancer Foundation such as MCF-7 cells (Soule HD, 1973). The MDA-MB-231 cell line is a tumourigenic breast epithelial cell line, first derived from a pleural effusion of a 51 year old female patient with breast adenocarcinoma (Cailleau R, 1978b). The MDA-MB-231 cell line models invasive disease *in vitro* and is traditionally considered to be “triple-negative” for hormone receptor status (ER-, PR-, HER2-).

The use of breast cancer cell lines as experimental models comes with significant limitations. Cell lines are prone to genetic and phenotypic drift over extended periods of continual culture (Burdall *et al.*, 2003). Changes in cell phenotype over time not only have implications for experimental reproducibility, but as cells drift further away from the phenotype of the original tumour, their relevance as accurate models of disease is diminished (Burdall et al., 2003). Other limitations include the physiological relevance of growing cells on plastic in two dimensions, as the network of interactions that exist between cells *in vivo* may be lost. This is exemplified by MDA-MB-231 cells, which are widely regarded as invasive *in vitro*, yet show poor metastatic potential in some *in vivo* xenograph models (Burdall et al., 2003). Despite the aforementioned limitations, cell lines remain a valuable experimental model, partly due to their unlimited replicative capacity and their relatively high degree of homogeneity.

1.3.6 Alternatively spliced isoforms as potential biomarkers and therapeutic targets in cancer

Tumour-specific alternative splice variants have been postulated as potential biomarkers for disease prognosis, or as novel therapeutic targets in cancer. Key examples include alternatively spliced isoforms of the androgen receptor (AR), which are strongly linked with disease progression in prostate cancer (PCa). In the early stages of PCa, prostate cancer cells depend on androgen hormones to drive proliferation via the androgen receptor (Feldman and Feldman, 2001). Current clinical treatments for PCa include androgen ablation therapy, however after an initial period of remission, many patients subsequently develop hormone-resistant or castrate-resistant prostate cancer (CRPCa) (Kohli and Tindall, 2010). Unlike the full-length AR which is localised in the cytoplasm and translocates to the nucleus upon androgen binding, some alternatively spliced isoforms of the AR lack the C-terminal ligand-binding domain and are permanently localised within the nucleus (Dehm and Tindall, 2011). Consequently, some alternatively spliced isoforms of the AR are constitutively active, even in the absence of androgens, suggesting alternatively spliced isoforms of the AR may play a role in the development of castrate-resistant prostate cancer.

Alternative splicing of the *human epidermal growth factor receptor 2* (*HER-2*) gene may also have significant clinical importance in breast cancer disease progression and drug resistance (Jackson et al., 2013). Expression of the HER-2 protein is up-regulated

in approximately 20-30% of breast cancers (Rubin and Yarden, 2001) and confers increased proliferative and survival advantages via oncogenic signalling pathways involving the PI3K/AKT pathway (increased survival/reduced apoptosis) and the RAS/RAF/MEK/MAPK pathway (increased proliferation) (Citri and Yarden, 2006). Consequently, HER-2 was identified as an ideal candidate for the development of novel targeted therapies, which include the monoclonal antibody trastuzumab (Herceptin) and the tyrosine kinase inhibitor lapatinib (Tyverb). HER-2 splice variants may be of significant prognostic value given a number of different HER-2 protein isoforms are functionally distinct. For example, the Δ16HER2 isoform (in which exon 16 is skipped) is associated with increased malignant transformation and trastuzumab resistance (Mitra et al., 2009). Consequently, patients expressing the Δ16HER2 variant may benefit from more aggressive treatment strategies. The truncated HER-2 p100 protein isoform is created by retention of intron 15 (Scott et al., 1993). HER-2 p100 is associated with reduced tumour cell proliferation, possibly due to reduced downstream signal transduction from the truncated HER-2 protein (Wimberly et al., 2014). A third HER-2 variant termed 'Herstatin' is produced following retention of intron 8, which also inhibits growth of HER-2 over-expressing cells (Doherty et al., 1999). Consequently, considering expression of *HER-2* splice variants in the future may help determine patient prognosis or be used to predict the effect of therapeutic agents which target HER-2.

Several strategies have been identified which may be useful when targeting splicing in the treatment of human disease. These include the use of synthetic antisense oligonucleotides (ASOs) which bind pre-mRNA and block splice site selection, the use of RNAi to target specific isoforms for degradation, or the use of monoclonal antibodies to specifically target aberrant protein isoforms and inhibit their function. However, currently the number of potential treatments which target splicing that have reached the clinical trial stage is limited. Most therapeutic strategies have been aimed at the treatment of monogenic diseases, such as Spinal Muscular Atrophy (SMA) (Kolb and Kissel, 2011). Interestingly, splice-switching oligonucleotides (SSO) have been successfully used to switch HER-2 mRNA splicing in cell lines, which lead to reduced proliferation and the induction of apoptosis (Wan et al., 2009). Targeting splice variants in the treatment of cancer may become more viable in the future if improved

delivery methods are developed and novel splice variants are identified which significantly contribute to disease initiation and progression (Pajares et al., 2007).

1.4 Methodologies for dissecting alternative splicing

1.4.1 Introduction

Given the complexity of splicing regulation, how can we further our current understanding of alternative splicing? To date, there have been many techniques used to dissect the molecular mechanisms of splicing, utilizing both *in vitro* and *in vivo* splicing assays. Traditionally, a reductionist approach has been used to create individual model systems which allow the roles of specific *cis*- or *trans*-acting factors to be characterised *in vivo*, such as the use of minigenes. This has been complemented by *in vitro* techniques, such as SELEX, which is used to identify consensus sequences recognised by RNA-binding proteins. Most recently however, high-throughput techniques have been developed which have facilitated the study of splicing on a much greater scale. These include the development of splicing-specific microarrays, high-throughput RT-PCR platforms, UV cross-linking and immunoprecipitation (CLIP) experiments coupled with next-generation sequencing, as well as RNA-seq.

1.4.2 Model systems using minigenes

The creation of model systems using minigenes is frequently used to replicate splicing regulation *in vivo* (Mardon et al., 1987) and has been used in numerous studies to investigate splicing regulation by Tra2β (Glatz *et al.*, 2006; Grellscheid *et al.*, 2011a). The construction of minigenes involves cloning of the genomic region of interest (often one or more alternatively spliced exons and the flanking intronic regions) into an exon-trap vector. The insert is cloned in between two constitutively spliced exons, which are downstream of a eukaryotic promoter that drives transcription. Subsequently, minigene constructs are then transfected into cells where they are subject to splicing regulation similar to the endogenous gene. Minigene experiments are particularly suited to the study of *cis*- and *trans*-acting factors which regulate a specific region within a gene. *Cis*-elements are often studied through mutagenesis of splice sites and auxiliary elements, or through the introduction of new regulatory sequences. Similarly,

trans-acting factors may be studied through the introduction of proteins using expression vectors or alternatively by depletion of endogenous proteins using RNAi.

1.4.3 Alternative splicing microarrays

The development of splice-specific microarrays allowed thousands of alternative splicing events to be analysed in parallel. Microarrays contain thousands of immobilised oligonuclotide probes which hybridise with specific RNA targets (Matlin et al., 2005). Traditionally, microarrays have been used to monitor the expression of mRNA without the capacity to distinguish between multiple splice variants (Ben-Dov et al., 2008). However, the development of isoform-specific microarrays permited global analyses of alternative splice events (Johnson et al., 2003).

1.4.4 SELEX

Systematic Evolution of Ligands by EXponential enrichment (SELEX) is an *in vitro* screening technique frequently used to determine the RNA sequences that are recognised by RNA-binding proteins. Briefly, it involves the progressive selection of specific RNA molecules from pools of random RNAs to identify a consensus binding sequence (Yang et al., 2007). SELEX was successfully used to identify the 'GAA' repeat sequence that is one of two RNA sequences specifically recognised by Tra2β (Tacke *et al.*, 1998a). The RNA-binding sites of SR-proteins identified through SELEX have also proven particularly useful in the development of search programmes such as ESEfinder, which predicts the presence of auxiliary elements within a given sequence (Cartegni et al., 2003). Subsequently, functional SELEX was developed to identify the sequences which actually influence splicing activity when bound to a specific protein, rather than simply associate with that protein (Gopinath, 2007).

1.4.5 UV cross-linking and immunoprecipitation (CLIP)

The RNA-binding motifs generated by SELEX are typically short (between 4-10 nucleotides) and highly degenerate. Given the high frequency with which these degenerate motifs occur in the genome, accurate prediction of functional targets based on the RNA-binding motifs alone has traditionally proven difficult (Elliott, 2010). Direct RNA-protein interactions can be identified through protein/RNA immunoprecipitation experiments, such as RNA immunoprecipitation (RIP) and UV

cross-linking and immunoprecipitation (CLIP). Subsequently, CLIP coupled with high-throughput sequencing (HITS-CLIP) was developed to identify the physiological targets of RNA-binding proteins on a genome-wide scale (Ule et al., 2005a).

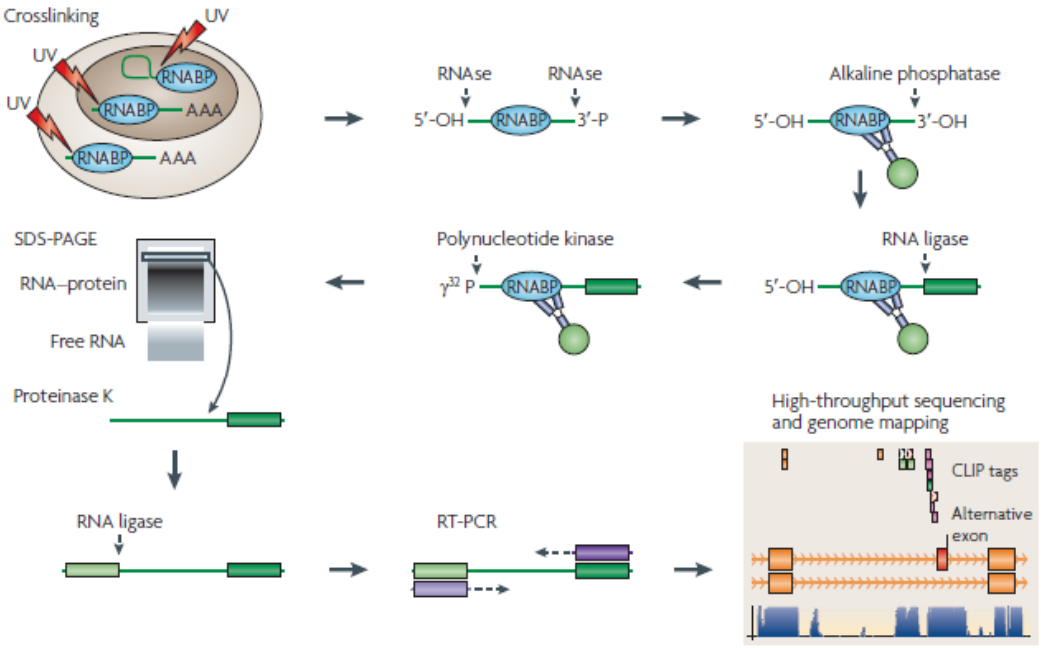


Figure 1.12 UV cross-linking and immunoprecipitation coupled with high-throughput sequencing (HITS-CLIP). HITS-CLIP utilises UV-irradiation to penetrate living cells and induce covalent cross-links between RNA and proteins in direct contact. The RNA–protein complexes are then immunoprecipitated and purified under stringent conditions. Following purification, the RNA-binding protein is digested, leaving the purified RNA substrate. This is followed by linker ligation and RT-PCR to amplify the original RNA sequences. Sequencing may then be performed using high-throughput sequencing technologies. This image is adapted from a recent review by (Licatalosi and Darnell, 2010).

CLIP was originally used to identify RNA targets of the brain-specific splicing regulator Nova in mice (Ule et al., 2003). Following this, Ule et al. combined CLIP with splicing-specific microarrays to create an 'RNA map', which could predict position-dependent splicing patterns of Nova-regulated targets (Ule et al., 2006). Subsequent studies using CLIP have identified the RNA targets for numerous RNA-binding proteins, including FOX2 (Yeo et al., 2009), SRSF1 (SF2/ASF) (Sanford et al., 2008), hnRNPC (Konig et al., 2010) and SRSF3 and SRSF4 (Anko et al., 2012). Most recently, adaptations to the original CLIP protocol have facilitated the identification of protein-RNA interactions at near individual nucleotide resolution, including individual-nucleotide resolution CLIP

(iCLIP) (Konig et al., 2011) and PhotoActivatable-Ribonucleoside-enhanced CLIP (PAR-CLIP) (Hafner et al., 2010).

1.4.6 RNA-seq

Most recently, the development of RNA-sequencing (RNA-seq) has facilitated the analysis of alternative splicing on a transcriptome-wide scale. RNA-seq was developed to allow whole transcriptome profiling using deep-sequencing and allows simultaneous analysis of gene expression and alternatively spliced isoforms. Briefly, a population of RNA (either total or fractionated, such as poly(A)+) is fragmented and reverse transcribed into a cDNA library. Sequencing adaptors are added to the cDNA fragments, which are then sequenced using next-generation sequencing technology. The sequencing reads are then processed and mapped to a reference genome to create an expression profile for each gene (Wang et al., 2009). RNA-seq has recently been applied to analyse the transcriptome of different breast cancer cell subtypes, revealing novel splicing alterations which were sub-type specific (Eswaran et al., 2013). Other recent applications include combining CLIP datasets with RNA-seq, to investigate how RNA-binding proteins such as hnRNPC and U2AF65 compete to regulate the transcriptome (Zarnack et al., 2013).

1.5 Research aims and objectives

The identification and characterisation of novel RNA targets of the RNA-binding proteins Tra2α and Tra2β was a key objective in this thesis.

Prior to the start of my project, a previous PhD student had recently used HITS-CLIP to identify direct RNA targets of Tra2β from the mouse testis (Liu, 2009). Therefore an early objective was to validate and characterise splicing regulation of exons identified from the previous HITS-CLIP experiment, using minigenes. I also investigated whether two different isoforms of Tra2β may be functionally distinct in splicing regulation using over-expressed proteins and whether direct RNA-binding was required for Tra2β-mediated splicing regulation of the newly identified exons.

Tra2β is up-regulated in several human cancers and we hypothesised that Tra2β may regulate alternative splicing of genes with functional importantance in breast cancer. Therefore for the majority of this project, my aim was to identify and characterise RNA targets in the human invasive breast cancer cell line MDA-MB-231. Two transcriptome-wide approaches were used in this study. Initially, I used iCLIP to map the transcriptome-wide binding sites of Tra2β in MDA-MB-231 cells. Subsequently, I used RNA-seq to investigate functional changes in the transcriptome following Tra2 protein depletion. By combining the iCLIP and RNA-seq data, the objective was to identify target exons which were directly bound by Tra2β and functionally responsive to Tra2 protein depletion.

The iCLIP and RNA-seq data facilitated further investigation into a number of unanswered questions regarding vertebrate Tra2 protein biology and splicing regulation. For example, do Tra2α and Tra2β regulate the same endogenous target exons? Do Tra2 proteins regulate constitutive exons in addition to their known role in alternative splicing? Do Tra2 proteins regulate splicing of genes in functionally coherent pathways? What are the phenotypic consequences of Tra2 protein depletion and deregulated splicing in MDA-MB-231 cells? This thesis will attempt to address some of these questions regarding vertebrate Tra2 protein biology.

Chapter 2: Functional dissection of splicing regulation by the RNA-binding protein Tra2β

2.1 Introduction

Tra2β is implicated in male germ cell development and an on-going project in the Elliott lab has been to identify physiological targets of Tra2β in the testis. The development of new technologies such as UV cross-linking and immunoprecipitation coupled with high-thoughput sequencing (HITS-CLIP) has facilitated the identification of direct targets of RNA-binding proteins on a transcriptome-wide scale (Ule et al., 2005b). Prior to the start of my study, a previous PhD student in the lab (Dr. Yilei Liu) had recently performed a HITS-CLIP experiment to identify direct RNA targets of Tra2β in the mouse testis. In this chapter, I use data generated from that original HITS-CLIP experiment to validate splicing regulation of novel Tra2β target exons and further dissect their regulation using minigenes and binding site mutagenesis.

The *Tra2b* gene itself is alternatively spliced to produce five known mRNA transcripts, encoding two known protein isoforms (please refer to Figure 1.8) (Stoilov *et al.*, 2004). Full-length Tra2β (termed Tra2β-1) is ubiquitously expressed and contains a single, central RNA recognition motif (RRM) which is flanked by N-terminal and C-terminal RS-domains (Dauwalder *et al.*, 1996; Beil *et al.*, 1997). A second, truncated protein isoform (termed Tra2β-3) is also produced via alternative splicing, which lacks the N-terminal RS1 domain (Stoilov et al., 2004). It is currently unclear whether Tra2β-1 and Tra2β-3 are functionally distinct in splicing regulation. Therefore in this chapter, I also investigate the functional importance of the RS1 of Tra2β in splicing regulation using over-expressed proteins.

Tra2β is up-regulated in several human cancers and we hypothesised that Tra2β may regulate alternative splicing of functionally important genes in breast cancer. In the final part of this chapter, I investigate splicing regulation of a panel of breast cancer-associated alternative splice events using minigenes. I also analyse the splicing profile of endogenous targets exons following siRNA-mediated depletion of Tra2β in MDA-MB-231 cells.

2.2 Aims

The aims of this chapter were to:

1. Validate splicing regulation of target exons identified from the Tra2β HITS-CLIP experiment
2. Investigate whether multiple Tra2β binding sites are required for splicing activation of a target exon from *Nasp*
3. Investigate whether Tra2β-1 and Tra2β-3 may have distinct functions in splicing regulation using over-expressed proteins
4. Investigate whether direct RNA-binding is required for splicing activation of target exons by Tra2β
5. Investigate whether Tra2β regulates splicing of a selection of breast cancer-associated alternative splicing events
6. Investigate the splicing profile of endogenous target exons (corresponding to the Tra2β-responsive minigenes) following Tra2β protein depletion using RNAi

2.3 Materials and Methods

2.3.1 Cell culture

2.3.1.1 Cell lines

Three human cell lines were used in this chapter; HEK-293 cells (catalogue number: ATCC-CRL-1573), MCF-7 cells (catalogue number: ATCC-HTB-22) and MDA-MB-231 cells (catalogue number: ATCC-HTB-26). All three cell lines were originally purchased from the American Type Culture Collection and LGC Standards, Europe.

2.3.1.2 HEK-293

HEK-293 is a transformed human embryonic kidney cell line (Graham *et al.*, 1977). This cell line is routinely used in splicing assays due to the high transfection efficiency observed in this cell line.

2.3.1.3 MCF-7

MCF-7 is a tumourigenic breast epithelial cell line originally derived from pleural effusion of a 69 year old female patient with breast adenocarcinoma (Soule HD, 1973). The phenotypic characteristics of this cell line are early-stage, non-invasive, ER positive and PR positive breast cancer.

2.3.1.4 MDA-MB-231

MDA-MB-231 is a tumourigenic breast epithelial cell line originally derived from pleural effusion of a 51 year old female patient with breast adenocarcinoma (Cailleau R, 1978b). The phenotypic characteristics of this cell line are invasive, triple-negative (ER-, PR-, HER2-) breast cancer.

2.3.1.5 Routine cell passage

Cell culture was performed under aseptic conditions in a Class II laminar flow microbiological safety cabinet. All cell lines were routinely cultured in 75cm^2 and 25cm^2 tissue culture flasks at 37°C in a humidified atmosphere containing 5% CO^2. HEK-293 cells were maintained in DMEM (with phenol red) (PAA) supplemented with 10% FBS (Sigma-Aldrich). MDA-MB-231 cells and MCF-7 cells were maintained in

DMEM (without phenol red) (PAA) supplemented with 10% FBS and 1% Penicillin Streptomycin (Sigma-Aldrich).

2.3.1.6 Cell line maintenance

Cells were passaged every 3 to 5 days at approximately 70-80% confluency. Cell passage was performed by removing growth media, rinsing the cell monolayer with sterile 1× phosphate buffered saline (PBS) (Sigma-Aldrich) and incubating with 2mM trypsin-EDTA (Sigma-Aldrich) for 5 minutes at 37°C. Complete growth media was added to neutralise the effect of trypsin-EDTA and detached cells were collected by centrifugation at 200×g for 5 minutes. The supernatant was discarded before the pelleted cells were resuspended in complete growth media and passaged at a ratio of approximately 1 to 4.

2.3.1.7 Cryopreservation of cells

Cells were routinely cryopreserved at early passage numbers to generate a continuous stock of frozen cells from early passage. Cryopreservation was performed in 1ml aliquots of cryoprotective media and stored in cryovials (Sigma-Aldrich). Cryoprotective media consisted of 95% FBS with 5% dimethyl sulphoxide (DMSO) (Sigma-Aldrich). Cells were stored at -80°C for long term storage. When required, frozen stocks were rapidly thawed in a 37°C water bath and freezing media was removed by centrifugation at 200×g for 5 minutes. Cells were then resuspended in complete growth medium and plated in sterile tissue culture flasks.

2.3.1.8 Cell counting

Cells were counted prior to experiments using a Neubauer chamber haemocytometer. Cell pellets were resuspended in complete growth media and a 10µl aliquot of single cell suspension placed onto the haemocytometer. The number of cells overlying the ruled grid area was counted using low power magnification (×10). The number of cells per mililiter was calculated and the cell suspension was diluted appropriately to seed the correct number of cells for each experiment.

2.3.2 Minigene splicing assays

2.3.2.1 Minigene transfections

The pXJ41 minigene constructs were co-transfected with several pGFP3 expression vectors into HEK-293 cells for splicing analysis. All minigene transfections were performed using three biological replicates. HEK-293 cells were seeded into 6-well plates at approximately 50-80% confluency 24 hours before transfection. A standard co-trasfection protocol consisting of 200ng minigene together with either 200ng GFP-only expression vector or 500ng expression vector was used throughout this chapter, unless otherwise stated. DNA plasmids were incubated at room temperature for 20 minutes with 3µl GeneJammer and 97µl Opti-MEM (Life Technologies), before adding the mix to HEK-293 cells grown in 6-well plates. Cells were then incubated for 24 hours at 37°C. Prior to harvesting, cells were observed under a fluorescent microscope to confirm efficient expression of GFP-tagged proteins. To harvest cells the media was removed, cells were gently washed with 1ml PBS and incubated with 300µl Trypsin-EDTA for 5 minutes at 37°C. Once cells had detached from the plate surface, 1ml DMEM was added to each well. The resuspended cells from each well were split into two aliquots and centrifuged at 3,000rpm to form cell pellets for RNA and protein analysis. Cell pellets for RNA analysis were thoroughly resuspended in 100µl Trizol (Life Technologies) prior to extraction, whilst cell pellets for protein analysis were directly lysed in 30-50µl 2X SDS sample loading buffer.

2.3.2.2 RNA extraction

A standard Trizol RNA extraction was performed using 100µl Trizol (Life Technologies) per cell pellet. Cells were resuspended in 100µl Trizol and incubated at room temperature for 5 minutes. 20µl chloroform was added and thoroughly mixed. Samples were incubated at room temperature for 10 minutes and centrifuged at 13,000rpm for 10 minutes. The top aqueous layer was removed (approximately 50µl), transferred to a new eppendorf and an equal volume of isopropanol was added. Samples were incubated once more at room temperature for 15 minutes and centrifuged at 13,000rpm for 10 minutes at 4°C, forming a white RNA pellet. The supernatant was carefully removed, and the pellet washed with 70% ethanol and allowed to air dry. Finally, samples were resuspended in 30µl Diethylpyrocarbonate

(DEPC)-treated dH_20. The RNA concentration was quantified via a Nanodrop spectrophotometer. RNA samples were diluted to 50ng/µl with RNase-free dH_20 prior to RT-PCR.

2.3.2.4 One-step RT-PCR

RNA samples were reverse transcribed and PCR-amplified using two pXJ41 minigene-specific primers (pXJRTF and pXJB1) in a single one-step RT-PCR reaction, using a One-step RT-PCR kit (Qiagen) following the manufacturer's instructions. A standard 5µl One-step RT-PCR reaction was used per sample and is provided in Figure 2.1, together with the pXJRTF and pXJB1 primer sequences.

5X reaction buffer	1µl
Q-solution	1µl
pJXRTF primer (10µM)	0.3µl
pJXB1 primer (10µM)	0.3µl
dNTPS (10mM)	0.2µl
Enzyme mix	0.2µl
RNA (50ng/µl)	2µl
TOTAL	5µl

Primer	Sequence
pXJRTF (forward)	GCTCCGGATCGATCCTGAGAACT
pXJB1 (reverse)	AGCAGAACTTGTTTATTGCAGC

Figure 2.1 Standard 1X 5µl One-step RT-PCR master mix used in minigene splicing assays (upper table). pXJRTF and pXJB1 primer sequences (lower table).

2.3.2.5 Capillary gel electrophoresis (QIAxcel)

The 5µl RT-PCR samples were diluted with 5µl QIAxcel DNA dilution buffer (Qiagen) and electrophoresed using the QIAxcel multi-capillary electrophoresis system (Qiagen) for analysis. Samples were analysed using the QIAxcel Biocalculator software (Qiagen) to determine the size of each PCR product (bp) and the relative concentration of each band (ng/ul).

2.3.2.6 Calculation of percentage splicing inclusion (PSI)

The percentage splicing inclusion (PSI) used throughout this thesis was calculated using the following formula:

$$\frac{\text{Concentration of 'exon included' PCR band (ng/ul)}}{\text{(Concentration of 'exon included' + 'exon excluded' PCR bands (ng/ul))}} \text{ X100 = PSI (\%)}$$

2.3.3 Western Immunoblotting

Efficient GFP-tagged protein expression was determined by Western Immunoblotting, using a polyclonal mouse α-GFP antibody (Abcam, ab1218) (1:2000 dilution) and α-mouse HRP-linked secondary antibody (Amersham, NA931VS) (1:2000 dilution). Cell pellets were lysed in 2X SDS sample loading buffer and denatured at 100°C for 5 minutes. Proteins were separated on 10% SDS-PAGE gels and transferred to a PVDF membrane (Hybond-P, GE). The membrane was first blocked in blocking solution (Tris-Buffered Saline with Tween 20 (TBST) containing 5% non-fat dry milk) for 1 hour and subsequently probed for 1 hour with the primary antibody diluted in blocking solution. Membranes were washed three times for 5 minutes with TBST and then incubated for 1 hour with a secondary antibody conjugated to horseradish peroxidise. Membranes were washed a further three times for 5 minutes with TBST before enhanced chemiluminescent (ECL) detection using an ECL Prime Western Blotting Detection Kit (Amersham). Efficient knockdown of endogenous Tra2β was also confirmed by Western Immunoblotting using a rabbit α-Tra2β antibody (Abcam, ab31353) (1:2000 dilution) and a mouse α-Tubulin antibody (Sigma-Aldrich, T5168) (1:2000 dilution) as a loading control.

2.3.4 Mutation of Tra2β binding sites

2.3.4.1 Site-directed mutagenesis of the *Nasp-T* minigenes

Site-directed mutagenesis was used to mutate Tra2β binding sites in the wildtype *Nasp-T* pXJ41 minigene. The original wildtype *Nasp-T* minigene was previously cloned by Mrs. Caroline Dalgliesh (Newcastle University). In total, seven variants of the *Nasp-T* minigene (designated M1, M2, M3, M4, M1+M2, M2+M3, M3+M4) were created. Mutagenic primers (Table 2.1) were designed to introduce 2 or 3 single base mutations

within Tra2β binding sites in each minigene. Site-directed mutagenesis utilises two PCR reactions (PCR A and PCR B) to create overlapping fragments containing the desired complementary mutation on opposite strands of the PCR fragments. A third reaction (PCR C) using both PCR A and PCR B as a template is then used to create a full-length insert with the mutation on both strands. The full-length mutated insert was then cloned into an empty pXJ41 vector. A schematic representation of the site-directed mutagenesis strategy is provided in Figure 2.2.

Primer	Sequence
NASP M1 F	GGGTGGACGATAAGACATGG
NASP M1 R	CCATGTCTTATCGTCCACCC
NASP M2 F	GTGAGCCTCAAGAGTAGCTCC
NASP M2 R	GGAGCTACTCTTGAGGCTCAC
NASP M3 F	GAATCCTCTGCATAGGCAAAAG
NASP M3 R	CTTTTGCCTATGCAGAGGATTC
NASP M4 F	GGACTGACTCAAGTTGAGGTCGC
NASP M4 R	GCGACCTCAACTTGAGTCAGTCC

*Bases highlighted RED indicate sites where single base substitutions were introduced

Table 2.1 Primer sequences used for site-directed mutagenesis of the wildtype *Nasp-T* minigene.

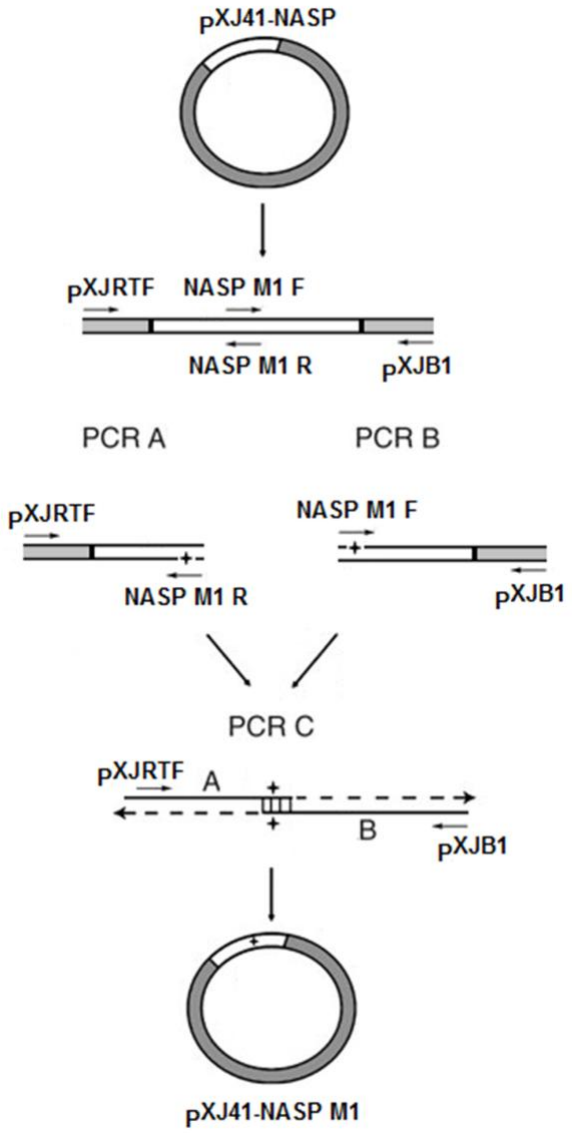


Figure 2.2 Schematic representation of site-directed mutagenesis of the wildtype *Nasp-T* minigene. Site-directed mutagenesis utilises two PCR reactions (PCR A and PCR B) to create overlapping fragments containing the desired complementary mutation on opposite strands of the PCR fragments. These fragments are mixed, then a third reaction (PCR C) using both PCR A and PCR B as a template and flanking primers (pXJRTF and pXJB1) is then used to create a full-length insert with the mutation on both strands. The full-length mutated insert was then cloned into an empty pXJ41 vector. This image is adapted from the published protocol by (Heckman and Pease, 2007).

Each PCR was performed in a 10µl reaction using a Phusion High-Fidelity DNA polymerase kit (Thermo Scientific). The PCR master mix and reaction conditions used for site-directed mutagenesis are provided in Table 2.2.

PCR reaction A

5X HF Reaction Buffer	2µl
Template (wildtype minigene, 20ng/µl)	0.5µl
pXJRTF Forward Primer (10µM)	1µl
Mutagenic Reverse Primer (10µM)	1µl
dNTPs (10mM)	0.2µl
dH_20	5.2µl
Phusion DNA polymerase	0.1µl
TOTAL	10µl

PCR reaction B

5X HF Reaction Buffer	2µl
Template (wildtype minigene, 20ng/µl)	0.5µl
Mutagenic Forward Primer (10µM)	1µl
pXJB1 Reverse Primer (10µM)	1µl
dNTPs (10mM)	0.2µl
dH_20	5.2µl
Phusion DNA polymerase	0.1µl
TOTAL	10µl

PCR reaction C

5X HF Reaction Buffer	2µl
Template 1 (PCR Fragment A)	0.5µl
Template 2 (PCR Fragment B)	0.5µl
pXJRTF Forward Primer (10µM)	1µl
pXJB1 Reverse Primer (10µM)	1µl
dNTPs (10mM)	0.2µl
dH_20	4.7µl
Phusion DNA polymerase	0.1µl
TOTAL	10µl

Thermocycler programme

1 Heat Activation	95°C	5 minutes
2 Denaturation	95°C	20 seconds
3 Annealing	58°C	30 seconds
4 Extension	72°C	1 minute (Cycling to step 2 x29)
5 Final Extension	72°C	10 minutes
6 Cooling	4°C	∞

Table 2.2 PCR master mix for reactions A, B and C, and thermocycler programme used in site-directed mutagenesis of minigenes.

The full-length PCR products from PCR reaction C were electrophoresed on a 2% agarose gel and isolated using a QIAquick Gel Extraction Kit (Qiagen) following the

manufacturer's instructions and eluted in 30µl dH_20. The purified inserts and the pXJ41 empty vector were then digested with two restriction enzymes *Cla*1 and *Bam*H1 (NEB). Restriction digest reactions were incubated at 37°C for 3 hours and subsequently purified using QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's instructions. The DNA concentration of each sample after the restriction digest was determined using a Nanodrop spectrophotometer. Reaction conditions for the restriction digests are provided below in Table 2.3.

PCR C or pXJ41 empty vector	26µl
Buffer 4	5µl
BSA (X10)	5µl
*Cla*1 restriction enzyme	2µl
*Bam*H1 restriction enzyme	2µl
TOTAL	50µl

Table 2.3 Reagents used for the restriction digest of PCR product C and the empty pXJ41 vector.

Following purification of the digested full-length insert (PCR product C) and the pXJ41 empty vector, the two fragments were ligated using a T4 DNA ligase (NEB). The ligation reaction conditions are provided below in Table 2.4. Ligation reactions were incubated at 16°C for 16 hours, followed by 65°C for 10 minutes for heat deactivation of the ligation enzyme.

Insert (PCR C)	6-8µl
Vector (empty pXJ41)	0.5-1µl
T4 DNA Ligase	1µl
T4 Buffer	1µl
dH_20	0-1.5µl
TOTAL	10µl

Table 2.4 Reagents used for ligation of PCR product C into the empty pXJ41 vector.

2.3.4.2 Molecular cloning

Ligated plasmids were transformed in α-Select Chemically Competent Cells (BIOLINE) using the manufacturer's heat shock protocol (2µl plasmid incubated with 20µl competent cells). LB agar plates containing 50µg/ml ampicillin were used to grow and select competent cells which had been transformed with the pXJ41 minigene containing an ampicillin-resistance gene. The LB agar plates were incubated at 37°C for

16 hours. Single colonies from the plates were picked and subsequently grown in 5 ml LB plus 50μg/ml ampicillin. Tubes were incubated in a shaker for an overnight incubation at 37°C. The cultures were subsequently centrifuged at 5,000rpm for 2 minutes to pellet the cells. Plasmid DNA was extracted from cells using a QIAprep Miniprep Kit (Qiagen) following the manufacturer's instructions.

2.3.4.3 Sequencing

Plasmids were sent for DNA sequencing at Source Bioscience Lifescience Services. Samples were diluted to 100ng/μl and sequenced using both the pXJRTF forward primer and pXJB1 reverse primers (10μM) to confirm the correct sequences had been cloned into each minigene.

2.3.5 *In silico* prediction of Tra2β targets from a panel of breast cancer associated alternative splicing events

Potential targets for splicing regulation by Tra2β were identified from a breast cancer and ovarian cancer-associated alternative splicing resource (Venables et al., 2009a). Exons were first ranked in order of the percentage increase in exon inclusion in breast cancer tumours compared to their corresponding normal tissue. The 10 most frequently occurring K-mers from the Tra2β HITS-CLIP experiment were then mapped onto the exon sequences of the top 50 exons with increased splicing inclusion using a word search tool. In total, six exons (which were significantly up-regulated in breast cancer tumours compared to corresponding normal tissue) were identified which contained multiple predicted Tra2β binding sites and were therefore selected as candidate targets of splicing regulation by Tra2β.

2.3.6 Molecular cloning of the breast cancer associated exons into the pXJ41 minigene

Six breast cancer-associated alternative splice events and 2 exons with alternative 3' splice sites (identified as candidates from the murine Tra2β HITS-CLIP experiment) were cloned into the pXJ41 minigene for splicing analysis. Inserts were created using primers that flank the alternative exons and approximately 300bp upstream and downstream intronic regions using the UCSC genome browser. Each primer contained an additional *Bam*HI restriction site for cloning into the pXJ41 minigene (primers are

provided below in Table 2.5). Either 50ng human DNA (breast cancer-associated alternative splice events) or 50ng mouse DNA (alternative 3' splice sites) was used as a template for a 10µl PCR reaction using a Phusion High-Fidelity DNA polymerase (Thermo Scientific). PCR products were electrophoresed on a 2% agarose gel, isolated using QIAquick Gel Extraction kit following manufacturer's instructions and eluted in 30µl dH_20. Restriction digests using the *Bam*HI restriction enzyme were incubated at 37°C for 3 hours and subsequently purified using QIAquick PCR Purification Kit following the manufacturer's instructions. The digested inserts were then cloned into the pXJ41 minigene as described earlier in section 2.3.4.2.

Primer	Sequence
MLYRM1F	AAAAAAAAAGAATTCGGAGATGGCAAGTGCCTTTA
MLYRM1R	AAAAAAAAAGAATTCACCAGAAGGCTTTGAATGCT
MADAM17F	AAAAAAAAAGAATTCTTTTAAACCAGAACACTTCCTCA
MADAM17R	AAAAAAAAAGAATTCAGGCAGAGGCAGTAGGATCA
HPDLIM5F	AAAAAAAAAGAATTCTGACACTGATGAGGGGCTTT
HPDLIM5R	AAAAAAAAAGAATTCCAATGCTGTGTCGGGACTTA
HSLC10A7F	AAAAAAAAAGAATTCCCCAAAGTTGCACTTGGAAT
HSLC10A7R	AAAAAAAAAGAATTCTTGGTGACACAGTGCTTAATCT
HMCATF	AAAAAAAAAGAATTCAACAGCCCAGGCAGATTAGA
HMCATR	AAAAAAAAAGAATTCGAAATGAACTCCAGGGCAAA
HSDHCF	AAAAAAAAAGAATTCGTTGCCCAGGCTTGAGTG
HSDHCR	AAAAAAAAAGAATTCGGGAACTCTGTATGTTATTTCTGC
HCLK1F	AAAAAAAAACAATTGGTTTTGTTGCCTTGCCACTT
HCLK1R	AAAAAAAAACAATTGATGATCGATGCACTCCACAA
HSYTL2F	AAAAAAAAAGAATTCTGGCCTTGAGTTGAAACAGA
HSYTL2R	AAAAAAAAAGAATTCGCACATACAACTTCACCTTATGC

Table 2.5 Primer sequences used to clone exons and flanking intronic regions into the pXJ41 minigene.

2.3.7 siRNA transfection

Depletion of the endogenous Tra2α and Tra2β proteins was achieved by transfecting MDA-MB-231 cells with Silencer Select Pre-designed siRNAs (Ambion). siRNAs targeting the *TRA2A* mRNA (Ambion IDs: s26664 and s26665) and the *TRA2B* mRNA (Ambion IDs: s12749 and s12751) were transfected using siPORT NeoFX Transfection Agent (Ambion). Control cells were transfected with the same concentration of negative control siRNA (Ambion Cat#: 4390843). Untreated cells and cells treated with the transfection reagent only were also used as additional controls in some

experiments. Cells were transfected with either 24µl of 10µM negative control siRNA, 12µl of 10µM siRNA targeting *TRA2A* or *TRA2B* (single Tra2α or Tra2β depletion) or 12µl of 10µM siRNA targeting *TRA2A* and 12µl of 10µM siRNA targeting *TRA2B* (joint Tra2α and Tra2β depletion). For each transfection, 30µl siPORT NeoFX was added to 600 µl Opti-MEM media and incubated for 10 minutes. The appropriate siRNA was diluted in a further 600µl Opti-MEM and subsequently added to the transfection reagent mix and incubated for a further 20 minutes at room temperature. Each siPORT/siRNA mix was then added to a 100mm tissue culture plate, to which 10ml of MDA-MB-231 cell suspension was added. Cells were seeded at a density of approximately $3x10^5$ cells per 100mm plate. Cells were incubated for 72 hours following siRNA transfection, after which cells were harvested for both RNA extraction and Western Immunoblotting as described earlier.

2.3.8 Splicing analysis of endogenous target exons

RNA was extracted using the standard Trizol RNA extraction described in section 2.3.2.2. Subsequently, cDNA was synthesised from 500ng total RNA in a 10µl reaction, using a SuperScript VILO cDNA synthesis kit (Life Technologies) following the manufacturer's instructions. The splicing profiles of endogenous targets were monitored by PCR, using primers designed within flanking constitutive exons. The primer sequences for the endogenous splicing analysis are provided in Table 2.6. In each PCR, 1µl of diluted cDNA (1 in 8 dilutions) was used as the template in a standard 10µl PCR reaction using a Phusion High-Fidelity PCR Kit (Thermo Scientific) following the manufacturer's instructions. The splicing profiles were monitored and quantified using the QIAxcel capillary gel electrophoresis system (Qiagen) and the percentage splicing inclusion (PSI) of endogenous targets were calculated as described earlier in section 2.3.2.6.

Primer	Sequence
CLK1 F	AGCAAACACAGGATTCACCAC
CLK1 R	TGATCGATGCACTCCACAAC
PDLIM5 F	AGCCGGTTCCTGTTCAAAAG
PDLIM5 R	TGCTGTTTACTGTCACCCTGG
LYRM1 F	GAGAAGGAGAAGCCAGCAAA
LYRM1 R	CCAATTTCAATCCTGGCTGT
TRA2A F	TTGCCGACTCTTTCCTCTTC
TRA2A internal	CGTGTATTCTTCCCCAATTCA
TRA2A R	AGGAGTTCCCGTTGGAGATT
NASP F	GAATGGTGTGTTGGGAAACG
NASP R	TTTGGCATTTCTTCGGTCTT
NASP internal	TTCCACCCTTCTCCATTTCA

Table 2.6 Primer sequences used in the splicing analyses of endogenous targets exons.

2.4 Results

2.4.1 Investigating splicing regulation of exons identified from the Tra2β HITS-CLIP experiment

2.4.1.1 Investigating splicing regulation of candidate cassette exons using a minigene splicing assay

Prior to the start of my study, an analysis of the Tra2β HITS-CLIP experiment (previously performed by Dr. Yilei Liu) identified 741 cassette exons containing CLIP sequencing reads or 'CLIP tags' from the mouse testis. I began my PhD project following on from this initial HITS-CLIP experiment, with the aim of identifying functionally responsive exons from the catalogue of exons containing Tra2β CLIP tags. The work in this chapter contributed to a larger project, which has now been published in PLoS Genetics (Grellscheid et al., 2011a).

To test whether Tra2β directly regulated splicing inclusion of a selection of exons containing Tra2β CLIP tags, 8 cassette exons (together with approximately 300bp of flanking intronic sequences), had been previously cloned into the pXJ41 minigene vector by Mrs. Caroline Dalgliesh (Newcastle University) for splicing analysis. I subsequently transfected each of the minigenes into HEK-293 cells, along with pGFP3 expression constructs encoding either a GFP-only control or full-length Tra2β-GFP. The splicing pattern of the RNA expressed from each minigene was then analysed by RT-PCR and capillary gel electrophoresis. Efficient expression of the GFP-tagged constructs was confirmed by fluorescence microscopy and Western Immunoblotting. Each transfection was performed in triplicate to determine whether any observed changes in splicing were reproducible and statistically significant. Co-expression with 500ng of the Tra2β-GFP expression vector was found to significantly increase the percentage splicing inclusion (PSI) of 6 out of the 8 minigenes when compared to transfection with the GFP-only control (Figure 2.3). The Tra2β-responsive minigenes contained cassette exons cloned from *Nasp*, *Krba1*, *Fabp9*, *Creb*, *Lin28b* and *Tra2a* (see Figures 2.3 A, B, D, E, F, H respectively). A minigene containing a cassette exon from *Pank2* responded weakly to co-transfection with the Tra2β-GFP expression vector, but this change was not statistically significant (Figure 2.3 G). A minigene containing *Creb* exon 2 did not respond to co-expression of Tra2β-GFP (Figure 2.3 C).

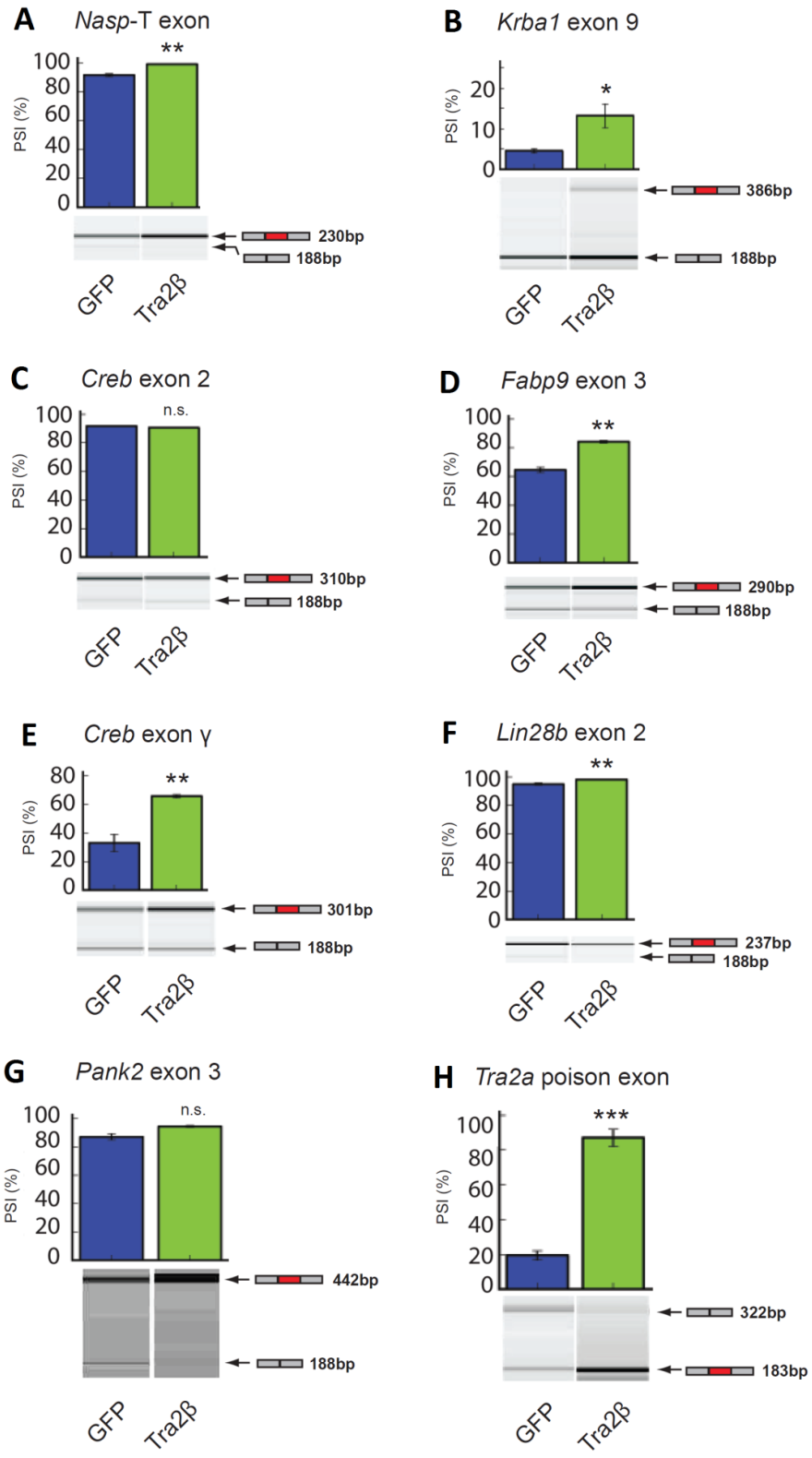


Figure 2.3 Splicing patterns of 8 minigenes containing candidate exons analysed by RT-PCR and capillary gel electophoresis. The percentage splicing inclusion (PSI) of each minigene exon was compared between cells transfected with either GFP-only (blue) or full-length Tra2β-GFP (green). PCR product sizes are provided on right-hand side of each gel. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is adapted from (Grellscheid et al., 2011a).

2.4.1.2 Tra2β activates splicing inclusion of a poison exon from *Tra2a*

Intriguingly, the most highly responsive minigene to co-expression of the full-length Tra2β-GFP protein contained a 306bp poison exon from the *Tra2a* gene [N.B. An internal screening primer was used for the *Tra2a* minigene RT-PCR, therefore counterintuitively, the PCR product respresenting poison exon inclusion was shorter (183bp) than the PCR product respresenting exon exclusion (322bp) for this particular minigene] (Figure 2.3 H). Poison exons introduce premature termination codons (PTCs) into mRNA transcripts, thereby targeting the transcript for degradation through the non-sense mediated decay (NMD) pathway (Lareau *et al.*, 2007; Ni *et al.*, 2007). Intriguingly, this inital minigene data suggested that Tra2β may negatively regulate expression of the Tra2α protein, by promoting inclusion of a highly responsive poison exon within the *Tra2a* mRNA. The *Tra2a* poison exon is studied in further detail in chapter 3.

2.4.1.3 Investigating Tra2β mediated splicing regulation of the *Nasp-T* exon using mutagenesis

One of the six Tra2β-responsive minigene exons was a particularly large (975 nucleotide) cassette exon from the *Nasp* gene. This alternatively spliced exon encodes part of a testis-enriched isoform of the Nasp protein and is subsequently referred to as the *Nasp-T* exon (Alekseev *et al.*, 2003) (Figure 2.5 A). To investigate the distribution of Tra2β binding sites within this large exon, the top 25 K-mers (25 most frequently recovered sequences) from the Tra2β HITS-CLIP experiment were mapped onto the *Nasp-T* exon (Figure 2.4). This Tra2β binding site analysis was done in collaboration with Dr. Sushma Grellscheid (School of Biological and Biomedical Sciences, Durham University). This analysis indicated that the *Nasp-T* exon contained a particularly high density of Tra2β binding sites throughout the exon.

Figure 2.4 The *Nasp-T* exon contains a high density of Tra2β binding sites throughout the exon. The top 25 K-mers (25 most frequently recovered sequences) identified from Tra2β HITS-CLIP experiment were mapped back to the *Nasp-T* exon to represent Tra2β binding sites within the exon. K-mers are shaded green. This figure is adapted from (Grellscheid et al., 2011a).

To investigate whether multiple Tra2β binding sites were required for splicing activation of the *Nasp-T* exon, a selection of Tra2β binding sites were mutated within the wildtype *Nasp-T* minigene using site-directed mutagenesis. I focused on an upstream region of the exon (close to the 3' splice site), which contained the highest number of Tra2β CLIP tags (see Figure 2.5 B). In order to mutate Tra2β binding sites without inadvertently creating splicing silencer sequences within the exon, I used the sequence analysis programme *Cis*-plotter (unpublished, developed by Dr. Sushma Grellsheid, Durham University). *Cis*-plotter predicts splicing enhancer and splicing silencers within a given sequence. *Cis*-plotter predicted four potential exonic splicing enhancer (ESE) sequences within this region, which directly overlapped with Tra2β CLIP tags (Figure 2.5 C). Each of the four predicted Tra2β binding sites were mutated to a sequence that would prevent Tra2β binding, but would not inadvertently introduce exonic splicing silencer (ESS) sequences. The four mutated versions of *Nasp-T* minigenes were then transfected into HEK-293 cells to assess the effect of each mutation on exon inclusion using RT-PCR and capillary gel electrophoresis (Figure 2.7).

2.4.1.4 Tra2β promotes splicing inclusion of the *Nasp-T* exon through multiple redundant binding sites

Mutation of single Tra2β binding sites (individual mutations are termed M1, M2, M3 and M4) within the *Nasp-T* minigene only had a minor effect on exon inclusion when co-transfected into HEK-293 cells with a GFP-only control. All four mutated versions of the *Nasp-T* minigene showed slightly reduced exon inclusion when compared to the wildtype minigene, although only the effect of mutations M2 and M4 were statistically significant (compare blue bars, Figure 2.6). Consequently, I created mutated versions of the *Nasp-T* minigene which combined two Tra2β binding site mutations (combined mutations are termed M1+M2, M2+M3 and M3+M4). A capillary gel electrophoresis image showing a comparison between the splicing pattern of the wildtype *Nasp-T* exon and the M3+M4 version of the *Nasp-T* minigene is shown in Figure 2.7.

Mutating two Tra2β binding sites resulted in a substantial reduction in exon inclusion for each combined mutation when compared to the wildtype minigene. Interestingly, different combinations of paired Tra2β binding site mutations within the *Nasp-T* exon had distinct effects on exon inclusion (compare version M2+M3 with GFP-only to version M3+M4 with GFP-only in Figure 2.6), suggesting the position of each binding site was important, distinct from the total number of binding sites mutated. In particular, the M3+M4 mutation strongly reduced exon inclusion to just 21%, whereas the M2+M3 mutation had less effect, reducing exon inclusion to approximately 78% (mean inclusion of the wildtype *Nasp-T* exon was 93%). Each of the combined mutations had a greater effect on splicing inclusion of the *Nasp-T* exon than the single mutations alone when transfected with the GFP-only control. However, each minigene containing a combined mutation still responded strongly to co-transfection of full-length Tra2β-GFP, suggesting Tra2β could still activate splicing inclusion through the other binding sites within the exon. Together, this data suggests that Tra2β is able to directly regulate splicing inclusion of the *Nasp-T* exon through multiple, redundant binding sites.

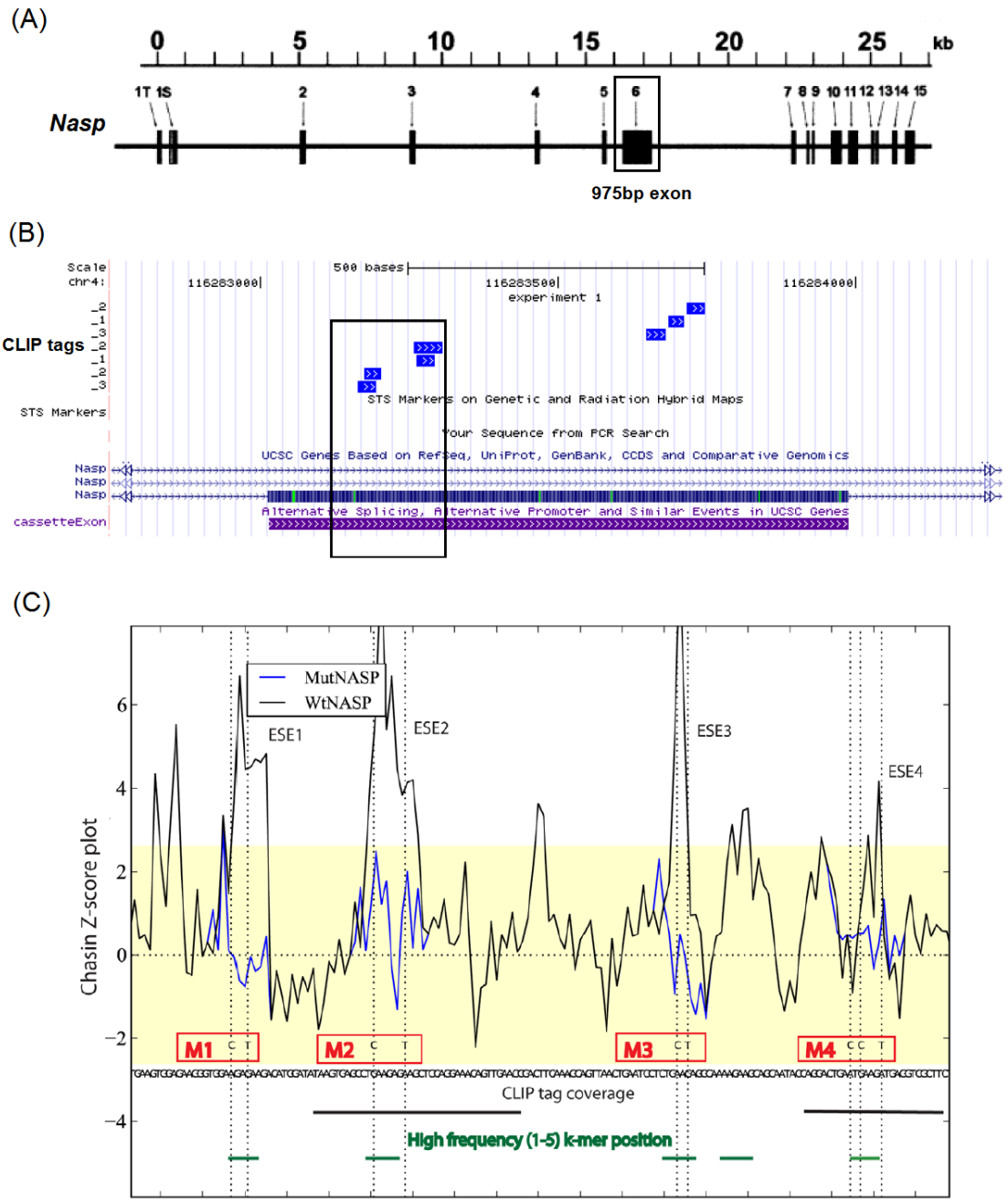


Figure 2.5 Mutagenesis of Tra2β binding sites within the wildtype *Nasp-T* minigene. (A) The 975bp *Nasp-T* exon within the *Nasp* gene. (B) Multiple Tra2β CLIP tags (blue bars) from the murine Tra2β HITS-CLIP experiment mapped directly within the *Nasp*-T exon. (C) The sequence analysis programme *Cis*-plotter was used to predict ESE and ESS sequences within the region of the *Nasp-T* exon containing the highest number of CLIP tags. Four ESE sequences were predicted (ESE1-4) and then mutated (red) to inhibit binding by Tra2β (M1, M2, M3 and M4). The Tra2β binding sites were mutated to a neutral sequence (shown in blue) to prevent inadvertently introducing splicing silencer sequences within the exon. Figure 2.5 part C is adapted from (Grellscheid et al., 2011a).

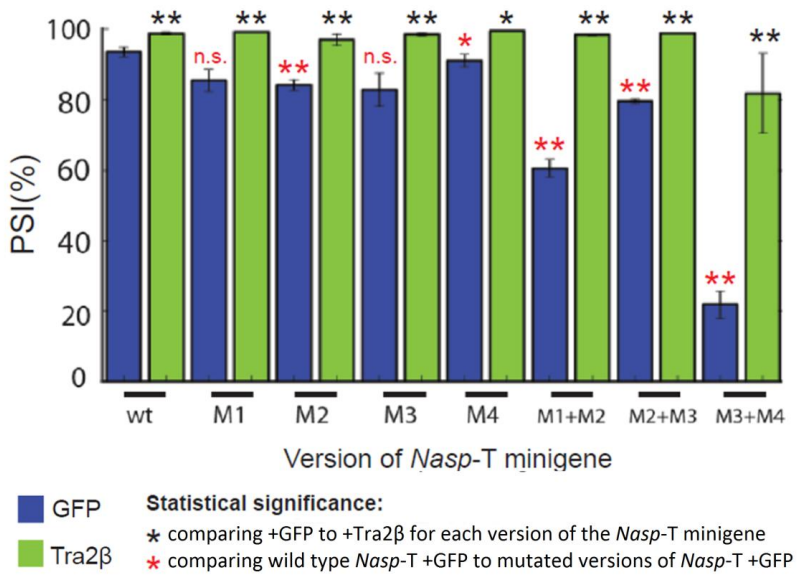


Figure 2.6 The percentage splicing inclusion (PSI) of different versions of the *Nasp-T* minigene exon. Each version of the *Nasp-T* minigene was co-transfected with either a GFP-only control (blue) or full-length Tra2β-GFP (green). Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is adapted from (Grellscheid et al., 2011a).

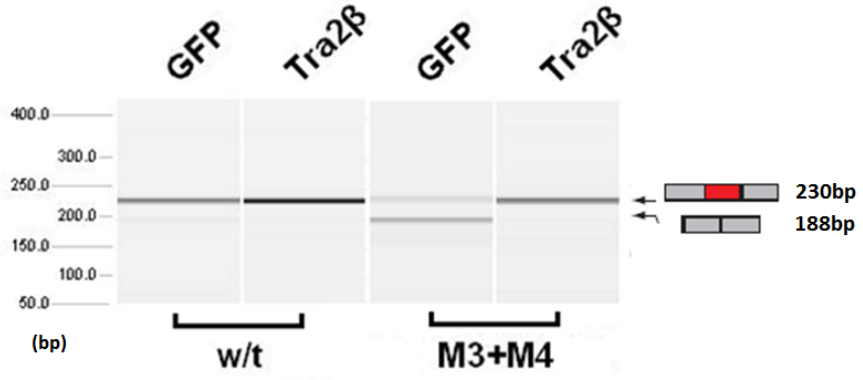


Figure 2.7 Capillary gel electrophoresis image comparing splicing patterns of the wildtype *Nasp-T* minigene to a mutated version (*Nasp-T* M3+M4). Due to the large size of the *Nasp-T* exon, an internal screening primer was used to produce a shorter PCR product (230bp) corresponding to exon inclusion. Each version of the minigene was co-transfected with either a GFP-only control or full-length Tra2β-GFP. Mutation of Tra2β binding sites (in version M3+M4) significantly reduced splicing inclusion of the *Nasp-T* exon at endogenous concentrations of Tra2β (compare columns 1 and 3). However, the M3+M4 version remained highly responsive to over-expression of full-length Tra2β (compare columns 3 and 4).

To determine whether the M3+4 version of the *Nasp-T* minigene would respond less efficiently to lower levels of ectopically expressed Tra2β, HEK-293 cells were co-transfected with either the wildtype *Nasp-T* minigene or the M3+M4 mutant version and a lower concentration gradient of ectopically expressed Tra2β (ranging from 0-50ng of Tra2β-GFP, instead of the typical 500ng) (Figure 2.8). Splicing inclusion of the wildtype *Nasp*-T exon was maximal (approximately 100% inclusion) after co-transfection with no more than 20ng of the Tra2β-GFP expression construct. In contrast, the mean PSI of the M3+M4 version increased more slowly across the concentration gradient. These results suggest that removal of just two Tra2β binding sites (from approximately 37 potential binding sites in total) is sufficient to significantly alter the concentration range of Tra2β to which the *Nasp-T* exon responds.

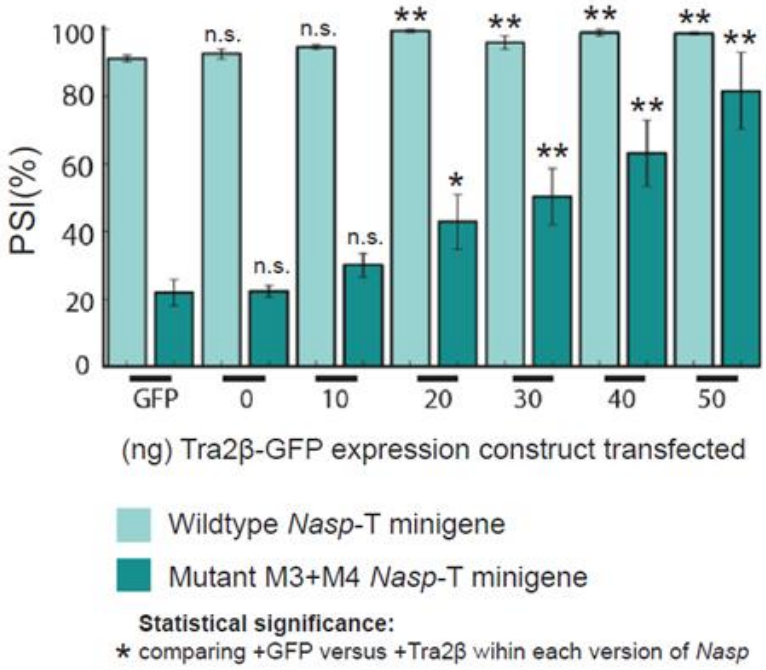


Figure 2.8 Tra2β activates inclusion of the *Nasp-T* exon in a concentration dependent manner. Mutating Tra2β binding sites (mutant M3+M4) significantly altered the sensitivity range to which the *Nasp-T* exon responded to Tra2β. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is adapted from (Grellscheid et al., 2011a).

To analyse direct protein-RNA interactions between Tra2β and the *Nasp-T* exon, an Electrophoretic Mobility Shift Assay (EMSA) was carried out by Mrs. Caroline Dalgliesh (Newcastle University) using purified full-length Tra2β protein and short radioactive RNA probes corresponding to a 155bp region of the *Nasp-T* exon (Figure 2.9). RNA probes corresponsing to the wildtype *Nasp-T* sequence and probes containing the M2 and M3+M4 mutations were used to determine the effect of the mutations on the affinity of Tra2β for this region of the *Nasp-T* exon. The mutant RNA probes were cloned from the *Nasp-T* M2 and M3+M4 minigenes I constructed earlier. The sequence of each RNA probe is shown in Figure 2.9 A.

The wildtype RNA probe was efficiently shifted by just 10ng of full-length Tra2β protein, demonstrating a strong interaction between Tra2β and the wildtype *Nasp-T* sequence (other Tra2β target exons required up to 200ng of full-length Tra2β protein to exert a similar shift) (Grellscheid et al., 2011a). The RNA probe containing the single M2 mutation shifted to a similar extent as the wildtype probe. However, the RNA probe containing the combined M3+M4 mutations shifted less efficiently than the wildtype and M2 probes, though it was still shifted considerably by just 10ng of full-length Tra2β protein.

The data from the EMSA experiment suggests that Tra2β has a very high affinity for the *Nasp-T* exon and directly interacts with the exon through multiple binding sites. The M2 and M3+M4 mutations did not substantially diminish the interaction between the purified Tra2β protein and the mutated *Nasp* RNA probes, suggesting Tra2β was still able to bind the RNA efficiently, most likely through the other predicted binding sites. This is consistent with the data from the *Nasp-T* minigenes, in which none of the binding sites were essential for Tra2β mediated splicing activation of the *Nasp-T* exon. Overall, the combined minigene and EMSA data suggest that the *Nasp-T* exon is highly sensitive to Tra2β protein expression and that Tra2β promotes splicing inclusion of the *Nasp-T* exon through multiple, redundant binding sites.

A

wild type *Nasp-T* RNA probe

GAAGUGGAGAAGGGUGGAAGAGAAGACAUGGAUAUAAGUGAGCCUGAAGAGAA
GCUCCAGGAAACAGUUGAACCGACUUCAAAGCAGUUAACUGAAUCCUCUGAAG
AGGCAAAAGAAGCAGCAAUACCAGGACUGAAUGAAGAUGAGGUCGCUUC

M2 *Nasp-T* RNA probe

GAAGUGGAGAAGGGUGGAAGAGAAGACAUGGAUAUAAGUGAGCCUCAAGAUAA
GCUCCAGGAAACAGUUGAACCGACUUCAAAGCAGUUAACUGAAUCCUCUGAAG
AGGCAAAAGAAGCAGCAAUACCAGGACUGAAUGAAGAUGAGGUCGCUUC

M3+M4 *Nasp-T* RNA probe

GAAGUGGAGAAGGGUGGAAGAGAAGACAUGGAUAUAAGUGAGCCUGAAGAGAA
GCUCCAGGAAACAGUUGAACCGACUUCAAAGCAGUUAACUGAAUCCUCUGAAG
AGGCAAAAGAAGCAGCAAUACCAGGACUGAAUGAAGAUGAGGUCGCUUC

B

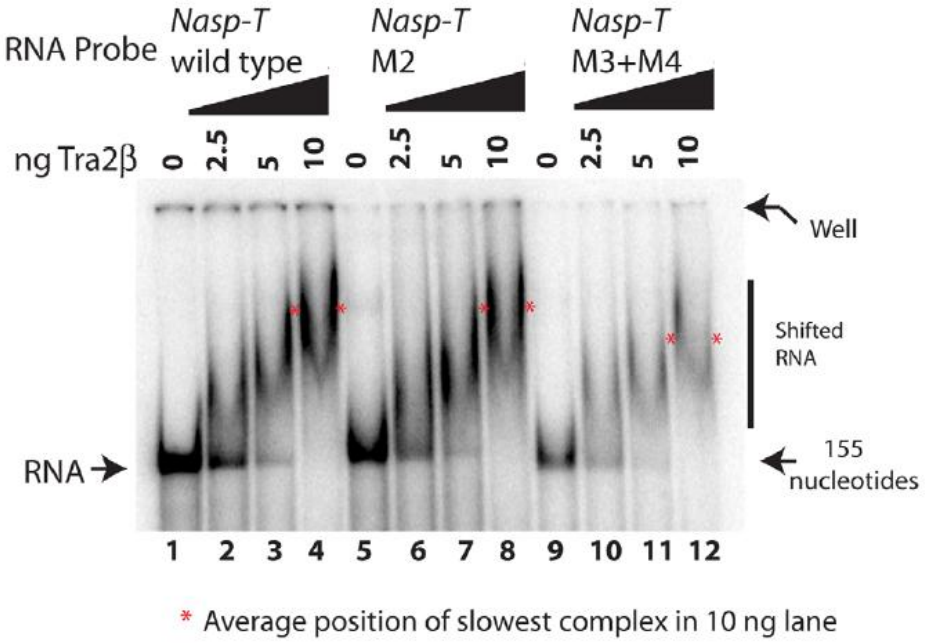


Figure 2.9 An Electrophoretic Mobility Shift Assay (EMSA) was used to examine direct interactions between Tra2β and different versions of the *Nasp-T* RNA. The Electrophoretic Mobility Shift Assay (EMSA) was carried out entirely by Mrs. Caroline Dalgliesh (Newcastle University) using purified full-length Tra2β protein and short radioactive RNA probes corresponding to a 155bp region of the *Nasp-T* exon. (A) The sequence of each of the three *Nasp* RNA probes (Tra2β binding sites highlighted in green, single base mutations highlighted in red). (B) The wildtype RNA probe was efficiently shifted by 10ng full-length Tra2β protein. The M2 probe shifted to a similar extent as the wildtype probe. The M3+M4 probe was less efficiently shifted than the wildtype and M2 probes. Figure 2.9 part B is adapted from (Grellscheid et al., 2011a).

2.4.2 Investigating the functional effect of removing the RS1 domain or RRM of Tra2β in splicing regulation

Having identified a panel of Tra2β-responsive exons, I next used these minigenes to investigate what effect removing either the RS1 domain or RRM from the Tra2β protein would have on splicing regulation. The panel of Tra2β responsive minigenes were co-transfected with pGFP3 expression vectors encoding truncated versions of the Tra2β protein. Five expression vectors were used in this study, which had been cloned previously by Mrs. Caroline Dalgliesh (Newcastle University). The pGFP3 expression vectors encode a GFP-only control, full-length Tra2β-GFP, full-length Tra2α-GFP, Tra2βΔRRM-GFP (lacking the RRM) and Tra2βΔRS1-GFP (lacking the RS1 domain). I confirmed efficient expression of each of the GFP-tagged proteins in HEK-293 cells by Western Immunoblotting using a GFP-specific antibody (Figure 2.10 A).

2.4.2.1 Direct RNA-binding was required for splicing activation of most Tra2β responsive minigenes

To investigate whether direct RNA-binding was required for splicing activation by the Tra2β protein, the splicing pattern of the Tra2β-responsive minigenes was compared between HEK-293 cells co-transfected with either full-length Tra2β-GFP (green columns) or Tra2βΔRRM-GFP (red columns) (Figure 2.10). Splicing activation of the *Tra2a* poison exon (Figure 2.10 B), *Lin28b* exon 2 (Figure 2.10 H), *Fabp9* exon 3 (Figure 2.10 I) and *Creb* exon γ (Figure 2.10 J) was completely lost when cells were co-transfected with Tra2βΔRRM-GFP as a substitute of full-length Tra2β-GFP (compare green and red columns, Figure 2.10). This data suggests that direct RNA-binding is required for efficient splicing activation of these exons. However, co-transfection with Tra2βΔRRM-GFP did significantly activate splicing inclusion of the wildtype *Nasp-T* minigene (Figure 2.10 C) and also slightly increased inclusion of the M3+M4 *Nasp-T* minigene (though this did not reach statistical significance and was significantly less than full-length Tra2β-GFP) (Figure 2.10 D). Therefore direct RNA-binding may not be essential for splicing activation of all target exons. Tra2β may also function as a splicing co-activator (enhancing exon recognition without directly binding to the RNA sequence) of some exons such as *Nasp-T*. The *Creb* exon 2 minigene, which did not respond to full-length Tra2β-GFP, similarly did not respond to co-transfection with Tra2βΔRRM-GFP (Figure 2.10 G).

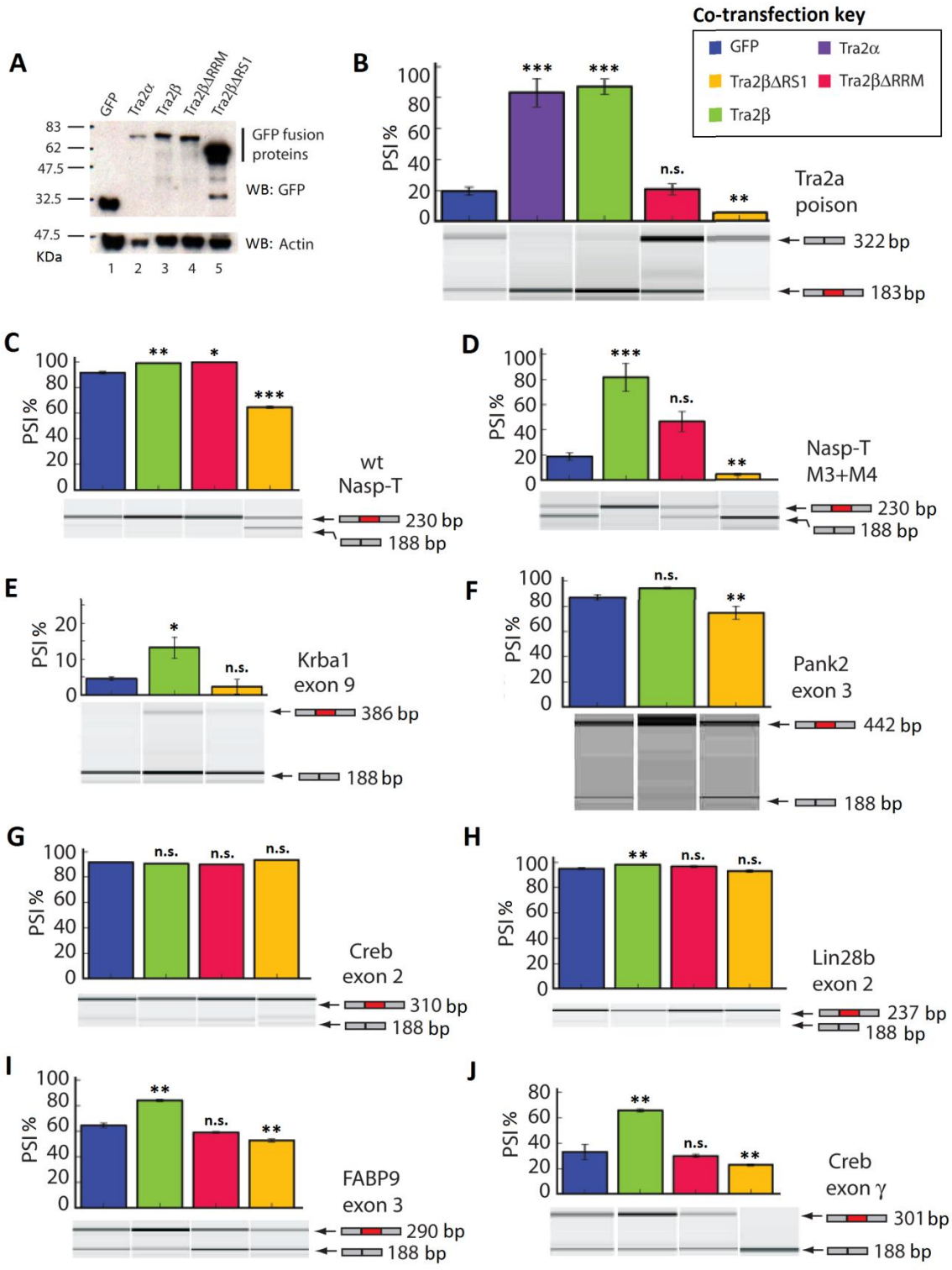


Figure 2.10 Co-transfection with different Tra2β expression constructs can activate, co-activate or repress splicing of the same target exons. (A) Efficient expression of the GFP-tagged fusion proteins was confirmed by Western Immunoblotting. (B-I) Splicing pattern of minigenes when co-transfected with GFP-only (blue), full-length Tra2β-GFP (green), full-length Tra2α-GFP (purple), Tra2βΔRRM-GFP (red) and Tra2βΔRS1-GFP (yellow). Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is adapted from (Grellscheid et al., 2011a).

2.4.2.2 A Tra2β isoform lacking the RS1 domain functioned as a splicing repressor

To investigate whether the RS1 domain of Tra2β was required for efficient splicing activation, the splicing pattern of the same Tra2β-responsive minigenes was compared between cells co-transfected with either GFP-only (blue columns), full-length Tra2β-GFP (green columns) or Tra2βΔRS1-GFP (yellow columns) (Figure 2.10). Unexpectedly, co-expression with Tra2βΔRS1-GFP (which lacks the RS1 domain) not only failed to activate splicing inclusion, but strongly repressed inclusion of nearly all Tra2β-responsive minigenes when compared to co-transfection with the GFP-only control (compare blue columns to yellow columns, Figure 2.10 B, C, D, E, F, I, J). Co-transfection with Tra2βΔRS1-GFP had no significant effect on splicing of the *Creb* exon 2 and *Lin28b* exon 2 minigenes compared to the GFP-only control. This data indicated that the RS1 domain of Tra2β is essential for splicing activation of the Tra2β target exons from this study. This data also suggests a possible novel function for the endogenous Tra2β-3 isoform (which also lacks the RS1 domain), which could potentially repress inclusion of target exons normally activated by the full-length Tra2β-1 protein. However, it is also possible that over-expression of the Tra2βΔRS1-GFP isoform in this minigene model system may amplify the effect of the Tra2βΔRS1 isoform. Further investigation is required to establish whether the endogenous Tra2β-3 protein exerts similar repressive activity.

2.4.3 Investigating regulation of alternative 3' splice sites identified from the Tra2β HITS-CLIP experiment

After cassette exons, the second most common class of alternative events associated with CLIP tags from the Tra2β HITS-CLIP experiment was alternative splice sites. In total, 432 alternative splice sites were identified within close proximity to Tra2β CLIP tags. To investigate whether Tra2β was involved in selection of alternative 3' splice sites, I cloned two exons from *Lyrm1* and *Adam17* which contained Tra2β CLIP tags mapping close to alternative 3' splice sites, together with their flanking intronic sequences, into the pXJ41 minigene vector for splicing analysis. An example of Tra2β CLIP tags (red bar) mapping close to the alternative 3' splice site of an exon from *Lyrm1* is shown in Figure 2.11.

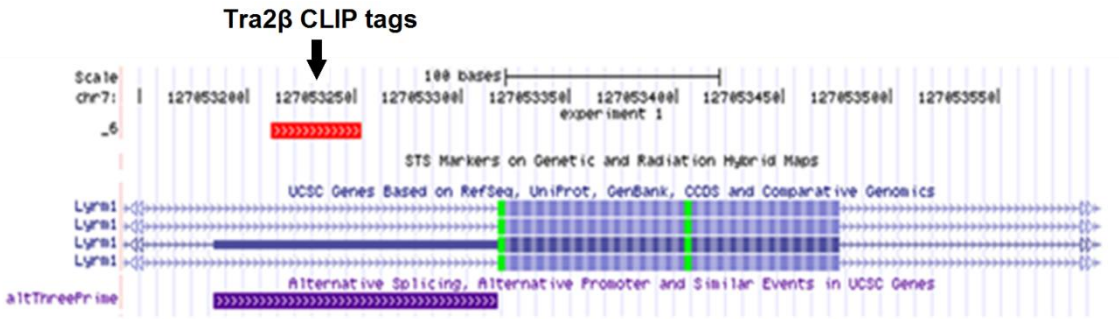


Figure 2.11 Tra2β CLIP tags (red bar) mapped within close proximity of an alternative 3' splice site of a cassette exon from *Lyrm1*.

Splicing analysis of the *Adam17* minigene by RT-PCR and capillary gel electrophoresis revealed 0% splicing inclusion of the *Adam17* cassette exon when transfected into HEK-293 cells with the GFP-only control (Figure 2.12 A, column 1). Co-transfection with full-length Tra2β-GFP had no effect on exon inclusion of the *Adam17* exon (Figure 2.12 A, column 2). Splicing analysis of the *Lyrm1* minigene revealed just two PCR bands, corresponding to the predicted sizes for exon inclusion (347bp) or exon exclusion (188bp) (Figure 2.12 B). The longer version of the cassette exon created using the alternative 3' splice site was not observed (predicted size 480bp). However, co-transfection with full-length Tra2β-GFP significantly increased splicing inclusion of the *Lyrm1* cassette exon when compared to co-transfection with the GFP-only control (compare columns 1 & 2, Figure 2.12 B). Co-transfection with Tra2βΔRS1-GFP also significantly repressed exon inclusion when compared to GFP-only control (compare columns 1 & 3, Figure 2.12 B). Although I did not detect selection of the alternative 3' splice site, this data suggests that the cassette exon within *Lyrm1* is another target of splicing regulation by Tra2β. As selection of the alternative 3' splice sites was not detected in the *Adam17* and *Lyrm1* minigenes, I decided to focus on other areas for further study.

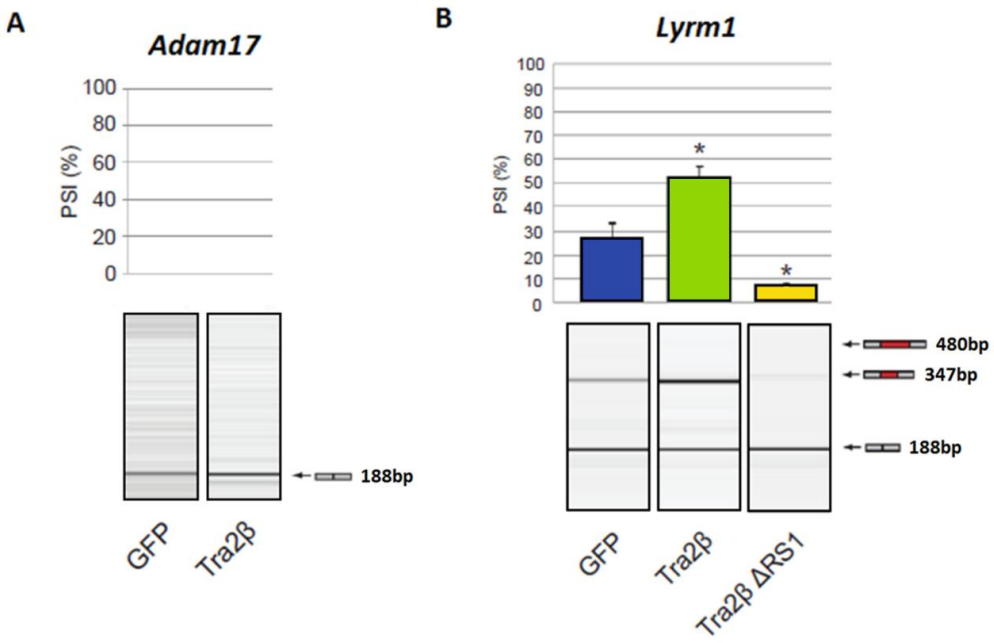


Figure 2.12 Splicing analyses of the *Adam17* and *Lyrm1* minigenes. Tra2β CLIP tags mapped within close proximity to alternative 3' splice sites of cassette exons with the *Adam17* and *Lyrm1* genes. (A) The *Adam17* minigene exon was not included (0% splicing inclusion) when transfected into HEK-293 cells and did not respond to co-expression with full-length Tra2β-GFP. (B) Splicing inclusion of the *Lyrm1* cassette significantly increased after co-transfection with full-length Tra2β-GFP and was significantly respressed by co-transfection with Tra2βΔRS1-GFP. I did not detect any isoform in which the alternative 3' splice site was selected from this minigene (predicted size 480bp). Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001.

2.4.4 Investigating splicing regulation of breast cancer-associated alternative splicing events by Tra2β using minigenes

In sections 2.4.1 to 2.4.3, I investigated splicing regulation of Tra2β target exons identified from the mouse testis. Following on from that work, I next investigated the possibility of using the consensus Tra2β binding site (derived from the Tra2β HITS-CLIP experiment and used to map Tra2β binding sites within the *Nasp-T* exon) to predict possible targets of splicing regulation from a publicly available dataset of cancer-associated alternative splicing (Venables et al., 2009a). The data from Venables et al. was generated using a high-throughput RT-PCR screen of alternative splicing events in human breast and ovarian tumours (Venables et al., 2009a). My overall aim was to investigate whether Tra2β was involved in splicing regulation of a selection of human breast cancer-associated alternative splicing events using minigenes.

Accordingly, the top 25 K-mers from the Tra2β HITS-CLIP experiment were mapped to the top 50 up-regulated alternative splicing events identified from the panel of breast cancer tumours (Venables et al., 2009a). Six alternative exons were identified within *PDLIM5, MCAT, SLC10A7, SYTL2, CLK1, SDHC* which contained multiple predicted Tra2β binding sites and these were subsequently cloned into the pXJ41 minigene vector, together with approximately 300bp of flanking intronic sequences for splicing analysis. The six minigenes were then transfected into HEK-293 cells for splicing analysis. Each minigene was initially co-tranfected with full-length Tra2β-GFP or the GFP-only control to determine whether each minigene was responsive to full-length Tra2β protein expression (Figure 2.13).

From an initial screen using single biological samples, two of the six minigenes visibly responded to co-transfection with full-length Tra2β-GFP. The Tra2β-responsive minigenes contained an exon with an alternative 3' splice site from *PDLIM5* (Figure 2.13 A) and a cassette exon from *CLK1* (Figure 2.13 E). Unexpectedly, co-transfection with full-length Tra2β-GFP actively repressed splicing inclusion of both splicing events in the *PDLIM5* and *CLK1* minigenes. The remaining four minigenes containing exons from *MCAT* (Figure 2.13 B), *SLC10A7* (Figure 2.13 C), *SYTL2* (Figure 2.13 D) and *SDHC* (Figure 2.13 F) did not visibly respond to co-transfection with full-length Tra2β-GFP when compared to the GFP-only control.

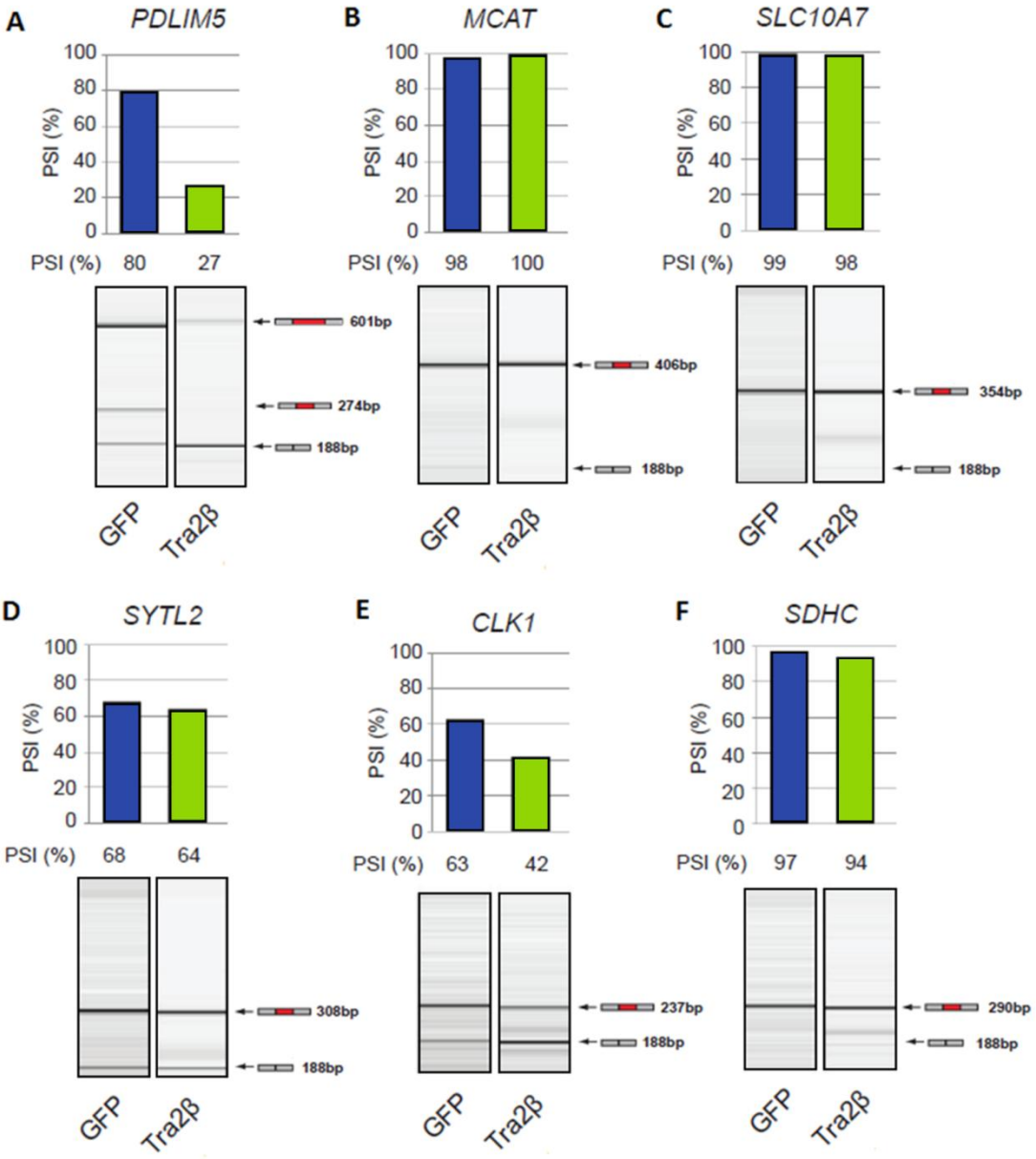


Figure 2.13 Splicing analyses of six minigenes containing breast cancer-associated alternative splicing events. Each minigene was co-transfected with either a GFP-only control or full-length Tra2β-GFP. This initial screen used data from single biological samples. Splicing inclusion of the *PDLIM5* (A) and *CLK1* (E) exons was visibly reduced following co-transfection with full-length Tra2β-GFP compared to the GFP-only control.

I repeated the *PDLIM5* and *CLK1* minigene transfections in biological triplicate to confirm that the observed changes in splicing were reproducible and statistically significant (Figure 2.14). Co-transfection of the *CLK1* minigene with either full-length Tra2β-GFP or Tra2βΔRS1-GFP significantly repressed exon inclusion when compared to the GFP-only control (Figure 2.14 A). Co-transfection of full-length Tra2β-GFP significantly repressed inclusion of both versions of the *PDLIM5* exon (a larger exon is

created through use of an alternative 3' splice site), whilst Tra2βΔRS1-GFP had no significant effect on inclusion of either size exon (Figure 2.14 B).

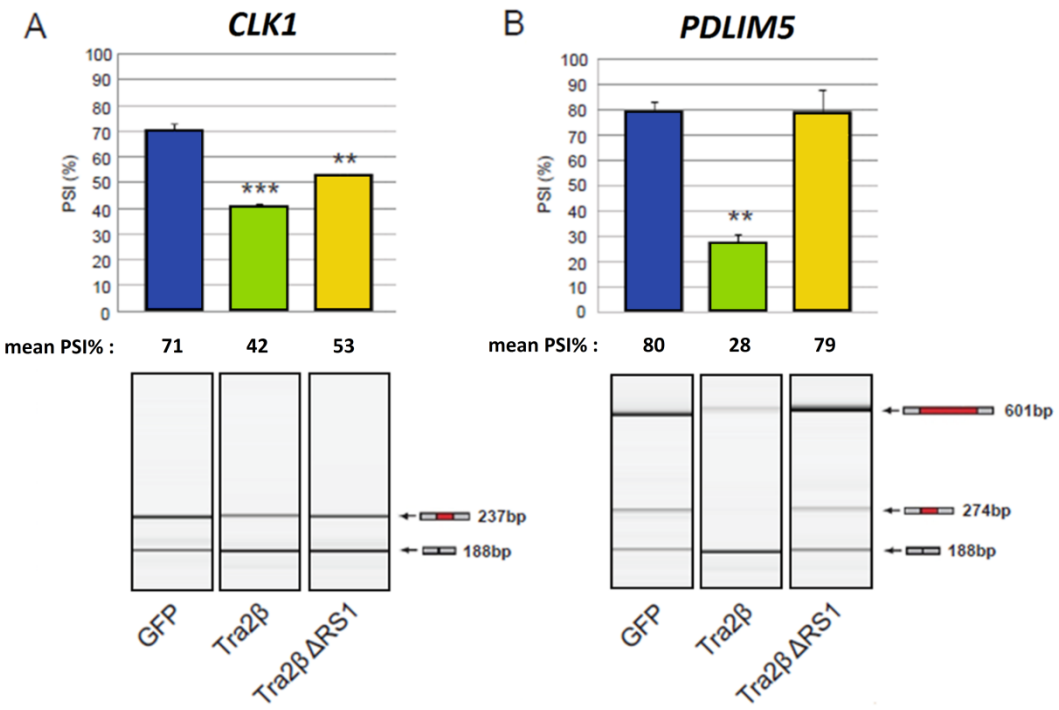


Figure 2.14 Splicing analyses of the *CLK1* and *PDLIM5* minigenes. Co-transfection with full-length Tra2β-GFP significantly repressed exon inclusion in the *CLK1* (A) and *PDLIM5* (B) minigenes compared to the GFP-only control. Co-transfection with Tra2βΔRS1-GFP also significantly reduced splicing inclusion of the *CLK1* exon, but not in the *PDLIM5* minigene. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001.

2.4.5 Tra2β protein depletion via siRNA transfection

To investigate whether the corresponding endogenous exons from *LYRM1*, *PDLIM5* and *CLK1* were regulated by Tra2β, I knocked down endogenous Tra2β protein expression in the invasive breast cancer cell line MDA-MB-231 using siRNA transfection. Knockdown of endogenous Tra2β protein was confirmed by Western Immunoblotting, using a rabbit polyclonal antibody against Tra2β and a mouse monoclonal antibody against α-Tubulin as a loading control. Optimal Tra2β protein depletion was observed 72 hours after siRNA transfection. Tra2β protein depletion was observed in cells transfected with TRA2B siRNA when compared to untreated control cells, negative control siRNA transfected cells or TRA2A siRNA transfected cells (Figure 2.15).

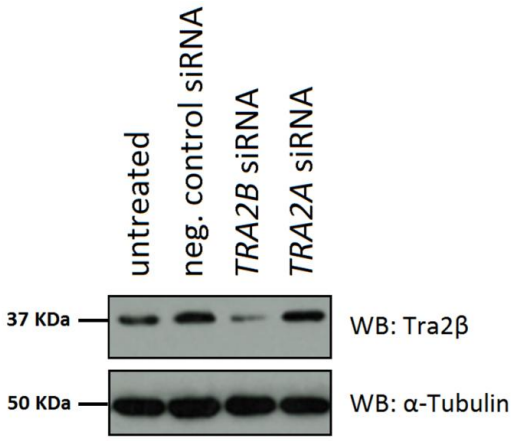


Figure 2.15 Depletion of endogenous Tra2β was confirmed in MDA-MB-231 cells by Western Immunoblotting. Expression of α-Tubulin was monitored as a loading control.

2.4.6 Splicing analysis of endogenous target exons following Tra2β depletion

To determine whether depletion of endogenous Tra2β would affect splicing of the endogenous target exons, MDA-MB-231 cells were harvested 72 hours after siRNA transfection. Splicing patterns of the endogenous RNA targets was analysed by RT-PCR and capillary gel electrophoresis, using primers located within flanking constitutive exons. Unexpectedly, the splicing inclusion of all three endogenous exons from *PDLIM5* (Figure 2.16 A), *CLK1* (Figure 2.16 B) and *LYRM1* (Figure 2.16 C) was unaffected by Tra2β depletion when compared to untreated control cells or negative control siRNA transfected cells. As a positive control, I analysed splicing inclusion of the human homologs of two target exons identified earlier in the chapter; the *NASP-T* exon (Figure 2.16 D) and the *TRA2A* poison exon (Figure 2.16 E). Although this initial RT-PCR was performed using single biological replicates, inclusion of the endogenous *NASP-T* exon and *TRA2A* poison exon was visibly reduced in cells transfected with *TRA2B* siRNA when compared to untreated control cells or negative control siRNA transfected cells. This suggested that depletion of endogenous Tra2β had at least been sufficient to induce splicing changes to these two target exons. It also showed that Tra2β-mediated splicing regulation of the *NASP-T* exon and *TRA2A* poison exon was conserved between mice and humans.

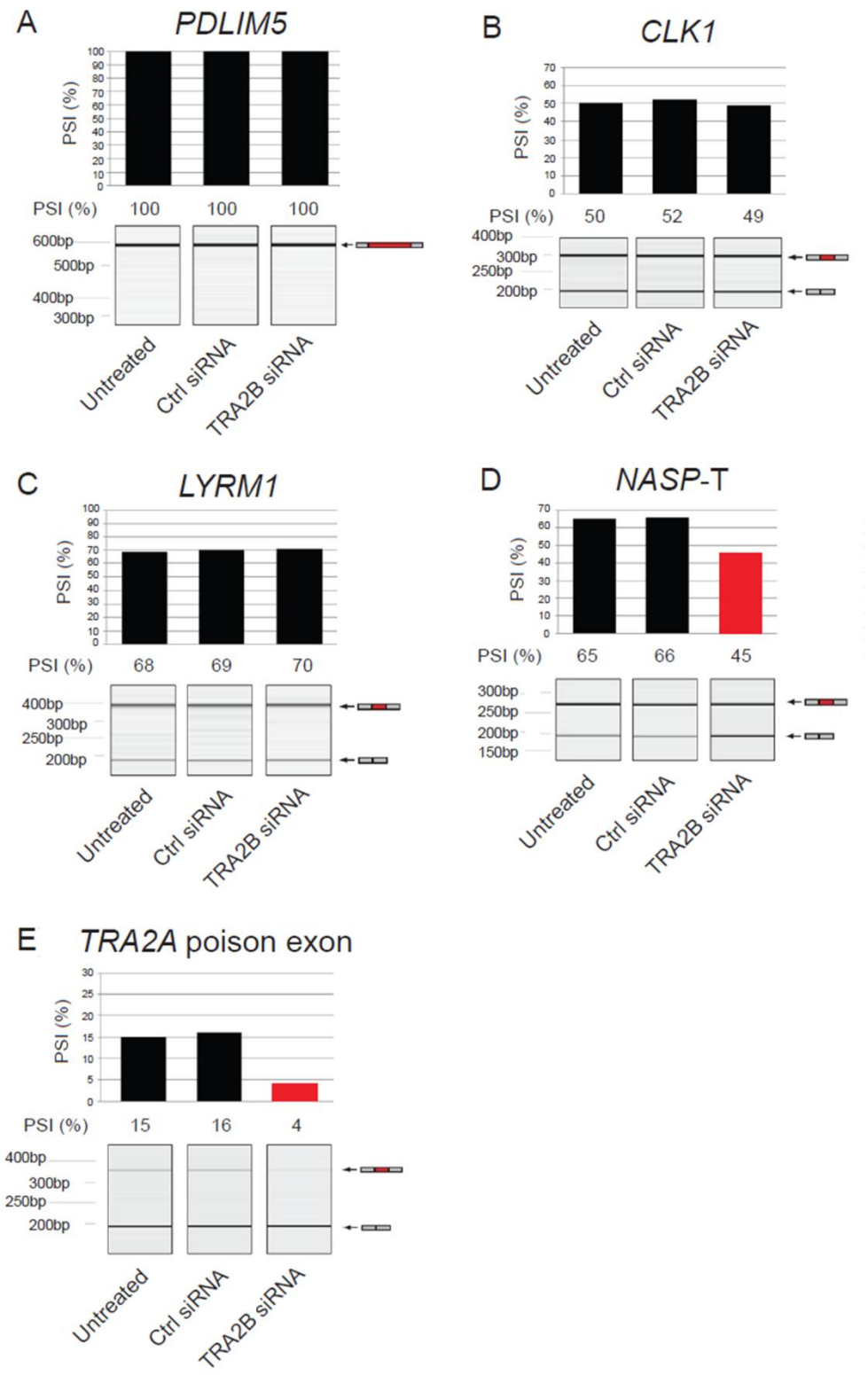


Figure 2.16 Splicing analyses of endogenous target exons following depletion of Tra2β in MDA-MB-231 cells. The PSI of endogenous exons within (A) *PDLIM5*, (B) *CLK1*, (C) *LYRM1*, (D) *NASP* and (E) *TRA2A* was determined by RT-PCR and capillary gel electrophoresis. Knockdown of Tra2β had no effect on splicing inclusion of the endogenous exons within *PDLIM5*, *CLK1* and *LYRM1* when compared with untreated control cells or cells transfected with negative control siRNA. However, knockdown of Tra2β did reduce inclusion of the endogenous *NASP-T* and *TRA2A* poison exons. Data shown is from single biological samples only.

2.4.7 Splicing analysis of endogenous target exons after joint depletion of Tra2α and Tra2β

Subsequent experiments (discussed in chapter 3) indicated that joint depletion of both Tra2α and Tra2β proteins was required to observe robust changes in splicing of Tra2β target exons, due to paralog compensation between Tra2α and Tra2β. Therefore in a much later experiment, I monitored splicing of the endogenous *CLK1*, *LYRM1* and *PDLIM5* exons following joint depletion of both Tra2α and Tra2β proteins in MDA-MB-231 cells. Consistent with this hypothesis, joint depletion of both Tra2 proteins (but not depletion of Tra2β alone) led to a significant reduction in splicing inclusion of the *CLK1* exon (Figure 2.17 A) and the *LYRM1* exon (Figure 2.17 B). Joint depletion of Tra2α and Tra2β had no significant effect on inclusion of the *PDLIM5* exon.

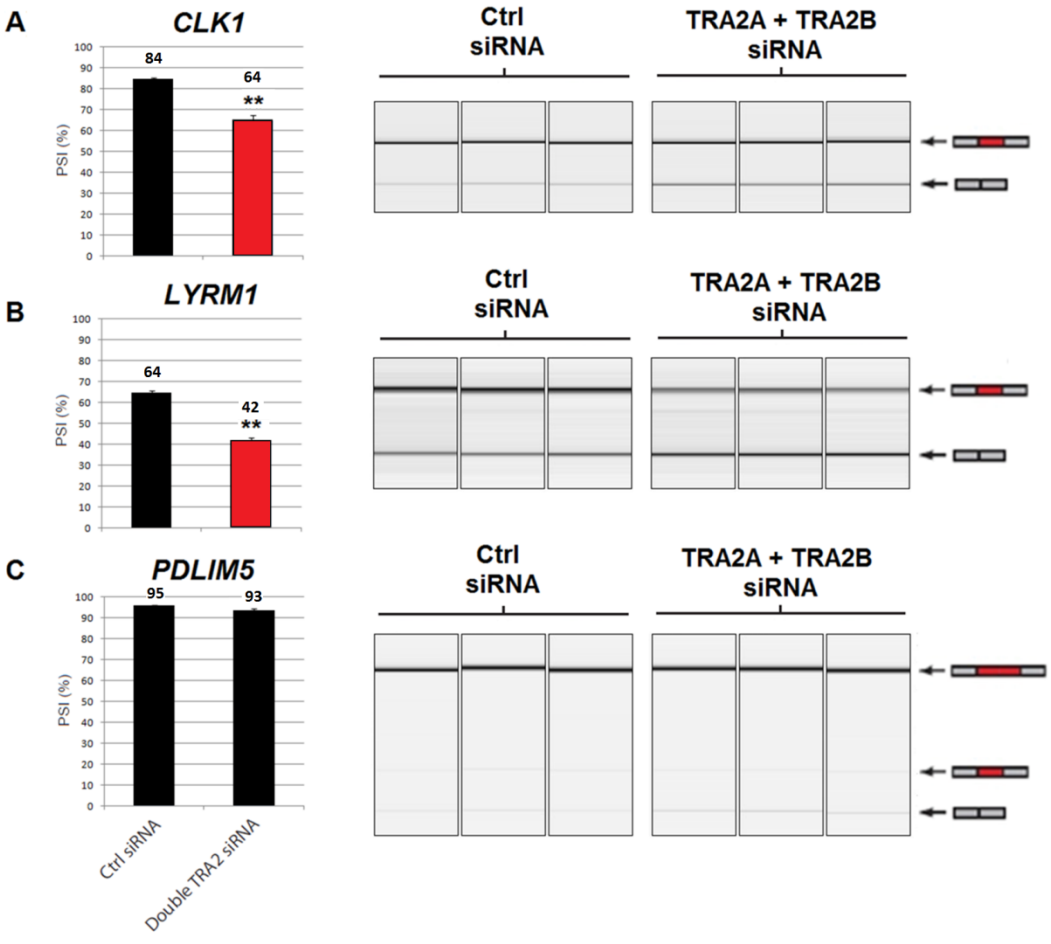


Figure 2.17 Splicing analyses of endogenous target exons after joint depletion of Tra2α and Tra2β proteins. Splicing inclusion of the endogenous *CLK1* exon (A) and *LYRM1* exon (B) was significantly reduced following depletion of both Tra2α and Tra2β proteins, but not depletion of Tra2β alone. Inclusion of the endogenous *PDLIM5* exon was not affected by joint Tra2 protein depletion. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001.

2.5 Discussion

2.5.1 Validation of novel Tra2β target exons from the mouse testis

To further understand the biological functions of Tra2β, a key priority was to identify novel RNA targets which were physiologically regulated by this RNA-binding protein. New technologies including UV cross-linking and immunoprecipitation (CLIP) of RNA-binding proteins coupled to deep sequencing are providing new insights into the targets of many RNA-binding proteins on a transcriptome-wide scale, but these techniques require functional validation. The murine Tra2β HITS-CLIP experiment carried out previously provided valuable information about Tra2β-RNA interactions in the mouse testis, which I have used in this chapter to functionally validate Tra2β-mediated splicing regulation using minigenes.

To determine whether Tra2β directly regulated splicing of cassette exons identified from the Tra2β HITS-CLIP experiment, a panel of candidate exons were cloned into the pXJ41 minigene vector for splicing analysis. The minigene assays identified a number of novel exons which responded to co-transfection with full-length Tra2β *in cellulo,* including exons with important functions from *Nasp* and *Tra2a*. Other novel Tra2β-responsive exons were identified from *Creb*, *Krba1*, *Lin28b*, *Fabp9* and *Lyrm1*.

One advantage of using minigene models is that it readily permits manipulation of the target sequence. This allows binding sites to be moved or mutated to provide key functional information about protein-RNA interactions. On the other hand, the use of minigenes significantly limited the number of targets which could be analysed. Ultimately, just 10 exons were screened from a transcriptome-wide study of Tra2β binding. Subsequent studies have since analysed Tra2β splicing regulation on a more global scale, using strategies including splicing-sensitive microarrays after digitoxin treatment (digitoxin is a negative regulator of Tra2β protein expression) (Anderson et al., 2012) or using RNA from a neuronal-specific *tra2b* knockout mouse model (Storbeck et al., 2014). RIP-seq has also been used to investigate transcriptome-wide binding sites of Tra2β (Uren et al., 2012).

2.5.1.1 The *Tra2a* posion exon

From the initial panel of minigenes, the most highly responsive exon to Tra2β co-expression was a poison exon from *Tra2a*. Poison exons introduce premature termination codons (PTC) into mRNA transcripts, thereby targeting the mRNA for nonsense-mediated decay (NMD) (Lareau et al., 2007). Poison exons therefore provide a mechanism for post-transcriptional regulation of gene expression via alternative splicing (Ni et al., 2007). Tra2β was previously shown to auto-regulate its own expression by promoting inclusion of a poison exon within the *Tra2b* mRNA (Stoilov et al., 2004). This new data suggested that Tra2β may also directly regulate expression of Tra2α, by promoting inclusion of a posion exon within the *Tra2a* mRNA. Extensive cross-regulation through non-productive splicing has previously been described for other RNA-binding proteins. This includes cross-regulation between members of the SR-protein family, as well as between the Polypyrimidine Tract Binding protein (PTB) and its paralog nPTB (Wollerton *et al.*, 2004; Spellman *et al.*, 2007). Cross-regulation between Tra2α and Tra2β is investigated further in chapter 3.

2.5.1.2 The *Nasp-T* exon

Tra2β was found to directly regulate a large exon from *Nasp* which contained multiple Tra2β binding sites. *Nasp* encodes a histone chaperone protein required for the nuclear import of histones during the G1-S phase transition in the cell cycle (Richardson et al., 2006). The Nasp protein occurs in two major isoforms; a large, testis-enriched 'tNASP' isoform which is expressed in germ, embryonic and transformed cells and a smaller, 'sNASP' isoform expressed in all rapidly dividing somatic cells (Alekseev OM, 2011). The tNASP isoform is produced by inclusion of the 975bp cassette exon we identified as directly regulated by Tra2β. Increased expression of the tNASP isoform correlates with an increase in Tra2β protein expression during male germ cell development, suggesting Tra2β may regulate the ratio of Nasp protein isoforms during the development of meiotic cells (Welch and O'Rand, 1990; Alekseev *et al.*, 2009). The tNASP isoform is induced in transformed cells and specific ablation of the tNASP isoform in prostate cancer cells was found to inhibit proliferation and induce apoptosis, suggesting induction of the tNASP isoform may play a role in tumour progression (Alekseev OM, 2011).

Sequence analysis of the *Nasp-T* exon revealed a particularly high density of Tra2β binding sites throughout the exon. Site-directed mutagenesis uncovered a redundancy between which Tra2β binding sites were used to activate splicing, as mutated versions of the *Nasp-T* exon were still efficiently activated by co-expression with Tra2β. The redundancy of Tra2β binding sites within the *Nasp-T* exon is most consistent with the kinetic probability model of splicing regulation proposed by (Grellscheid *et al.*, 2011b). In the kinetic probability model, the higher the density of Tra2β binding sites within an exon, the higher the probability of activation at any given nuclear concentration of Tra2β, yet each individual binding site is non-essential for splicing activation. Reducing the number of Tra2β binding sites within the *Nasp*-T via mutagenesis decreased exon activation at endogenous concentrations of Tra2β, but inclusion of the exon could still be activated by over-expressing Tra2β. The number of Tra2β binding sites within an exon such as *Nasp-T* may therefore establish the sensitivity of the exon to endogenous concentrations of Tra2β. However, mutation of different pairs of Tra2β binding sites did not exert an exactly equal effect, highlighting an underlying positional importance of the Tra2β binding sites, distinct from the total number of sites within the exon. Inclusion of both the *Nasp-T* exon and *Tra2a* poison exon has subsequently been shown to be significantly reduced in a neuronal-specific *Tra2b* knockout mouse model, confirming the *Tra2a* poison exon and *Nasp-T* exon as physiological targets *in vivo* (Grellscheid et al., 2011a).

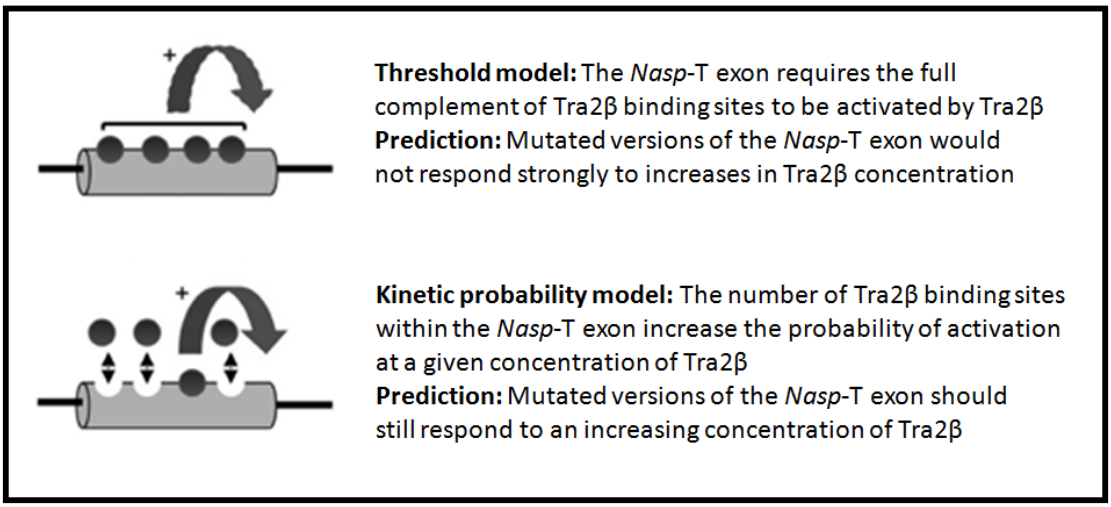


Figure 2.18 Two models of Tra2β mediated splicing activation. Splicing regulation of the *Nasp-T* exon was most consistent with the kinetic probability model of splicing regulation proposed by (Grellscheid *et al.*, 2011b). This figure is adapted from (Grellscheid et al., 2011b).

2.5.2 Co-expression of different Tra2β expression constructs can activate, co-activate or repress inclusion of the same target exons

To determine whether direct RNA-binding was required for Tra2β-mediated splicing activation of target exons, inclusion of the Tra2β-responsive minigenes was compared between cells co-transfected with full-length Tra2β and the Tra2βΔRRM isoform. The Tra2βΔRRM isoform was unable to activate splicing of all Tra2β-responsive minigenes except for *Nasp-T*, suggesting direct RNA interaction is required for activation of these exons. Mutation of Tra2β binding sites within the *Nasp-T* minigene reduced exon inclusion, indicating that Tra2β-mediated splicing regulation of this exon was direct. However, co-transfection of Tra2βΔRRM also significantly increased exon inclusion in the *Nasp-T* minigene (though to a lesser extent than full-length Tra2β), suggesting Tra2β may also be capable of activating splicing of some exons through an indirect mechanism. One possibility is that Tra2βΔRRM may interact with endogenous Tra2β or other RNA-binding proteins to facilitate splicing complex assembly, independent from direct RNA-binding. Tra2β may therefore have additional functions as a splicing co-activator for some exons.

Co-transfection with the Tra2βΔRS1 isoform not only failed to activate splicing, but surprisingly actively repressed splicing inclusion of nearly all Tra2β-responsive minigenes tested. The Tra2βΔRS1 isoform did not inhibit splicing of any exons which did not respond to Tra2β, suggesting Tra2βΔRS1 acts as a specific inhibitor of Tra2β-responsive exons, rather than as a general inhibitor of splicing. One possible model for this is that the Tra2βΔRS1 isoform acts as competitive inhibitor of full-length Tra2β (Figure 2.19). In a competitive inhibition model, the Tra2βΔRS1 isoform would occupy Tra2β binding sites within an exon (preventing full-length Tra2β from binding), but fail to facilitate splicing complex assembly through RS1 domain interactions. As a Tra2β-3 isoform (lacking the RS1 domain) is expressed endogenously, it is possible that different endogenous Tra2β protein isoforms may also act competitively to either activate or inhibit inclusion of the same target exons *in vivo.* The endogenous Tra2β-3 isoform has been detected in protein lysates from the brain and testis using an antibody specific to the C-terminus of the Tra2β protein (Stoilov et al., 2004). However, the Tra2β antibody used in my study is specific for the N-terminus of the Tra2β protein and therefore I did not detect the endogenous Tra2β-3 protein isoform in my study.

The use of a C-terminal specific primary antibody against Tra2β would facilitate the study of the endogenous Tra2β-3 protein isoform in future studies.

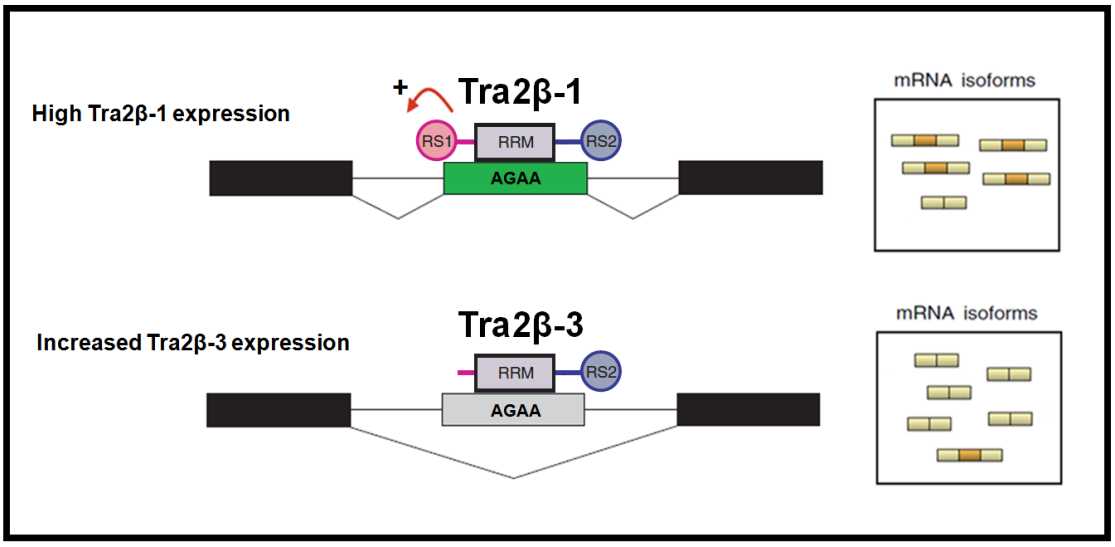


Figure 2.19 A competitive inhibition model of target exon regulation by Tra2β-1 and Tra2β-3. Full-length Tra2β (Tra2β-1) and Tra2βΔRS1 (Tra2β-3) compete for the same binding sites within an exon. Tra2β-1 is able to enhance spliceosome assembly through RS1-domain interactions. When Tra2β-3 occupies the same binding site, Tra2β-1 is unable to bind to the exon. However as Tra2β-3 lacks the functional RS1-domain, Tra2β-3 fails to enhance spliceosome assembly, leading to skipping of this exon.

2.5.3 Investigating regulation of alternative 3' splice site selection

To investigate whether Tra2β was involved in selection of alternative 3' splice sites, two exons with 3' alternative splice sites from *Adam17* and *Lyrm1* were cloned into pXJ41 minigenes for splicing analysis. Inclusion of the *Adam17* minigene exon was not detected when co-transfected with either GFP alone or full-length Tra2β-GFP. Inclusion of the *Lyrm1* minigene exon was positively regulated by full-length Tra2β-GFP, though use of the alternative 3' splice site was not detected. As use of the alternative 3' splice site was not detected for either minigene, I decided to focus on other aspects of Tra2β mediated splicing regulation for further investigation.

2.5.4 Investigating regulation of breast cancer-associated alternative splice events by Tra2β

To investigate whether Tra2β regulated a panel of breast cancer-associated alternative splice events, six alternative exons that contained multiple Tra2β binding sites were cloned into the pXJ41 minigene for splicing analysis. Surprisingly, co-transfection with

Tra2β was found to inhibit inclusion of a cassette exon from *CLK1*. *CLK1* encodes a dual specificity protein kinase which phosphorylates serine-arginine rich proteins involved in pre-mRNA processing (Colwill et al., 1996). Interestingly, a functionally related member of the CLK protein family; CLK2, has previously been reported to directly phosphorylate Tra2β (Stoilov et al., 2004). CLK2 mediated hyperphosphylation of Tra2β strongly reduces the ability of Tra2β to directly bind RNA (Stoilov et al., 2004). However, it is not known whether CLK1 also phosphorylates Tra2β. Interestingly, inclusion of the *CLK1* cassette exon was the most significantly altered splice event in the RT-PCR screen of breast cancer-associated splicing changes (Venables et al., 2009b), suggesting it may have clinical significance.

Another Tra2β-responsive alterative splicing event was an exon with an alternative 3' splice site from *PDLIM5,* which encodes a PDZ-containing scaffold protein associated with cardiomyocyte and neuronal growth (Krcmery et al., 2010). Tra2β was also found to repress inclusion of both versions of the exon (created using different 3' splice sites) in the *PDLIM5* minigene. Though Tra2β is most frequently reported to promote exon inclusion, Anderson et al. have since reported multiple exons shown to be repressed by Tra2β (2012).

2.5.5 Splicing regulation of endogenous target exons

To investigate splicing regulation of the corresponding endogenous exons, endogenous Tra2β was depleted in the invasive breast cancer cell line MDA-MB-231 using RNAi. Suprisingly, there was no significant difference in splicing inclusion of endogenous exons from *PDLIM5*, *LYRM1* and *CLK1* genes following depletion of Tra2β compared to negative control siRNA treated cells. However, inclusion of the endogenous *NASP-T* and *TRA2A* poison exons was visibly reduced after Tra2β depletion, confirming that these exons were endogenous targets and that splicing regulation of these exons is conserved between mice and humans. As the endogenous exons from *PDLIM5*, *LYRM1* and *CLK1* did not respond to depletion of Tra2β, they were not pursued further.

Subsequent experiments later in my study revealed that due to paralog compensation between Tra2α and Tra2β (discussed in chapter 3), joint depletion of both Tra2α and Tra2β proteins was required to observe robust splicing changes to most Tra2β targets exons. Inclusion of the endogenous exons within *CLK1* and *LYRM1* was significantly

reduced following joint depletion of both Tra2α and Tra2β (but not single depletion of Tra2β), whilst inclusion of endogenous *PDLIM5* exon remained unaffected. Over-expression of full-length Tra2β promoted inclusion of the *LYRM1* minigene exon in HEK-293 cells, whilst knockdown of Tra2β reduced inclusion of endogenous *LYRM1* exon in MDA-MB-231 cells. This data is consistent with full-length Tra2β actively promoting splicing inclusion of the *LYRM1* exon. The LYR motif containing 1 gene (*LYRM1*) encodes a nucleoprotein which is highly expressed in adipose and heart tissue (Zhu et al., 2010). The LYRM1 protein has been implicated both in regulation of cell growth during heart development (Zhu et al., 2010) and insulin-resistance in adipocytes (Qin et al., 2012).

Over-expression of full-length Tra2β in HEK-293 cells repressed inclusion of the *CLK1* minigene exon. However joint Tra2 protein depletion was also found to reduce *CLK1* exon inclusion in MDA-MB-231 cells. This apparent discrepancy regarding *CLK1* splicing regulation may be due to a number of underlying factors. For example, over-expression of Tra2β in the minigene model may inhibit the function of other RNA-binding proteins which positively regulate this exon. Cell specific differences in the expression of RNA-binding proteins between HEK-293 cells and MDA-MB-231 cells may also influence regulation of the *CLK1* exon. Other factors, such as the limited size of the intronic sequence cloned into the *CLK1* minigene could also account for differences in splicing regulation between the minigene exon and the endogenous exon.

2.5.6 Chapter Summary

In this chapter, I have validated splicing regulation of a panel of novel Tra2β target exons, including functionally important exons within *Nasp* and *Tra2a.* Tra2β was found to directly bind to the *Nasp-T* exon and promote splicing inclusion through multiple redundant binding sites. The *Tra2a* poison exon is studied in further detail in chapter 3.

The majority of target exons required direct binding for splicing activation by Tra2β, though Tra2β may also activate splicing inclusion of some exons such as the *Nasp-T* exon indirectly. The RS1-domain of the Tra2β protein was found to be essential for splicing activation for all Tra2β-responsive exons tested. Furthermore, the Tra2βΔRS1 isoform was found to efficiently repress inclusion of a number of exons normally activated by full-length Tra2β. Intruigingly, this suggests that the endogenous Tra2β-1 and Tra2β-3 isoforms could also be functionally distinct in splicing activity.

Chapter 3: Transcriptome-wide identification of Tra2β-RNA interactions using iCLIP

3.1 Introduction

Tra2β protein expression was previously found to be significantly up-regulated in invasive breast cancer tumours (Watermann et al., 2006) and we hypothesised that Tra2β may regulate alternative splicing of functionally important genes in breast cancer. Therefore in this chapter, my overall aim was to identify and characterise direct RNA targets of Tra2β in the human invasive breast cancer cell line MDA-MB-231.

In chapter 2, I used data from a previous Tra2β HITS-CLIP experiment to validate splicing targets of Tra2β from the mouse testis. Despite the successful identification of several target exons, the previous HITS-CLIP experiment generated a library of limited complexity (in total 177,000 reads mapped back to the mouse genome) (Grellscheid et al., 2011a). Primer extension assays have shown that cross-linked RNAs frequently cause reverse transcription reactions to stall, resulting in truncated cDNAs which were completely lost during the previous HITS-CLIP protocol (Urlaub *et al.*, 2002). To overcome this problem, Konig et al. recently developed an improved CLIP protocol termed iCLIP, which captures the truncated cDNAs to enable identification of the site of protein-RNA cross-linking at near individual nucleotide resolution (Konig *et al.*, 2011). iCLIP has been used successfully to identify targets for several RNA-binding proteins, including hnRNP C (Konig *et al.*, 2010), TIA1 and TIAL1 (Wang *et al.*, 2010) and SRSF3 and SRSF4 (Anko *et al.*, 2012). In this chapter, I use iCLIP to identify direct RNA targets of Tra2β in MDA-MB-231 cells, and subsequently investigate splicing regulation of these newly identified targets following Tra2 protein depletion using RNAi.

In chapter 2, we also identified a poison exon from *Tra2a* which was regulated by both Tra2α and Tra2β proteins. Tra2β was also previously found to auto-regulate its own protein expression via regulation of a posion exon within the *Tra2b* pre-mRNA (Stoilov et al., 2004). These data suggested that vertebrate Tra2 proteins may cross-regulate their protein expression. Therefore in this chapter, I also further examine whether Tra2α and Tra2β cross-regulate their protein expression and investigate their functional relationship in splicing regulation.

3.2 Aims

The aims of this chapter were to:

1. Investigate whether Tra2α and Tra2β cross-regulate their protein expression
2. Identify direct RNA targets of Tra2β in MDA-MB-231 cells using individual nucleotide-resolution CLIP (iCLIP)
3. Validate splicing regulation of newly identified targets using siRNA-mediated depletion of the Tra2 proteins in MDA-MB-231 cells
4. Investigate whether Tra2α protein expression can functionally compensate for loss of Tra2β expression in splicing regulation

3.3 Materials and Methods

3.3.1 siRNA transfection

In this chapter, siRNA transfections were used to deplete endogenous Tra2α, Tra2β or to jointly deplete Tra2α and Tra2β in MDA-MB-231 cells. The siRNA transfection protocol used in this chapter is described previously in chapter 2, methods section 2.3.7.

3.3.2 Western Immunoblotting

Endogenous proteins were detected by Western Immunoblotting using primary antibodies against Tra2α (Novus Biologicals, H00029896-B01P) (1:500 dilution), Tra2β (Abcam, ab31353) (1:2000 dilution) and α-Tubulin (Sigma-Aldrich, T5168) (1:2000 dilution). The standard protocol used for Western Immunoblotting is described in chapter 2, methods section 2.3.3.

3.3.3 Quantitative real-time PCR (qPCR)

Relative gene expression was determined by quantitative real-time PCR (qPCR) using a SYBR Green PCR Master Mix kit (Applied Biosystems) and an Applied Biosystems 7900HT Fast Real-Time PCR Machine. qPCR was used in this chapter to analysis *TRA2A* and *TRA2B* mRNA expression, as well as the expression of 10 variable exons and control constitutive exons within the *CD44* gene. RNA samples were prepared from MDA-MB-231 cells 72 hours after siRNA transfection using an RNeasy Mini Kit (Qiagen) following the manufacturer's instructions and treated with DNase using a DNA-free kit (Ambion) following the manufacturer's instructions. cDNA was generated from equal quantites (500ng) of total RNA for each sample using a SuperScript VILO cDNA synthesis kit (Invitrogen) following the manufacturer's instructions. The cDNA was diluted 1 in 20 using RNase-free water (Ambion) and 1µl of the diluted cDNA was used per reaction. All qPCR experiments were performed using a minimum of 3 biological samples, with a minimum of 3 technical replicates per biological sample. For each reaction, 9µl of PCR master mix was added per well to a 384-well qPCR plate containing 1µl diluted cDNA per well and sealed with a clear adhesive film. A 'no template control' (NTC) reaction was used for each master mix. As the primers used in the *CD44* experiment did not span exon junctions, a –RT control was used for each

RNA sample. An example of the standard 1X 9µl qPCR master mix is provided in Table 3.1. The thermocycling conditions used for qPCR were as follows: 50°C for 2 minutes, 95°C for 10 minutes, then 40 cycles of [95°C for 15 seconds and 60°C for 60 seconds]. Gene expression was calculated relative to three housekeeping genes *ACTB*, *GAPDH* and *TUBB*. Ct values for each sample were calculated using SDS 2.4 software (Applied Biosystems) and relative mRNA expression was calculated using the $2^{-\Delta\Delta Ct}$ method. All primers used for quantitative real-time PCR (qPCR) in chapter 3 are provided in Table 3.2.

1X SYBR Green Master Mix

2X SYBR Green PCR Master Mix	5µl
Forward primer (10µM)	1µl
Reverse primer (10µM)	1µl
RNase-free water	2µl
TOTAL	9µl

Table 3.1 Standard 1X SYBR Green Master Mix used in quantitative real-time PCR (qPCR).

Primer	Sequence
TRA2A For	TCCAATGTCTAACCGGAGAAG
TRA2A Rev	CCAAACACTCCAAGGCAAGT
TRA2B For	CGCCAACACCAGGAATTTAC
TRA2A Rev	TCATAGCCCCGATCATATCC
ACTB For (reference gene 1)	CATCGAGCACGGCATCGTCA
ACTB Rev (reference gene 1)	TAGCACAGCCTGGATAGCAAC
GAPDH For (reference gene 2)	AACAGCGACACCCATCCTC
GAPDH Rev (reference gene 2)	CATACCAGGAAATGAGCTTGACAA
TUBB For (reference gene 3)	CTTCGGCCAGATCTTCAGAC
TUBB Rev (reference gene 3)	AGAGAGTGGGTCAGCTGGAA
CD44 constitutive exon 2 For	CTACAGCATCTCTCGGACGG
CD44 constitutive exon 2 Rev	GTCTCAAATCCGATGCTCAGAG
CD44 constitutive exon 5 For	GAAAGGAGCAGCACTTCAGG
CD44 constitutive exon 5 Rev	ACTGTCTTCGTCTGGGATGG
CD44 variable exons 1&2 For	TGGGATTGGTTTTCATGGTT
CD44 variable exons 1&2 Rev	CAGCCATTTGTGTTGTTGTG
CD44 variable exon 3 For	CGTCTTCAAATACCATCTCAGC
CD44 variable exon 3 Rev	TCATCATCAATGCCTGATCC
CD44 variable exon 4 For	ACGGGCTTTTGACCACAC
CD44 variable exon 4 Rev	AGCACTTCCGGATTTGAATG
CD44 variable exon 5 For	AATGGCACCACTGCTTATGA
CD44 variable exon 5 Rev	TGTGGGGTCTCTTCTTCCTC
CD44 variable exon 6 For	AGGAACAGTGGTTTGGCAAC
CD44 variable exon 6 Rev	CGAATGGGAGTCTTCTTTGG
CD44 variable exon 7 For	CAGCCTCAGCTCATACCAG
CD44 variable exon 7 Rev	CCATCCTTCTTCCTGCTTG
CD44 variable exon 8 For	TGGACTCCAGTCATAGTATAACGC
CD44 variable exon 8 Rev	GCGTTGTCATTGAAAGAGGTCCTG
CD44 variable exon 9 For	AGCAGAGTAATTCTCAGAGC
CD44 variable exon 9 Rev	TGCTTGATGTCAGAGTAGAAGTTG
CD44 variable exon 10 For	ATAGGAATGATGTCACAGGTGG
CD44 variable exon 10 Rev	CGATTGACATTAGAGTTGGAATCTCC
CD44 standard isoform specific For	CCTCCAGTGAAAGGAGCAGCAC
CD44 standard isoform specific Rev	GTGTCTTGGTCTCTGGTAGCAGGGAT

Table 3.2 Primers used for quantitative real-time PCR (qPCR) in chapter 3.

3.3.4 Splicing analysis of endogenous RNA targets

The percentage splicing inclusion (PSI) of endogenous target exons was measured by conventional RT-PCR and capillary gel electrophoresis. RNA was extracted from cells 72

hours after siRNA transfection using the standard Trizol RNA extraction protocol described earlier in chapter 2, methods section 2.3.2.2. The cDNA was synthesised from 500ng total RNA in a 10µl reaction, using a SuperScript VILO cDNA synthesis kit (Life Technologies) following the manufacturer's instructions. The splicing profile of endogenous targets was monitored by PCR using primers designed within constitutive flanking exons. The primers used for the endogenous splicing assays performed in this chapter are provided in Appendix A. For each PCR, 1µl of diluted cDNA (1 in 8 dilution) was used as the template in a 10µl PCR reaction using Phusion High-Fidelity PCR Kit (NEB UK) following the manufacturer's instructions. A standard 10µl PCR reaction master mix is shown in chapter 2, Table 2.2. Splicing profiles were monitored and quantified using the QIAxcel capillary gel electrophoresis system (Qiagen) and the percentage splicing inclusion (PSI) of endogenous targets were calculated as described in chapter 2, methods section 2.3.2.6.

3.3.5 Individual nucleotide resolution CLIP (iCLIP)

Direct RNA targets of Tra2β in MDA-MB-231 cells were identified using a modified UV-cross-linking and immunoprecipitation technique, developed to capture protein-RNA interactions at individual nucleotide resolution (iCLIP). These experiments were based on the published iCLIP protocol by Konig et al. and were then optimised specifically for the Tra2β protein (Konig *et al.*, 2011).

3.3.5.1 UV cross-linking of MDA-MB-231 cells

MDA-MB-231 cells were grown on 100mm tissue culture plates to approximately 80% confluency. The media was removed and cells were covered in ice-cold 1×PBS and irradiated with 150-400mJ/cm^2 of UV at 254nm using a UV Stratalinker (Stratagene). The UV-irradiated MDA-MB-231 cells were harvested with a cell scraper into 1.5ml microcentrifuge tubes and the cells were precipitated by centrifugation for 10 seconds at 13,000×g at 4 °C. The 1×PBS was removed and cell pellets were snap frozen on dry ice until ready for use.

3.3.5.2 Preparation of magnetic beads

Magnetic beads were prepared by washing 100µl of Protein A Dynabeads (Invitrogen) twice with lysis buffer (Table 3.3 A) and resuspending them in 100µl lysis buffer with

5µg Tra2β antibody (Abcam). The magnetic beads were incubated at room temperature for 1 hour on a rotating platform and washed twice with lysis buffer prior to the addition of cell lysates.

3.3.5.3 Cell lysis, partial RNA digestion and immunoprecipitation

The UV-irradiated cell pellets were resuspended in lysis buffer and treated with Turbo DNase I (Ambion), together with either a high concentration (1:10 dilution) or a low concentration (1:500 dilution) of RNase I (Ambion). The cell lysates were incubated at 37°C for 3 minutes while shaking at 1,100 rpm (Eppendorf Thermomixer) and immediately stored on ice. The cellular debris was precipitated by centrifugation at 13,000×g and 4°C for 20 minutes, followed by careful collection of the supernatant. The supernatant was added to the magnetic beads for immunoprecipitation of Tra2β-RNA complexes and incubated at 4°C for 2 hours on a rotating platform. After incubation, the supernatant was removed and the magnetic beads were washed twice in high-salt buffer (Table 3.3 B), and then washed twice in wash buffer (Table 3.3 C).

3.3.5.4 Dephosphorylation of RNA 3' end and linker ligation

The 3' ends of the RNA were dephosphorylated in 20µl of Polynucleotide kinase (PNK) mix (Table 3.3 D) and incubated at 37°C for 20 minutes. The samples were washed once with wash buffer and once with high-salt buffer, followed two more washes in wash buffer. The pre-adenylated L3 linker (Table 3.4) was ligated to the 3' ends by resuspending the magnetic beads in 20µl linker ligation mix (Table 3.3 E) and incubated overnight at 16°C.

3.3.5.5 Radioactive labelling of RNA 5' ends

The magnetic beads were washed twice with 1ml wash buffer, before the magnetic beads were resuspended in 8µl of hot PNK mix containing ^{32}P-γ-ATP and incubated at 37°C for 5 minutes to radioactively label the 5' end of the RNA molecules. The hot PNK mix was carefully removed and the magnetic beads were resuspended in 20µl NuPAGE loading buffer (Invitrogen). The beads were incubated on a thermomixer at 70°C for 10 minutes to disassociate the protein-RNA complexes from the beads. The empty magnetic beads were then precipitated on a magnet, so that the supernatant could be loaded onto a 4-12% NuPAGE Bis-Tris gel (Invitrogen) for electrophoretic separation.

3.3.5.6 SDS-PAGE and membrane transfer

The samples were loaded onto a 4-12% NuPAGE Bis-Tris gel (Invitrogen) with 1× MOPS running buffer (Invitrogen). A pre-stained protein ladder (Invitrogen) was used as a reference molecular weight marker and electrophoresis was performed at 180V for 1 hour. After size separation, the protein-RNA complexes were then transferred to a nitrocellulose membrane (BioRad) using a Novex wet transfer apparatus (Invitrogen) for 2 hours at 30V. The membrane was rinsed in 1× PBS buffer and wrapped in cling film, prior to exposure to a BioMax XAR Film (Kodak) for detection. The membrane was exposed to film at -80°C for 30 minutes, 1 hour or overnight.

3.3.5.7 RNA isolation

The protein-RNA complexes were isolated from the membrane using the autoradiograph as a template. Protein-RNA complexes from the low concentration RNase sample were cut out from the diffuse radioactive signal seen above the single band observed in the high concentration RNase sample (which corresponded to the molecular weight of Tra2β, approximately 40kDa). The slices of membrane were placed in a 1.5ml microcentrifuge tube and incubated with 2mg/ml Proteinase K (Roche) diluted in PK buffer (Table 3.3 F) at 37°C for 20 minutes while shaking at 1100 rpm. A separate PK buffer containing 7M urea was then added to the tube and incubated at 37°C for a further 20 minutes. The sample, plus 400µl phenol/chloroform (Ambion) was added to a Phase Lock Gel Heavy tube (VWR). The solution was incubated at 30°C for 5 minutes while shaking at 1100 rpm prior to phase separation by centrifugation at 13,000×g for 5 minutes. The aqueous layer was transferred into a fresh tube and mixed with 0.5µl GlycoBlue (Ambion), 40µl 3M sodium acetate (pH 5.5) and 1ml 100% ethanol to precipitate the RNA overnight at -20°C.

3.3.5.8 Reverse transcription

The precipitated RNA was reverse transcribed in a 7.25µl RNA/primer mix (Table 3.3 G) containing 'Rclip' primers with individual barcode sequences for each replicate (Table 3.4). The RNA samples were incubated at 70°C for 5 minutes and cooled to 25°C before 2.75µl reverse transcription mix (containing SuperScript III reverse transcriptase (Invitrogen)) was added. The reverse transcription reaction was performed under the

following conditions: 25°C for 5 minutes, 42°C for 20 minutes, 50°C for 40 minutes and 80°C for 5 minutes before cooling to 4°C. The newly transcribed cDNAs were precipitated with the addition of 90µl TE buffer, 0.5µl GlycoBlue, 10µl 3M sodium acetate pH 5.5 and 250µl 100% ethanol, incubated overnight at -20°C.

3.3.5.9 cDNA purification

The precipitated cDNA was resuspended in 6µl of water and 2× TBE-urea loading buffer (Invitrogen) and incubated at 80°C for 3 minutes. The cDNA samples were loaded onto a 6% TBE-urea gel (Invitrogen) beside a low molecular weight marker and electrophoresed at 180V for 40 minutes. Three fragments corresponding to different cDNA size ranges were cut at approximately 120-200 nucleotides (high), 85-120 nucleotides (medium) and 70-85 nucleotides (low). The gel fragments were mixed with 400µl TE buffer and crushed using a 1ml syringe plunger. The crushed gel mixture was incubated at 37°C for 2 hours while shaking at 1100 rpm. The liquid portion of the sample was transferred into a Costar SpinX column (Corning Incorporated) containing two glass pre-filters (Whatman). The tubes were centrifuged at 13,000×g for 1 minute into a fresh 1.5ml microcentrifuge tube. The samples were added with 0.5µl GlycoBlue, 40µl 3M sodium acetate pH 5.5 and 1ml 100% ethanol to precipitate again overnight at -20°C.

3.3.5.10 Ligation of primer to the 5' end of the cDNA

The precipitated cDNA was resuspended in 8µl CircLigase ligation mix (Table 3.3 H) and incubated at 60°C for 1 hour to circularise the cDNA. In order to subsequently linearise the cDNA, a primer complementary to the *Bam*HI restriction site (Table 3.4) was annealed to the 5' end of the cDNA by adding 30µl of oligo annealing mix (Table 3.3 J) and incubated at the following conditions: 95°C for 1 minute, then temperature decreased by 1°C every 20 seconds until 25°C was reached. The circular cDNA was relinearised by restriction digest with *Bam*HI, by adding 2µl *Bam*HI (Fermentas) and incubating at 37°C for 30 minutes. Samples were mixed with 50µl TE buffer, 0.5µl GlycoBlue, 10µl 3M sodium acetate pH 5.5 and 250µl 100% ethanol to precipitate overnight at -20°C.

3.3.5.11 PCR amplification

The cDNA library was PCR-amplified by adding a PCR mix (Table 3.3 I) containing the P5 and P3 Solexa primers (Table 3.4). PCR was performed at the following conditions: 94°C for 2 minutes, 30 cycles of [94°C for 15 seconds, 65°C for 30 seconds, 68°C for 30 seconds], 68°C for 3 minutes and hold at 4°C. The size of the PCR-amplified fragments was determined by agarose gel electrophoresis.

3.3.5.12 High-throughput sequencing

Prior to sequencing the iCLIP libraries, the success of the experiment was monitored at two crucial steps: (1) the autoradiograph of protein-RNA complex after membrane transfer and (2) the gel image of the PCR-amplified products. Once verified, samples were submitted for high-throughput sequencing at the Institute of Genetic Medicine (Newcastle University, UK), in collaboration with Professor Bernard Keavney's research group. The samples were prepared for sequencing using the TruSeq Sample Preparation kit (Illumina) and the three replicates were sequenced on one lane of the Illumina Genome Analyser II system (GAIIx, Illumina).

3.3.5.13 Bioinformatic analysis of iCLIP sequencing data

The iCLIP sequencing data was transferred to Prof. Tomaz Curk (University of Ljubljana, Slovenia) for analysis and mapping to the human genome sequence using iCount (http://icount.fri.uni-lj.si/), a bioinformatics pipeline developed for the analysis of iCLIP data. The transcriptome-wide distribution of Tra2β binding sites was visualised using the UCSC Genome Browser (http://genome.ucsc.edu/). iCLIP data analysis, crosslink site identification and quantification, randomization of iCLIP positions and pentamer enrichment analysis were performed as described previously (Wang *et al.*, 2010). Briefly, we used the human genome annotation version hg19, and gene annotations from Ensembl 59. Experiment barcode and random barcodes were registered and removed from iCLIP reads. After trimming we ignored reads shorter than 11 nucleotides. Remaining trimmed reads were then mapped using Bowtie (Langmead *et al.*, 2009), allowing two mismatches and accepting only reads with single hits. Crosslink sites were initially identified as the first nucleotide upstream of the iCLIP tag, and then filtered to determine statistically significant crosslink sites and those which occurred in

clusters (clusters are defined as iCLIP sequencing reads mapping within 15 nucleotides intervals) and with a significant iCLIP-tag count, compared to randomised positions, as described in (Wang *et al.*, 2010).

(A) Lysis buffer
50mM Tris-HCl, pH 7.4
100mM NaCl
1% NP-40
0.1% SDS
0.5% sodium deoxycholate
Protease inhibitor

(B) High-salt buffer
50mM Tris-HCl, pH 7.4
1M NaCl
1mM EDTA
1% NP-40
0.1% SDS
0.5% sodium deoxycholate

(C) Wash buffer
20mM Tris-HCl, pH 7.4
10mM MgCl2
0.2% Tween-20

(D) PNK mix (1X)
15µl water
4µl 5× PNK pH 6.5 buffer
0.5µl PNK enzyme
0.5µl RNasin

(E) Linker ligation mix (1X)
9µl water
4µl 4× ligation buffer
1µl RNA ligase
0.5µl RNasin
1.5µl pre-adenylated linker [20µM]
4µl PEG400

(F) PK buffer (1X)
100mM Tris-HCl pH 7.4
50mM NaCl
10mM EDTA

(G) RNA/primer mix (1X)
6.25µl water
0.5µl Rclip primer [0.5 pmol/µl]
0.5µl dNTP mix [10mM]

(H) Circligase ligation mix (1X)
6.5µl water
0.8µl 10× CircLigase Buffer II
0.4µl 50mM MnCl2
0.3µl Circligase II

(I) Oligo annealing mix (1X)
26µl water
3µl FastDigest Buffer
1µl cut oligo [10µM]

(J) PCR mix (1X)
19µl cDNA
0.5 µl P3 primer [10µM]
0.5 µl P5 primer [10µM]
20µl Accuprime Supermix 1 enzyme

Table 3.3 Complete list of buffer components and reaction mixes used in the iCLIP experiments.

Primer	Sequence
L3 linker	/5rApp/AGATCGGAAGAGCGGTTCAG/3ddC/
Rclip 1	5'phosphateNNAACCNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
Rclip 2	5'phosphateNNACAANNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
Rclip 3	5'phosphateNNATTGNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC
Cut oligo	GTTCAGGATCCACGACGCTCTTCaaaa
P5 Solexa	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
P3 Solexa	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

Table 3.4 Primers used in the iCLIP experiment. Randomly generated nucleotides were included in the Rclip primers (yellow 'N') to differentiate between unique cDNAs and multiple copies of the same cDNA which had been PCR-amplified. A four-base barcode (red) was used to allow multiplexing of multiple experiments in a single sequencing lane. Location of the *Bam*HI restriction site is shown in green. The *Bam*HI restriction site allowed linearization of the single-stranded circular DNA once the 'cut oligo' had annealed to create a short region of double-stranded DNA.

3.4 Results

3.4.1 Tra2β efficiently suppresses Tra2α protein expression in MDA-MB-231 cells

In chapter 2, we identified a poison exon from *Tra2a* which was highly responsive to Tra2α and Tra2β protein expression using a minigene assay in HEK-293 cells (Figure 2.3 H). Co-expression of full-length Tra2α or Tra2β significantly increased splicing inclusion of the *Tra2a* minigene exon. I also confirmed that splicing inclusion of the endogenous *TRA2A* poison exon was reduced after knockdown of endogenous Tra2β in the MDA-MB-231 cell line, suggesting the *TRA2A* poison exon was a genuine endogenous target (Figure 2.16 E). As poison exons introduce premature termination codons (PTCs) into mRNA transcripts and inhibit translation of the full-length proteins (Stoilov *et al.*, 2004; Lareau *et al.*, 2007; McGlincy and Smith, 2008; Grellscheid *et al.*, 2011a), this data suggested that Tra2β may directly regulate Tra2α protein expression by promoting inclusion of a poison exon within the *TRA2A* mRNA. Tra2β has also been shown to promote inclusion of a poison exon within its own pre-mRNA (Stoilov et al., 2004). Therefore to investigate whether human Tra2 proteins cross-regulate their endogenous protein expression, I detected Tra2α and Tra2β protein expression by Western Immunoblotting, following siRNA-mediated depletion of the two endogenous Tra2 proteins.

Tra2α protein expression is relatively low in MDA-MB-231 cells, but significantly increased after siRNA-mediated depletion of Tra2β (Figure 3.1 A, top panel, compare lanes 1 and 3). Although weak, the Tra2α signal from the Western Immunoblot was of the predicted size and was almost completely eliminated following transfection with a *TRA2A* specific siRNA, suggesting that this band was specific for the Tra2α protein (Figure 3.1 A, top panel, compare lanes 1 and 2). Depletion of the Tra2α protein had a much smaller effect on Tra2β protein levels (Figure 3.1 A, middle panel). This result was reproducible in three independent experiments and the expression of Tra2α and Tra2β was quantified from each individual Immunoblot (relative to expression of the loading control α-Tubulin) using the Image J graphics software (Figure 3.1 B). Up-regulation of Tra2α protein expression following depletion of Tra2β was highly statistically significant. On the other hand, I observed a small increase in Tra2β protein expression following depletion of Tra2α, which was not statistically significant. This

effect was also confirmed using two independent sets of siRNA, targeting different regions of the *TRA2A* and *TRA2B* mRNAs (Figure 3.2).

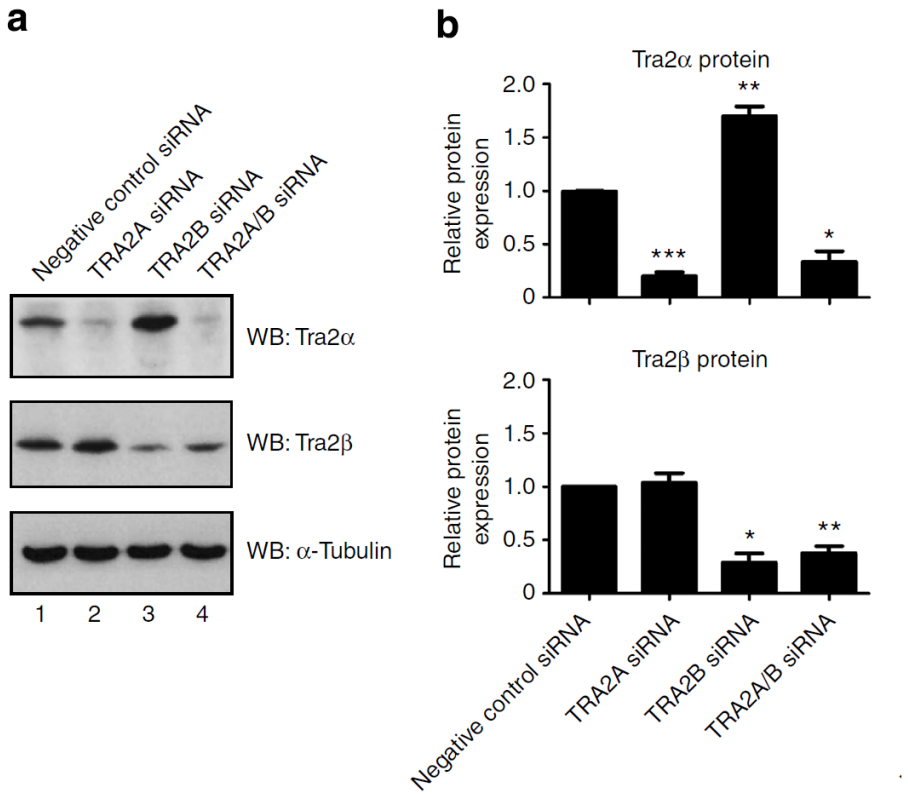


Figure 3.1 Depletion of Tra2β resulted in up-regulation of Tra2α protein expression. Endogenous Tra2α and Tra2β protein expression in MDA-MB-231 cells was detected by Western Immunoblot. (a) Depletion of Tra2β induced up-regulation of Tra2α protein expression, whereas depletion of Tra2α had no significant effect on Tra2β protein expression. (b) Western Immunoblots from three independent experiments were quantified (relative to expression of the loading control α-Tubulin) using Image J graphics software, which confirmed that up-regulation of Tra2α following depletion of Tra2β was reproducible and statistically significant. Conversely, depletion of Tra2α had no significant effect on Tra2β protein expression. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is taken from (Best et al., 2014b).

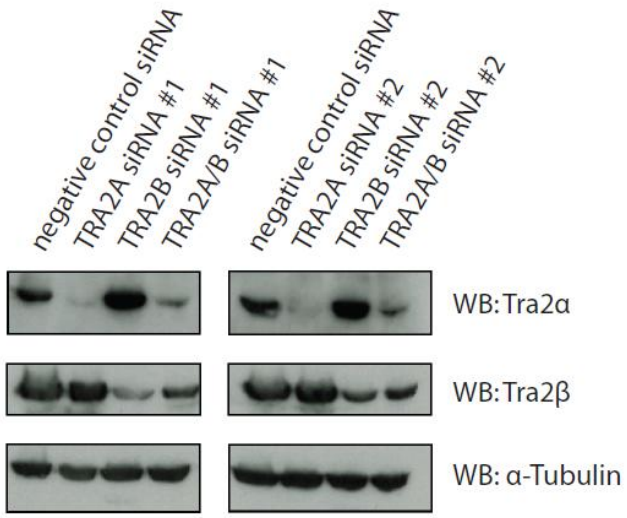


Figure 3.2 Depletion of Tra2β induced reciprocal up-regulation of Tra2α protein expression using 2 independent sets of siRNA. Depletion of Tra2α had no significant effect on Tra2β protein expression using either of the *TRA2A* siRNA sequences. This figure is taken from (Best et al., 2014b)

3.4.2 Cross-regulation of expression occurs at the RNA level through splicing regulation of reciprocal poison exons

To determine whether cross-regulation between Tra2α and Tra2β occurred at the RNA level, I monitored steady state mRNA expression of *TRA2A* and *TRA2B* by quantitative real-time PCR (qPCR) (Figure 3.3 A). Depletion of Tra2β led to a significant increase in *TRA2A* mRNA expression, whilst knockdown of Tra2α also resulted in a smaller but statistically significant increase in *TRA2B* mRNA expression, confirming that each Tra2 protein negatively regulated the expression of the other at the RNA level (Figure 3.3 A). The effect of Tra2β depletion on *TRA2A* mRNA expression was larger than the effect of Tra2α depletion on *TRA2B* mRNA expression, consistent with the asymmetrical effect observed at the protein level by Western Immunoblot.

To determine the effect of Tra2α knockdown on splicing inclusion of *TRA2B* poison exon and vice versa, splicing of the endogenous *TRA2A* and *TRA2B* poison exons was monitored by RT-PCR and capillary gel electrophoresis. Depletion of Tra2β significantly reduced inclusion of the *TRA2A* poison exon to almost undetectable levels (Figure 3.3 B, top panel), whilst depletion of Tra2α also resulted in a smaller but statistically significant decrease in inclusion of the *TRA2B* poison exon (Figure 3.3 B, lower panel). The larger effect of Tra2β depletion on inclusion of the *TRA2A* poison exon is consistent with the larger effect seen on *TRA2A* mRNA and Tra2α protein expression,

suggesting that both Tra2 proteins influence expression of the other protein through regulation of their reciprocal poison exons, but that this effect is asymmetrical in MDA-MB-231 cells.

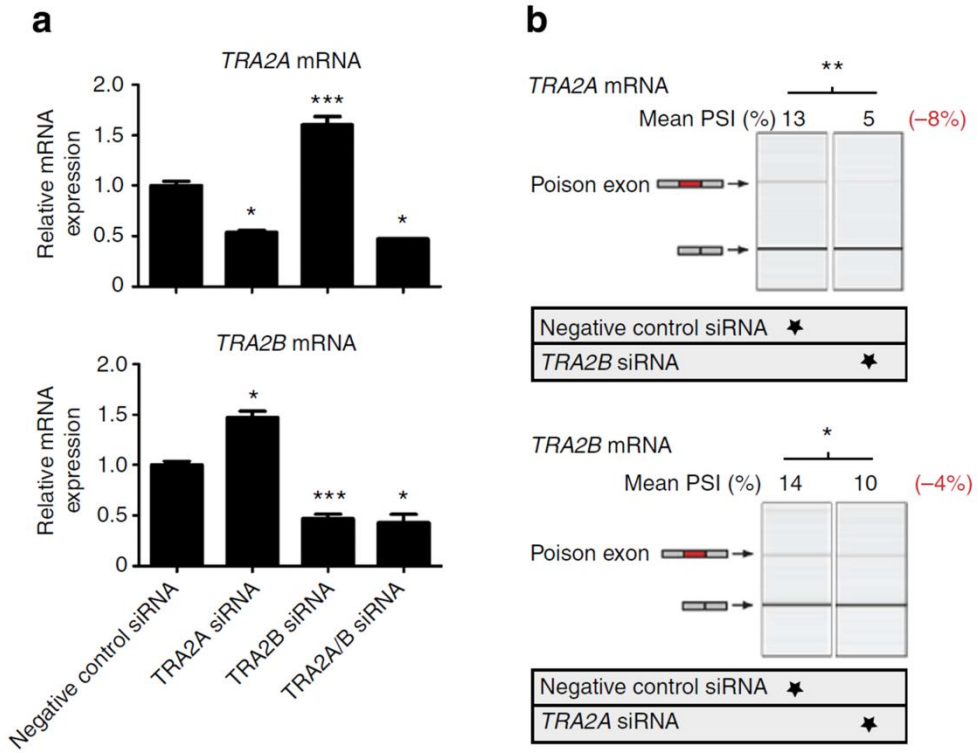


Figure 3.3 Cross-regulation between Tra2α and Tra2β occurs through splicing regulation of reciprocal posion exons. (a) *TRA2A* and *TRA2B* mRNA expression was determined by qPCR following siRNA-mediated depletion of the Tra2 proteins. Transfection with *TRA2B* siRNA led to significant increase in *TRA2A* mRNA, whilst transfection with *TRA2A* siRNA also led to a smaller but significant increase in *TRA2B* mRNA. (b) The percentage splicing inclusion (PSI) of poison exons within the *TRA2A* and *TRA2B* mRNA following knockdown of the reciprocal Tra2 protein. Both poison exons were significantly affected, but depletion of Tra2β had a greater effect on inclusion of the *TRA2A* poison exon than vice versa. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is taken from (Best et al., 2014b).

3.4.3 Identification of direct RNA targets of Tra2β in MDA-MB-231 cells using iCLIP

To identify endogenous RNA targets of Tra2β in MDA-MB-231 cells, I used individual nucleotide resolution cross-linked immunoprecipitation (iCLIP) according to the published iCLIP protocol (Konig et al., 2011). iCLIP was developed to study transcript regulation by RNA-binding proteins (RBPs) at the molecular level, by providing high-resolution positional information on protein-RNA interactions (Konig et al., 2011). iCLIP utilises UV cross-linking of proteins and RNA molecules followed by stringent purification steps to isolate RNA sequences directly bound by the RNA-binding protein of interest. Coupled with high-throughput sequencing, iCLIP facilitates the identification of protein-RNA interactions on a transcriptome-wide scale.

For a schematic representation of the iCLIP experiment, please refer to Figure 3.4. The experimental stages described in this paragraph correspond to Figure 3.4. During the iCLIP experiment, cells are irradiated with UV light to induce irreversible covalent cross-linking between proteins and RNA molecules in direct contact (1). The cells are subsequently lysed (2) and the RNA is partially digested using an RNase enzyme (3). The protein-RNA complexes are then immunoprecipitated using an antibody against the protein of interest (4). During the IP, the protein-bound RNA is dephosphorylated (5) and a DNA linker is added to the 3' end of the RNA (6), whilst the 5' end of the RNA is radio-labelled with ^{32}P (7). The protein-RNA complexes are then separated by size using SDS-PAGE and transferred to a nitrocellulose membrane to remove free RNA (8). The radio-labelled protein-RNA complexes can be detected by autoradiograph, which serves as a template to isolate the protein-RNA complexes of interest based on their molecular weight. The protein is subsequently digested using proteinase K, leaving a small polypeptide remaining at the cross-linked nucleotide (9). The remaining RNA is isolated using phenol/chloroform and the purified RNA is reverse transcribed into cDNA (10). The cDNA is then separated by size using gel electrophoresis (11) and appropriate size fragments are purified from the gel. As the majority of reverse transcription (RT) reactions terminate at the cross-linked polypeptide, the site of protein-RNA interaction can be preserved by circularising the cDNA using a ssDNA circligase enzyme (12), so that the site of protein-RNA interaction is now adjacent to the RT primer sequence. Another primer is then annealed to the circular cDNA (13), creating a double-stranded region to allow cleavage by the restriction endonuclease

*Bam*HI (14). Finally, the linearised cDNA is PCR-amplified using primers containing adapter sequences for paired-end sequencing to produce a cDNA library (15). Further details of the iCLIP protocol are provided in methods section 3.3.5.

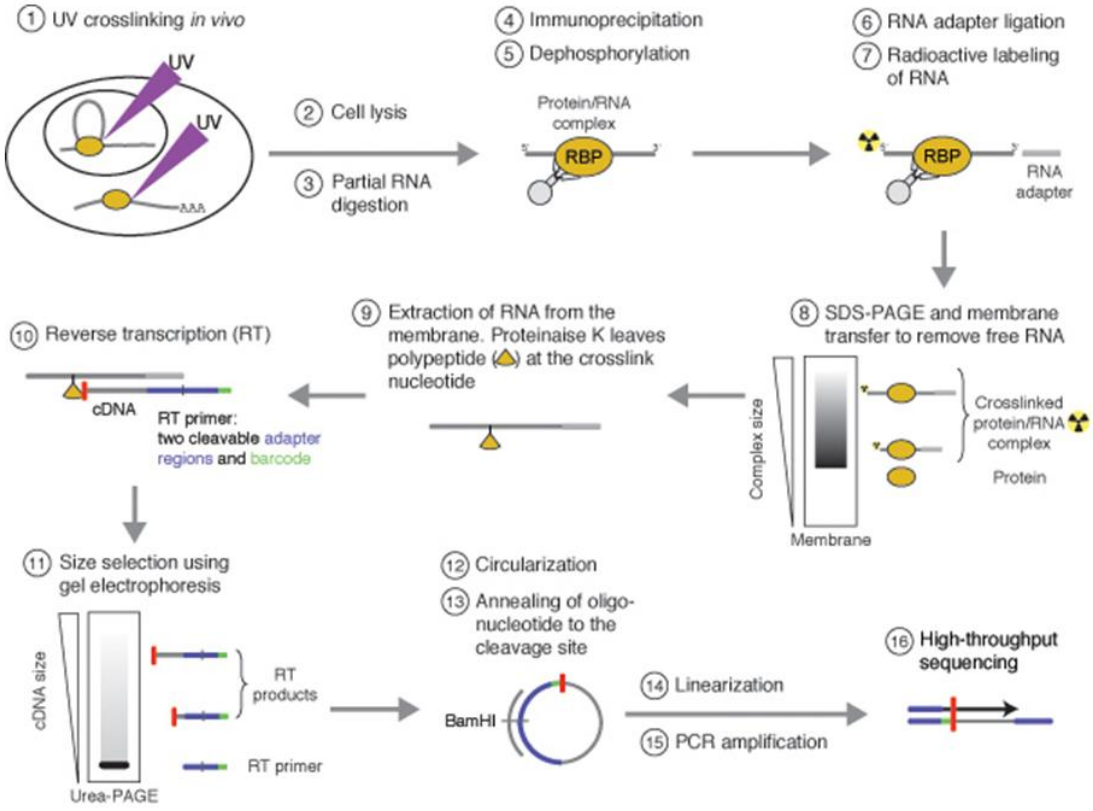


Figure 3.4 Schematic representation of the iCLIP protocol. This image is from the published iCLIP protocol by Konig et al. 2011.

3.4.3.1 Immunoprecipitation of endogenous Tra2β protein

An important component of the iCLIP experiment is the efficient and specific immunoprecipitation (IP) of the RNA-binding protein of interest. As a preliminary experiment to determine the specificity of the α-Tra2β antibody and to optimise the efficient IP of endogenous Tra2β protein, endogenous Tra2β was immunoprecipitated from MDA-MB-231 cells using varying amounts of α-Tra2β antibody (2-10µg) and then detected by Western Immunoblotting (Figure 3.5). A band corresponding to the known molecular weight of Tra2β (approximately 37kDa) was detected in the input sample and the α-Tra2β antibody IP sample, but not the mock IP samples using either rabbit IgG or no antibody as controls. 5µg of α-Tra2β antibody was determined as sufficient to remove detectable amounts of Tra2β from the supernatant sample, whilst Tra2β

remained detectable within the supernatants of both the rabbit IgG and no antibody controls. As the same rabbit α-Tra2β antibody used in the IP was also used as the primary antibody to probe the samples by Western Immunoblotting, a second band corresponding to the rabbit antibody light chain (approximately 25kDa) was also detected in both the α-Tra2β antibody IP sample and the rabbit IgG IP sample, but not in the no antibody IP sample. This band could have been eliminated by probing with a second α-Tra2β antibody raised from a different species, allowing a secondary antibody to be used which would not recognise the original rabbit antibody used in the IP.

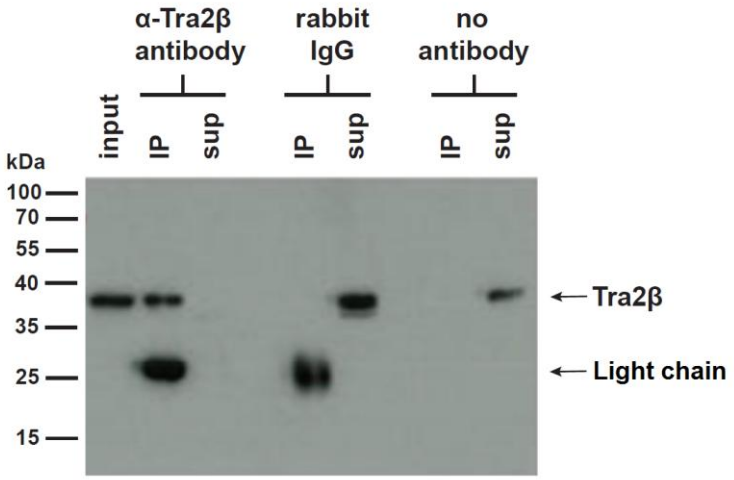


Figure 3.5 Immunoprecipitation of endogenous Tra2β protein from MDA-MB-231 cells.

3.4.3.2 Optimisation of the Tra2β iCLIP experiment

In order to optimise the Tra2β iCLIP experiment, MDA-MB-231 cells were UV-irradiated with 3 different energies (150, 250 and 400(X100mJ/cm^2)) as cross-linking efficiency can vary between individual proteins. Initially, one third of a 100mm plate of MDA-MB-231 cells at 50-80% confluency was used per sample and immunoprecipitated using 2μg of Tra2β antibody. Cells irradiated with 400(X100mJ/cm^2) produced the strongest signal, indicating there was a higher proportion of cross-linked RNA at this energy. Following SDS-PAGE and transfer to nitrocellulose membrane, the membrane was exposed to x-ray film for 30 minutes, 1 hour and overnight. Only after an overnight exposure could I detect a weak signal. The previous HITS-CLIP experiment (Grellscheid et al., 2011a) had suggested that for a successful CLIP experiment (in which sufficient

protein-RNA complexes are isolated), a strong signal should be detected after a 1-2 hour exposure, therefore I began to optimise the experiment by varying the quantity of cells, level of UV-irradiation, antibody concentration and RNase concentration, in order to increase the signal detected by autoradiograph (produced from the radio-labelled protein-RNA complexes).

For each experiment, two samples are run in parallel and treated with either a high or low concentration of RNase. The high concentration RNase sample serves as a control for antibody specificity; as the RNA is efficiently digested in this sample, a single band close to the molecular weight of the protein should be observed. An autoradiograph showing optimisation of the Tra2β iCLIP experiment is shown in Figure 3.6. In the high concentration RNase samples (1/10 dilution), a single band corresponding close to the molecular weight of Tra2β (approximately 37KDa) was observed, as most of the cross-linked RNA has been efficiently digested. However in the low concentration RNase sample (1/500 dilution), a diffuse radioactive signal is seen stretching above the molecular weight of the protein, as the cross-linked RNA is only partially digested, retarding the migration of the protein through the gel. Protein-RNA complexes immediately above the 40kDa band were isolated from the low RNase sample (highlighted red). The strongest signal was obtained using one 100mm plate of 50-80% confluent MDA-MB-231 cells, irradiated with 400(X100mJ/cm^2) UV, immunoprecipiated using 5µg of α-Tra2β antibody and digested using a 1/500 RNase I dilution (final RNase concentration 0.2 U/µl) (Figure 3.6, left hand side). A higher concentration of RNase (1/10 dilution) was necessary to control for antibody specificity (final RNase concentration 10 U/µl). The high concentration RNase sample indicated that a protein corresponding to the known molecular weight of Tra2β was being immunoprecipitated.

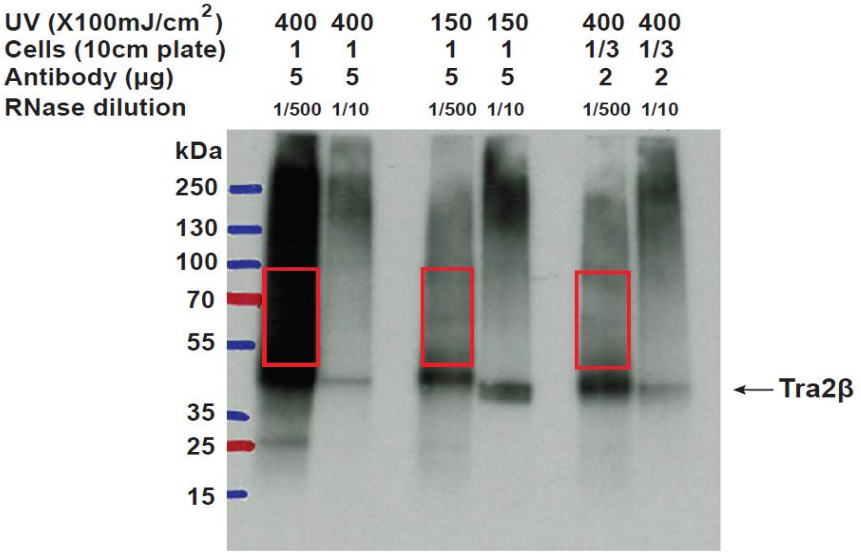


Figure 3.6 Autoradiograph showing optimisation of the Tra2β iCLIP experiments in MDA-MB-231 cells following a 1 hour exposure. Following immunoprecipitation of the Tra2β-RNA complexes, they were radio-labelled and then separated by denaturing polyacrylamide gel electrophoresis, transferred to a nitrocellulose membrane and exposed to film for detection. Protein-RNA complexes above the molecular weight of Tra2β were isolated from the low concentration RNase samples (highlighted red).

Following isolation of the protein-RNA complexes, the protein was digested using Proteinase K and the remaining RNA was reverse transcribed into cDNA. The cDNA was separated by size using gel electrophoresis and three cDNA fractions corresponding to 120-200nt (High), 85-120nt (Medium) and 70-85nt (Low) were excised from the gel and purified. Finally, the cDNA fractions were circularised and then digested with the restriction endonuclease *Bam*HI, to relinearize the cDNA before PCR amplification to produce a cDNA library for sequencing.

In order to optimise the cDNA library for sequencing, various PCR cycle numbers were tried during the PCR amplification of the cDNA library. From early experiments, only cDNA amplified for 35 cycles of PCR could be visibly seen on an agarose gel (Figure 3.7). However, following optimisation, PCR products were clearly visible as low as 22 cycles of PCR (Figure 3.8). Optimising the experimental conditions included increasing the amount of cross-linking (UV irradiation) from 150 to 400 (X100mJ/cm^2), increasing the quantity of cells per sample from one third of a 100mm tissue culture plate to one whole 100mm plate, and increasing the amount of Tra2β antibody used during immunoprecipiation from 2µg to 5µg. Reducing the PCR cycle number is essential for optimal sequencing results, as a lower cycle number significantly increases the number

of unique cDNA sequences and reduces PCR-amplification bias of more abundant transcripts.

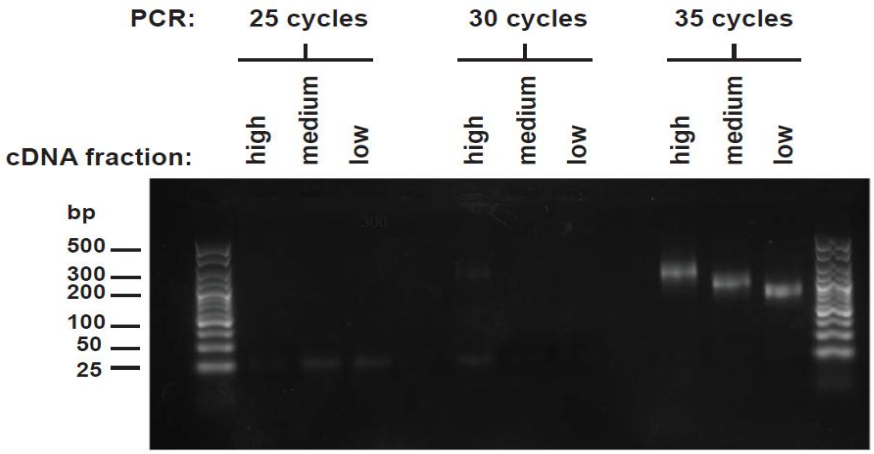


Figure 3.7 PCR amplification and agarose gel electrophoresis of cDNA from early Tra2β iCLIP experiments. In order to optimise the cDNA library for sequencing, several cycle numbers (25, 30 and 35) were tried during the PCR amplification of the cDNA. From early experiments, only cDNA amplified for 35 cycles could be visibly detected on an agarose gel.

Following optimisation, the Tra2β iCLIP experiment was repeated in three independent experiments with the addition of a no antibody control experiment. The final cDNA libraries were visualised by agarose gel electrophoresis to ensure the cDNA fragments were of appropriate size for sequencing (recommended 200-500bp) (Figure 3.8). As the high, medium and low size cDNA fractions were each within this range, all fractions were included for sequencing. Importantly, no PCR products were visible in the no antibody control experiment (Figure 3.8).

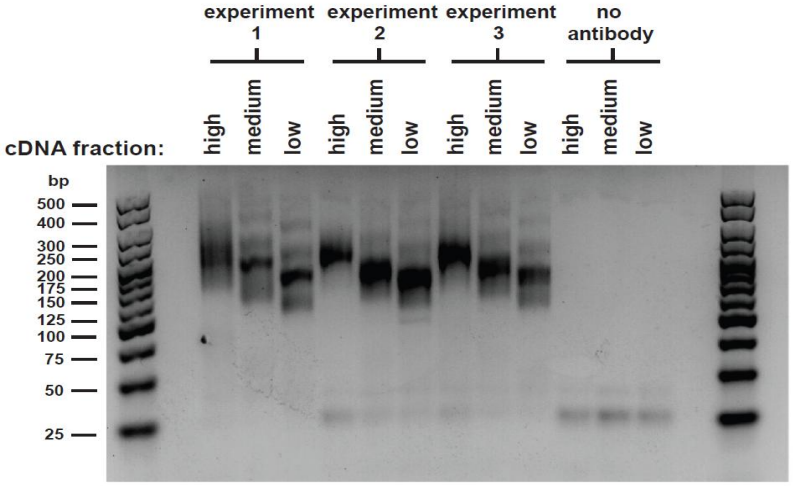


Figure 3.8 Final cDNA libraries from three independent Tra2β iCLIP experiments, visualised by agraose gel electrophoresis after 22 cycle PCR amplification. Importantly, no PCR products were visible in the no antibody control experiment.

Equal quantities of the 9 cDNA libraries (high, medium and low cDNA fractions from 3 independent experiments) were pooled together and submitted for sequencing. The final cDNA library was prepared by Mr Rafiq Hussain (Newcastle University) using the Illumina Truseq DNA kit, followed by 2X100bp paired-end sequencing using one lane on an Illumina Genome Analyser IIx. The sequencing data was then transferred to Prof. Tomaz Curk (University of Ljubljana) for analysis and mapping to the human genome sequence using iCount, a bioinformatics pipeline developed for the analysis of iCLIP data. In total, 7,443,903 reads were successfully mapped back to the human genome sequence, of which 3,338,710 were unique cDNAs considered for downstream analysis (subsequently referred to as iCLIP tags or cross-links).

3.4.4 Global distribution of Tra2β binding in MDA-MB-231 cells

The most abundant pentamers recovered in the iCLIP tags were highly enriched in 'GAA' nucleotide sequences (Figure 3.9), matching the Tra2β binding site identified from the previous Tra2β HITS-CLIP experiment in the mouse testis and from SELEX experiments using purified Tra2β proteins (Tacke *et al.*, 1998b; Grellscheid *et al.*, 2011a). This served as an early indication that the iCLIP experiments had successfully captured genuine Tra2β-RNA interactions. The combined human iCLIP data in MDA-MB-231 cells provided substantially more coverage than previously obtained from the HITS-CLIP experiment in mouse testis (in which just 177,457 reads were mapped back to the mouse genome). This greatly increased tag coverage enabled a much more

comprehensive map of Tra2β binding sites in human MDA-MB-231 cells than was previously obtained from the mouse testis.

Figure 3.9 The top 10 most frequently occurring pentamers recovered from the Tra2β iCLIP experiment in MDA-MB-231 cells. 'GAA' sequences are highlighted in red. This figure is taken from (Best et al., 2014b).

The genomic distribution of iCLIP tags is shown in Figure 3.10. In total, 1,546,290 (44.8%) of unique cDNAs mapped to intronic regions, suggesting the Tra2β iCLIP experiment largely captured Tra2β interactions with pre-mRNAs (Figure 3.10 A). A further 1,169,374 (33.8%) of the unique cDNAs mapped to exons (either 5'UTR, 3'UTR or ORF) (Figure 3.10 A), despite exons comprising approximately just 1% of the genome. After correcting for the relative size of each genomic region (by dividing the number of unique cDNAs mapping to each genomic region by the relative size of that region within the genome), I found that Tra2β binding is highly enriched within exons; 76.8% of Tra2β iCLIP tags mapped to exons relative to their size (5'UTR, 3'UTR or ORF), whilst a further 20.4% mapped to non-coding RNA (ncRNAs) (Figure 3.10 B). N.B. following the use of RNA-seq in subsequent chapters, it would now be possible to compare the enrichment of iCLIP tags relative to the MDA-MB-231 cell transcriptome - rather than the genome. This would likely provide a more accurate picture of Tra2β binding site enrichment by considering the expression profile of nuclear RNA, rather than the genomic DNA.

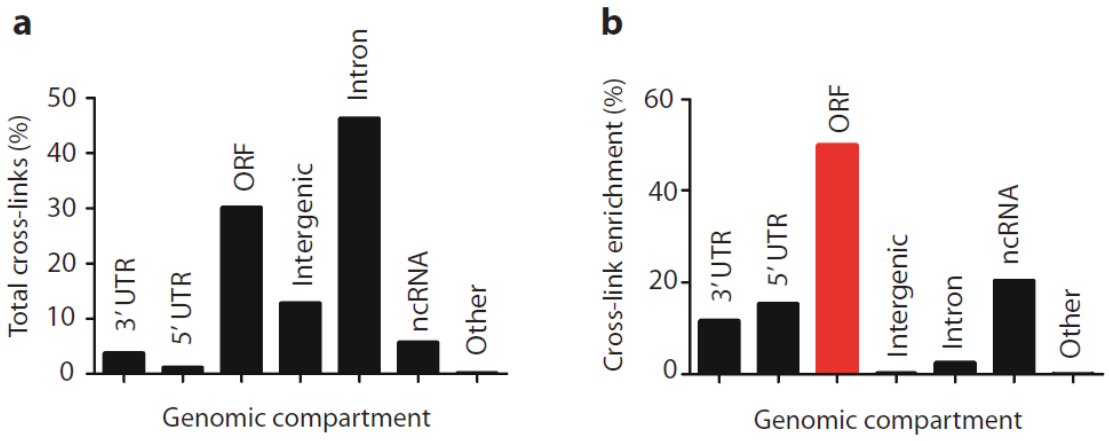


Figure 3.10 Global distributions of cross-links identified by iCLIP in MDA-MB-231 cells. (a) The percentage of total cross-links from the three independent iCLIP experiments that mapped to each genomic compartment. (b) The percentage of cross-links mapping to each genomic compartment after adjustment for the relative size of each genomic compartment within the human genome. After adjustment, Tra2β binding was found to be highly enriched within exons relative to their size (highlighted red). This figure is taken from (Best et al., 2014b).

3.4.5 Stratification of exons for functional validation

As Tra2β binding was highly enriched within exons, I initially focused on alternative exons for functional validation of splicing regulation. I began by stratifying alternative exons in order of exon junction sequences (defined as +/-300bp of either splice site) which contained the greatest number of iCLIP tags. However, as the iCLIP data does not consider relative gene expression, high numbers of iCLIP tags within an exon may be attributable to high gene expression, as well as enriched Tra2β binding. Taking this into consideration, I then stratified alternative exon junctions (+/-300bp of either splice site) with the highest number of iCLIP tags relative to iCLIP tag coverage across the whole gene, to identify exons with enriched Tra2β binding (Figure 3.11). An example of alternative exons from *ATRX*, *GLYR1* and *CEP95* identified using this method is shown in Figure 3.12.

$$\frac{\text{[iCLIP tags +/- 300bp of 3'ss] + [iCLIP tags +/- 300bp of 5'ss]}}{\text{total iCLIP tags across whole gene}}$$

Figure 3.11 Method used to stratify Tra2β target exons for validation. Exons with enriched Tra2β binding were identified by calculating the number of cross-links within or close to a particular exon, relative to the total number of cross-links across the whole gene.

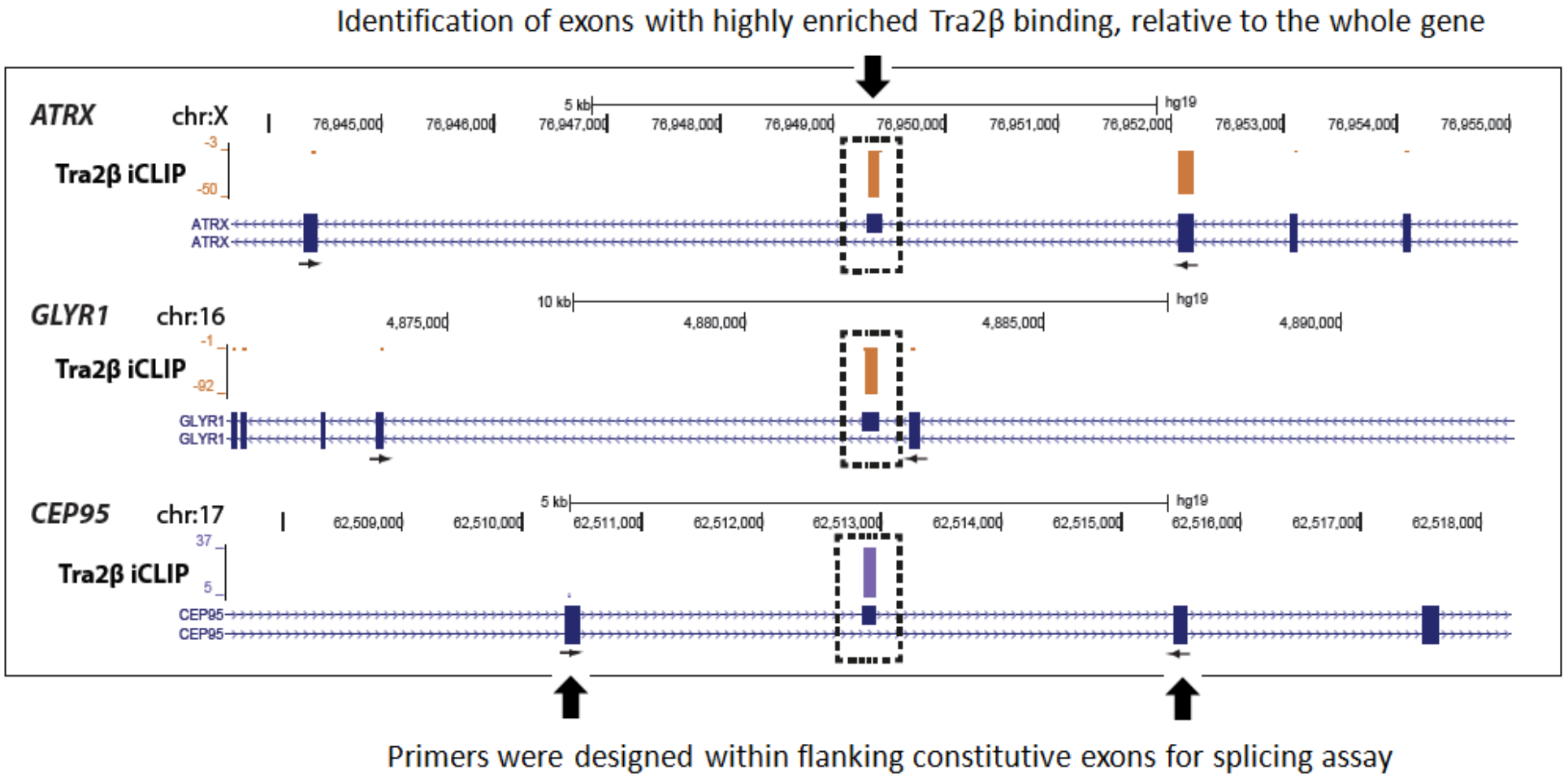


Figure 3.12 Alternative exons identified within *ATRX*, *GLYR1* and *CEP95* which contained a high density of Tra2β iCLIP tags relative to the whole gene. Primers were designed within flanking constitutive exons to measure percentage splicing inclusion (PSI) of target exons by RT-PCR.

3.4.6 Validating splicing regulation of endogenous target exons by human Tra2 proteins

Following stratification of alternative exons with enriched Tra2β binding, I designed primers within flanking constitutive exons to monitor the percentage splicing inclusion (PSI) of endogenous target exons by RT-PCR and capillary gel electrophoresis.

3.4.6.1 Up-regulation of endogenous Tra2α protein can functionally compensate for loss of Tra2β protein in splicing regulation

In the first part of this chapter, I observed that expression of endogenous Tra2α protein significantly increased following depletion of Tra2β protein. To determine whether up-regulation of Tra2α could functional compensate for loss of Tra2β protein in splicing regulation of target exons, the percentage splicing inclusion (PSI) of targets identified from the iCLIP experiment was measured after single depletion of either Tra2α or Tra2β, and after combined depletion of both Tra2α and Tra2β proteins. From an initial screen of 14 candidate exons, each of the Tra2β bound exons either did not respond or responded very weakly to single knockdown of either Tra2α or Tra2β. However, joint depletion of both Tra2α and Tra2β resulted in strong decreases in splicing inclusion for many exons and had a substantially greater effect than removing either protein alone.

The effect of removing both Tra2 proteins compared to depletion of Tra2α or Tra2β alone is shown in Figure 3.13. Clusters of Tra2β iCLIP tags mapped to alternative exons in the *ATRX*, *GLYR1* and *CEP95* genes (shown earlier in Figure 3.12). Single depletion of Tra2α protein resulted in a small but significant decrease in the PSI of the *ATRX* and *CEP95* cassette exons, whilst the mean PSI of the *GLYR1* exon was reduced but did not reach statistical significance. Single depletion of Tra2β protein resulted in a small but significant decrease in the mean PSI of all three cassette exons. Depletion of Tra2β protein had a larger effect than depletion of Tra2α for all exons tested, consistent with higher expression of the Tra2β protein in MDA-MB-231 cells. Joint depletion of both Tra2α and Tra2β resulted in a substantial decrease in the mean PSI of all three exons (highlighted red, Figure 3.13), which was consistently greater than the combined change in PSI of both single knockdowns. This data suggested that following depletion

of Tra2β, up-regulation of endogenous Tra2α protein could functionally compensate and largely maintain splicing inclusion of Tra2β target exons.

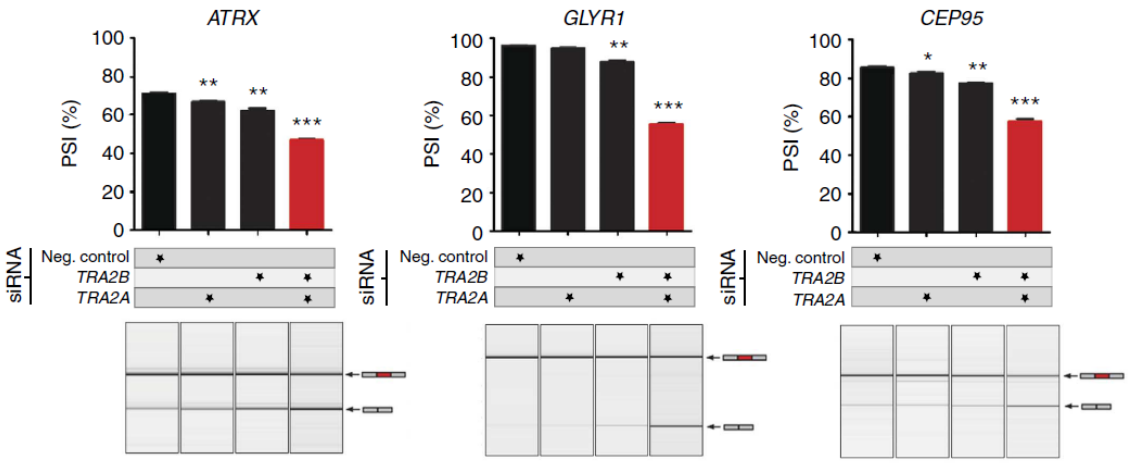


Figure 3.13 The mean percentage splicing inclusion of three novel Tra2β target exons from *ATRX*, *GLYR1* and *CEP95*. Splicing inclusion was only slightly affected by depletion of either endogenous Tra2α or Tra2β proteins, but was strongly affected by joint depletion of both Tra2α and Tra2β (red column). Splicing of endogenous targets was measured by RT-PCR and capillary gel electrophoresis (lower panels). Data represents the mean of three biological replicates ±s.e.m. (upper panels). Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is adapted from (Best et al., 2014b).

I tested this effect on a larger panel of candidate exons and obtained exactly similar results with 14 out of 14 novel Tra2β target exons tested in this way (mean PSI of Tra2 responsive exons is shown in Figure 3.14). In fact, some of the less responsive exons including cassette exons from *PAM* (Figure 3.14 K) and *BDP1* (Figure 3.14 D) only responded to depletion of both Tra2 proteins, and not to single depletion of Tra2β at all. Finally, I calculated the mean PSI change of all 14 exons tested for response to single depletion of either Tra2α or Tra2β, and combined depletion of both Tra2α and Tra2β proteins (Figure 3.14 O). There was no significant change in the average PSI change after single depletion of either Tra2 protein, but joint depletion led to a highly significant change in the average PSI change of target exons. This data from the larger panel is consistent with the hypothesis that up-regulation of endogenous Tra2α (following Tra2β depletion) is able to functionally substitute and largely maintain Tra2β target exon inclusion.

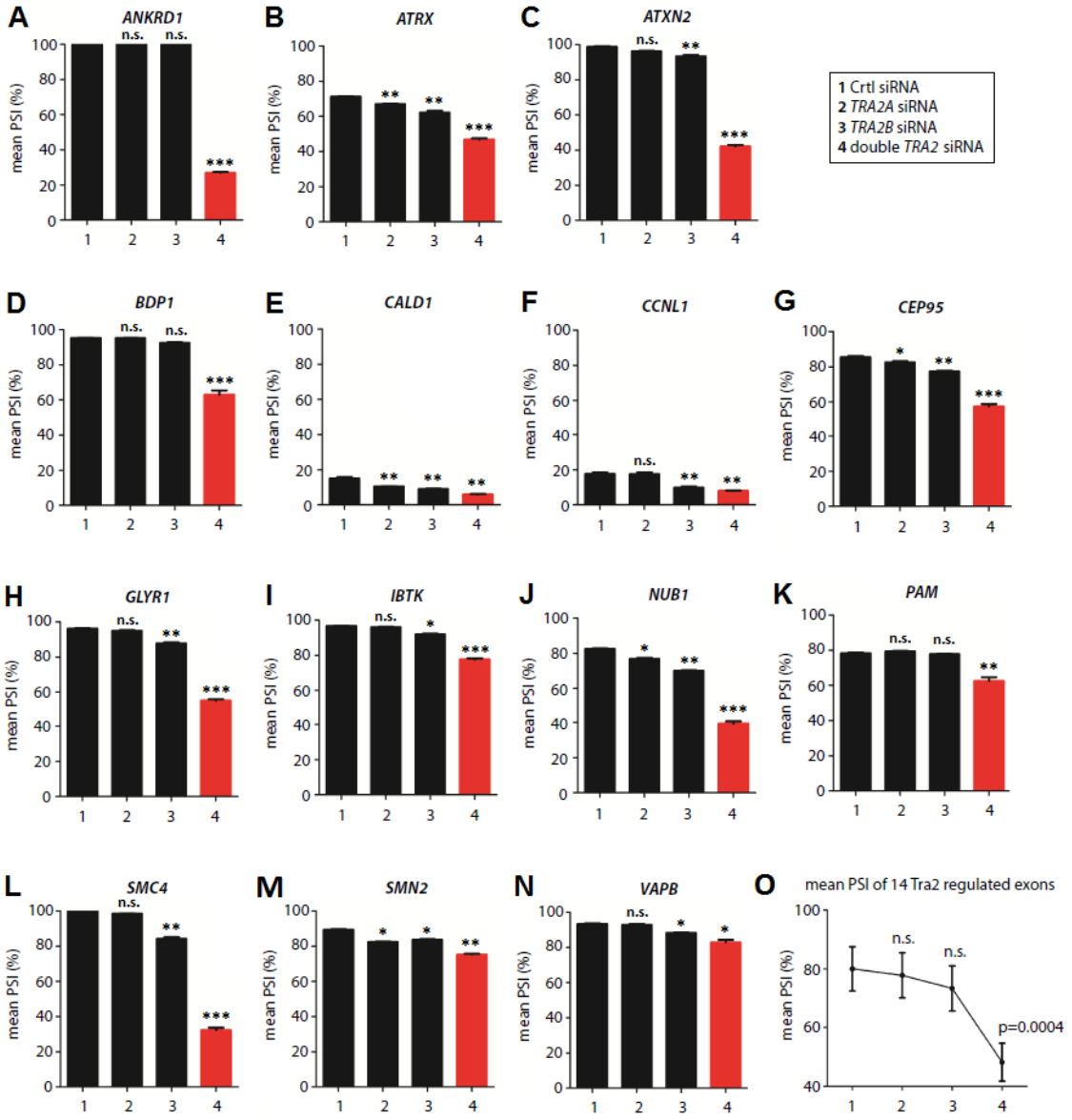


Figure 3.14 The mean PSI of fourteen Tra2 protein responsive exons identified using iCLIP following single or joint Tra2 protein depletion (A-N). Joint depletion of both Tra2α and Tra2β (red columns) was required to induce robust splicing changes in all Tra2 responsive exons tested. (O) The mean change in PSI of the 14 Tra2 responsive exons was not significant after single depletion of either Tra2α or Tra2β, but joint depletion of both Tra2α and Tra2β led to a highly significant reduction in the average PSI of the Tra2 target exons. Data represents the mean of three biological replicates ±s.e.m. (A-N). Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is adapted from (Best et al., 2014b).

3.4.6.2 Validation of additional targets exons following joint depletion of Tra2α and Tra2β

The previous data had demonstrated that depletion of both Tra2 proteins was required to observe robust changes in splicing of Tra2 protein dependent exons. Therefore I subsequently screened a more extensive panel of candidate exons by comparing control siRNA treated cells with joint Tra2α and Tra2β depleted cells. An example of a candidate exon showing highly specific Tra2β binding is shown below in Figure 3.15. A cluster of Tra2β iCLIP tags mapped highly specifically to a cassette exon within the *PRMT2* gene (Figure 3.15 A). Primers were designed within flanking constitutive exons for validation by RT-PCR. Splicing of the endogenous *PRMT2* cassette exon was measured by RT-PCR and capillary gel electrophoresis (Figure 3.15 B). Joint depletion of Tra2α and Tra2β led to a highly significant reduction is splicing inclusion of the *PRMT2* exon compared to control siRNA treated cells (-31% change in mean PSI). After screening candidate exons identified from the iCLIP data, I initially identified 32 Tra2 protein responsive exons (representative gel images and mean PSI from three biological replicates shown in Figure 3.16). High resolution iCLIP maps for all Tra2 responsive exons are provided in Appendix B.

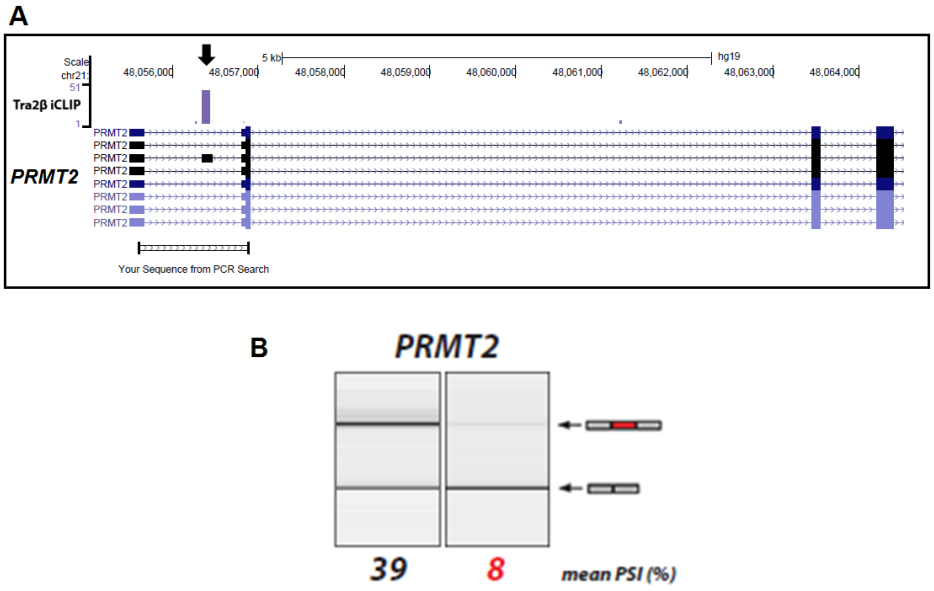


Figure 3.15 (A) Identification of a cassette exon within *PRMT2* with highly enriched Tra2β iCLIP tags relative to the whole gene. (B) Percentage splicing inclusion (PSI) of the endogenous *PRMT2* cassette exon was compared by PCR and capillary gel electrophoresis between control siRNA treated cells (39%, black) and joint Tra2 protein depleted cells (8%, red).

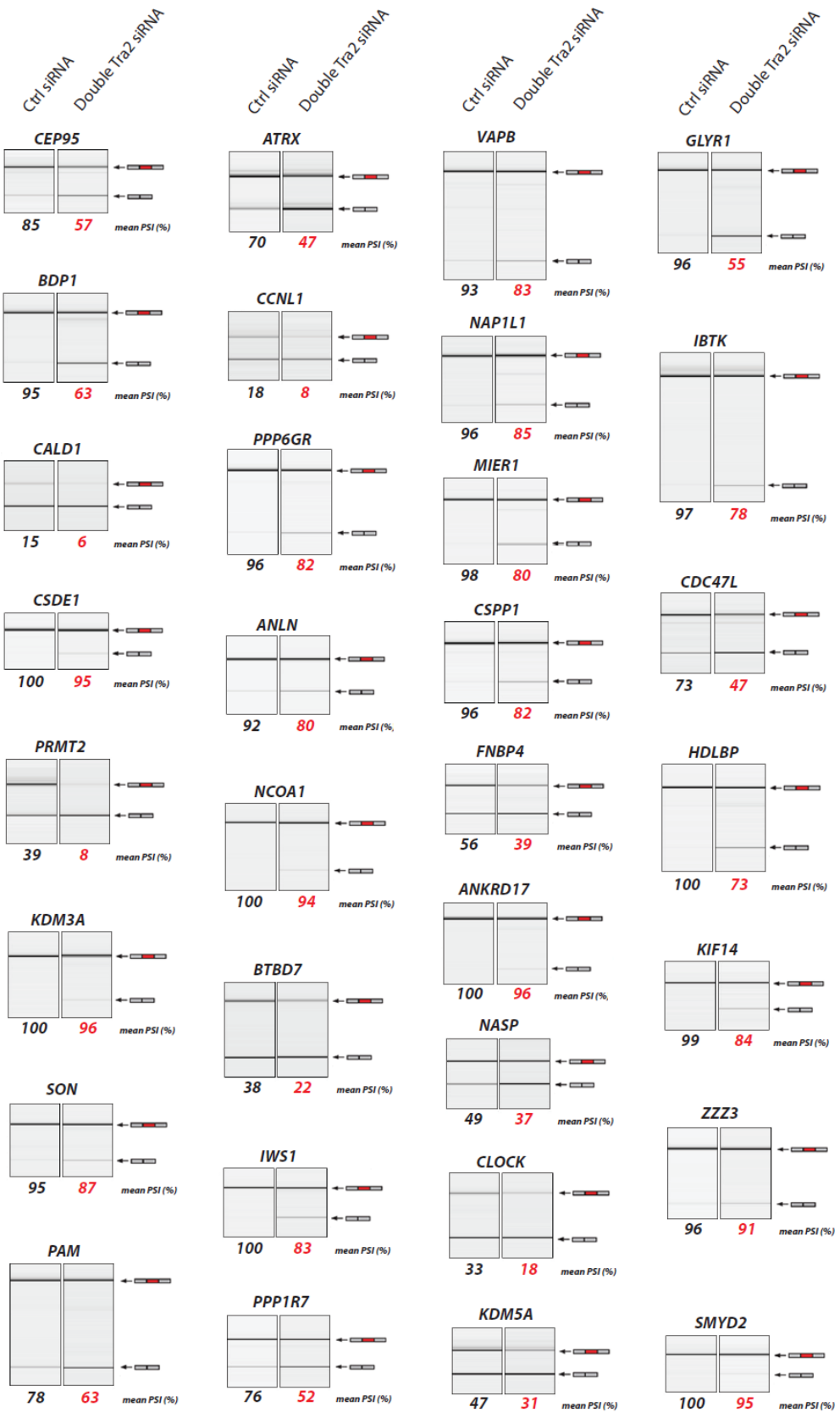


Figure 3.16 Validation of 32 Tra2 protein responsive exons by RT-PCR and capillary gel electrophoresis after joint Tra2 protein depletion. The mean PSI was calculated from three biological replicates from negative control siRNA treated cells (black) and joint *TRA2A* and *TRA2B* siRNA treated cells (red). This figure is adapted from (Best et al., 2014b).

3.4.6.3 Tra2β binding alone was insufficient to accurately predict splicing regulation

Despite successfully identifying many Tra2 responsive exons, a large panel of Tra2β bound exons did not significantly respond to either single or joint depletion of Tra2α and Tra2β. In total, 39 exons identified using iCLIP and tested by RT-PCR and capillary gel electrophoresis did not respond to joint depletion of Tra2α and Tra2β (information regarding non-responsive exons is available as supplementary data in the following publication (Best et al., 2014b). This data corroborates previous studies in which direct binding *per se* was found to be insufficient to elicit a functional response in splicing (Grellscheid et al., 2011b). The differences between functionally responsive and non-responsive exons identified by iCLIP are investigated further in chapter 4.

3.4.6.4 Intronic Tra2β binding is associated with exon repression in *CD44*

Tra2β binding was highly enriched within exons and depletion of Tra2α and Tra2β reduced the splicing inclusion on all exons tested that showed exonic binding (the Tra2 proteins therefore functioned to promote exon inclusion).

Tra2β iCLIP tags were also found to map to a variable region of the *CD44* gene. *CD44* undergoes extensive alternative splicing within a central variable region of the gene, which is flanked by two constant regions containing constitutive exons (Figure 3.17 A). Clusters of iCLIP tags were associated with the exonic sequence of variable exon 3 (v3), but unusually they mapped largely to the intronic regions surrounding variable exons 4 and 5 (v4 and v5) (Figure 3.17 B). As the variable region within *CD44* contains 10 consecutive alternative exons, it was not feasible to monitor splicing of the *CD44* variable region by conventional RT-PCR splicing analysis (using primers within flanking constitutive exons). Therefore to investigate splicing regulation of the *CD44* variable region, the relative expression of each variable exon, together with two control constitutive exons, was determined by qPCR in control siRNA treated cells and joint Tra2 protein depleted MDA-MB-231 cells.

A significant reduction in *TRA2A* and *TRA2B* mRNA was confirmed by qPCR in the *TRA2A* and *TRA2B* siRNA treated cells (red columns, Figure 3.17 C), relative to the negative control siRNA treated cells (blue columns, Figure 3.17 C). Depletion of Tra2α and Tra2β resulted in a small but statistically significant reduction in expression of variable exon 3 (v3), consistent with Tra2 proteins promoting v3 exon inclusion

through exonic binding. However, depletion of Tra2α and Tra2β led to a significant increase in expression of both the v4 and v5 exons. There was no significant change in expression of any of the other variable exons. Importantly, there was no change in expression of two control constitutive exons (exon 2 and exon 5), suggesting that the observed changes in exon expression resulted from alternative splicing, rather than changes in *CD44* gene expression.

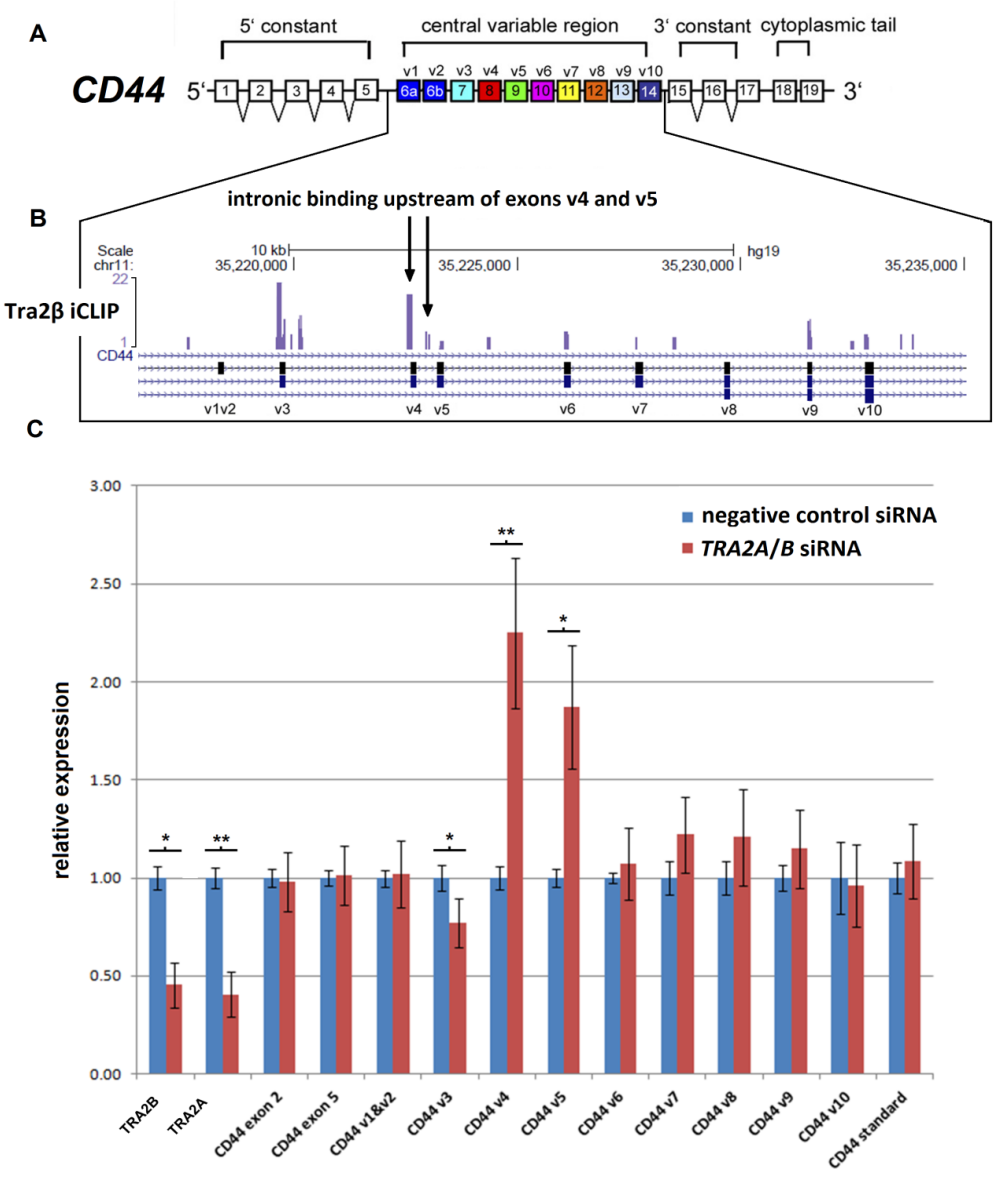


Figure 3.17 Intronic Tra2β binding was associated with repression of the variable exons v4 and v5 from *CD44*. (A) Structure of the *CD44* gene. (B) Tra2β iCLIP tags mapping to the central variable region of *CD44*. Tra2β iCLIP tags mapped directly to exon v3. Tra2β iCLIP tags also mapped to intronic regions surrounding exons v4 and v5. (C) qPCR analysis of *CD44* exon expression relative to three housekeeping genes, following depletion of Tra2α and Tra2β. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001.

3.5 Discussion

3.5.1 Tra2β efficiently suppresses endogenous Tra2α protein expression in MDA-MB-231 cells

In chapter 2, we identified a poison exon from the *TRA2A* gene which was directly regulated by Tra2β, suggesting that Tra2β might regulate protein expression of Tra2α. In concordance with that hypothesis, in this chapter I observed that the expression of Tra2α protein in MDA-MB-231 cells is normally low, but was significantly up-regulated in response to depletion of Tra2β. Conversely, knockdown of Tra2α had no significant effect on Tra2β protein expression. Knockdown of Tra2β strongly reduced splicing inclusion of the poison exon within the *TRA2A* mRNA and led to a significant increase in steady state *TRA2A* mRNA levels. This data suggested that Tra2β directly regulates Tra2α protein expression by promoting poison exon inclusion within the *TRA2A* mRNA.

NMD is a surveillance pathway conserved in all eukaryotes, which reduces errors in gene expression by targeting mRNA transcripts containing premature termination codons (PTCs) for degradation (Baker and Parker, 2004). NMD is thought to be a cellular defence mechanism, as translation of shorter mRNA transcripts could result in truncated protein isoforms which may have a detrimental gain-of-function or dominant-negative effect (Chang *et al.*, 2007). NMD also provides a mechanism for post-transcriptional regulation of gene expression, as inclusion of poison exons during unproductive splicing can also target mRNA for degradation via the NMD pathway (Ni *et al.*, 2007). DNA sequences leading to unproductive splicing are often highly conserved between mice and humans, and remarkably all genes encoding members of the SR-protein family have been found to produce non-productive mRNA isoforms via alternative splicing (Lareau *et al.*, 2007). A strikingly similar description of cross-regulation between two paralogous splicing factors was previously reported by Spellman et al., in which the splicing repressor (PTB) was found to regulate its own expression by promoting poison exon inclusion within the *PTB* mRNA, as well as the mRNA of its neuronally-restricted paralog nPTB (Spellman *et al.*, 2007). As observed in my data, a large degree of functional overlap was found between PTB and nPTB. Similarly, cross-regulation also occurs between the mammalian Fox proteins (Fox-1 (*A2BP1*), Fox-2 (*RBM9*) and Fox-3 (*HRNBP3*) (Damianov and Black, 2010).

3.5.2 Tra2β iCLIP in MDA-MB-231 cells

Cross-linking and immunoprecipitation of RNA-binding proteins coupled to deep sequencing is proving a powerful technique to study protein-RNA interactions on a transcriptome-wide scale. More recent variations of the original HITS-CLIP protocol (Ule et al., 2005b) including Photoactivatable-ribonucleoside-enhanced CLIP (PAR-CLIP) (Hafner *et al.*, 2010) and Individual-nucleotide resolution CLIP (iCLIP) (Konig *et al.*, 2011) have greatly increased the precision at which these protein-RNA interactions can be resolved on a global scale. PAR-CLIP relies on the incorporation of photoreactive ribonucleoside analogs such as 4-thiouridine (4-SU) into RNA transcripts. UV cross-linking after 4-SU incorporation induces thymidine to cytidine mutations, allowing the precise site of protein-RNA interaction to be identified. Alternatively, iCLIP captures mRNA transcripts truncated at the site of cross-linking, preserving the site of protein-RNA interaction at individual nucleotide resolution.

The Tra2β iCLIP experiments in this chapter provide the first global analysis of human Tra2β-RNA interactions. Initial visualisation of the Tra2β iCLIP autoradiographs showed a single band corresponding to just above the molecular weight of Tra2β in the high concentration RNase experiment, suggesting the Tra2β antibody was highly specific and the high stringency purification steps had isolated genuine Tra2β-RNA interactions. During the iCLIP protocol, the incorporation of randomly generated 'barcode' sequences into individual cDNA molecules overcame the problem of PCR amplification bias by allowing identification of 'unique cDNAs'. Following sequencing, non-unique cDNAs were filtered from the data set, so that only unique cDNAs were used for downstream analysis. The data generated from the iCLIP experiments was also highly reproducible and only Tra2β-RNA interactions consistent in all three biological replicate experiments were used for validation, contributing to a high quality data set.

Analysing the genome wide distribution of iCLIP tags revealed Tra2β binding was highly enriched with exons, consistent with the previous Tra2β HITS-CLIP experiment in the mouse testis (Grellscheid *et al.*, 2011a). The top 10 penatmers derived from the Tra2β iCLIP experiment were highly enriched in GAA-motifs, matching the known Tra2β binding site and serving as an early indication we had captured genuine Tra2β-RNA interactions. Tra2β iCLIP tags were also enriched within non-coding RNAs (ncRNAs), including the metastasis associated lung adenocarcinoma transcript 1 (*MALAT1*) and

nuclear paraspeckle assembly transcript 1 (*NEAT1*) (Figure 3.18). *MALAT1* is a long non-coding RNA which has recently been implicated in alternative splicing regulation, as *MALAT1* directly interacts with SR-proteins to promote phosphorylation, influencing SR-protein distribution within nuclear speckles (Tripathi *et al.*, 2010b). *MALAT1* is also over-expressed in many human cancers including breast cancer (Perez *et al.*, 2008) and prostate cancer (Ren *et al.*, 2013). *MALAT1* is also thought to have an important role in tumourigenesis (Li *et al.*, 2009) and was found to be a critical regulator of the metastatic phenotype of lung cancer cells (Gutschner *et al.*, 2013).

Similarly, the lncRNA *NEAT1* is also implicated with a role in nuclear paraspeckles, as *NEAT1* is essential to maintain the nuclear paraspeckle structure (Clemson *et al.*, 2009). This preliminary data suggests that Tra2β associates with several long ncRNAs associated with nuclear paraspeckles, including *MALAT1* and *NEAT1*. *MALAT1* and *NEAT1* are not alternatively spliced and their interaction with Tra2β may have a regulatory function, for example influencing Tra2β sub-nuclear localisation or protein turnover. Investigating the functional role of Tra2β/lncRNA interactions may prove an interesting area for future study.

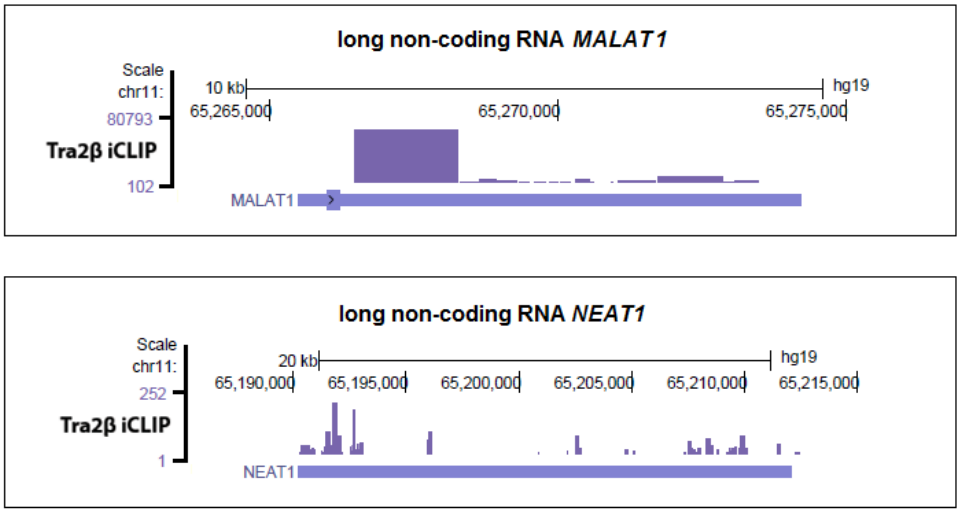


Figure 3.18 Tra2β iCLIP tags were enriched within the long ncRNAs *MALAT1* and *NEAT1*.

3.5.3 Validating splicing regulation of endogenous target exons

Splicing regulation of target exons was validated by RT-PCR and capillary gel electrophoresis in Tra2α and Tra2β depleted MDA-MB-231 cells. I initially validated 32 Tra2 protein responsive exons in this manner. Genes containing the most highly responsive exons included the α-thalassemia mental retardation X-linked gene (*ATRX)*, which encodes a member of the SWI/SNF family of chromatin remodelling proteins (De La Fuente *et al.*, 2011). Mutations in the *ATRX* gene are heavily associated with the X-linked mental retardation (XLMR) syndrome and alpha-thalassemia (ATRX) syndrome (Gibbons *et al.*, 2008). The Tra2 responsive cassette exon encoded a region close but not overlapping with the functional PHD-like domain of *ATRX* which is frequently mutated in disease (Badens *et al.*, 2006). However, the region encoded by the Tra2β responsive exon currently has no defined function. Other novel targets of Tra2β included cassette exons within *MIER1,* which encodes a transcriptional co-factor (McCarthy *et al.*, 2008), *PMRT2*, encoding a protein arginine methyltransferase (Ganesh *et al.*, 2006) and *CLOCK*, a key upstream regulator of the circadian rhythms (Alhopuro *et al.*, 2010). The general functions of Tra2 regulated genes are investigated further in chapter 4, using a Gene Ontology (GO) enrichment analysis.

3.5.4 Functional redundancy between Tra2α and Tra2β largely maintains inclusion of Tra2 target exons

To determine whether Tra2α could maintain splicing inclusion of Tra2β target exons, the splicing inclusion of 14 Tra2β target exons was measured after single depletion of either Tra2α or Tra2β, and after joint depletion of both Tra2α and Tra2β. Single depletion of either Tra2α or Tra2β had minimal effect on exon inclusion, however joint depletion of both Tra2α and Tra2β substantially reduced inclusion of all exons tested in this way. This data suggests that following depletion of Tra2β, up-regulation of Tra2α can largely maintain Tra2β target exon inclusion, a phenomenon coined “paralog compensation”. This data also demonstrates a large degree of functional redundancy between Tra2 proteins, as nearly all Tra2β target exon tested was responsive to both Tra2 proteins. A model of paralog compensation between Tra2 proteins is shown in Figure 3.19.

Although the splicing changes after single Tra2α and Tra2β depletion were small, they were often individually statistically significant (e.g. in the *ATRX* gene). This fine tuning of splicing profiles by joint Tra2 protein concentration might have physiological significance in whole organisms and at particular points of development (e.g. in brain development). Moreover, the physiological effect of even individually small splicing defects might be cumulative over hundreds of Tra2β target exons. This may explain why *Tra2b* knockout mice are embryonic lethal despite expressing the *Tra2a* gene (Mende *et al.*, 2010a; Grellscheid *et al.*, 2011a; Roberts *et al.*, 2013). Another possibility is there are some specific cell types in which *Tra2b* is not expressed.

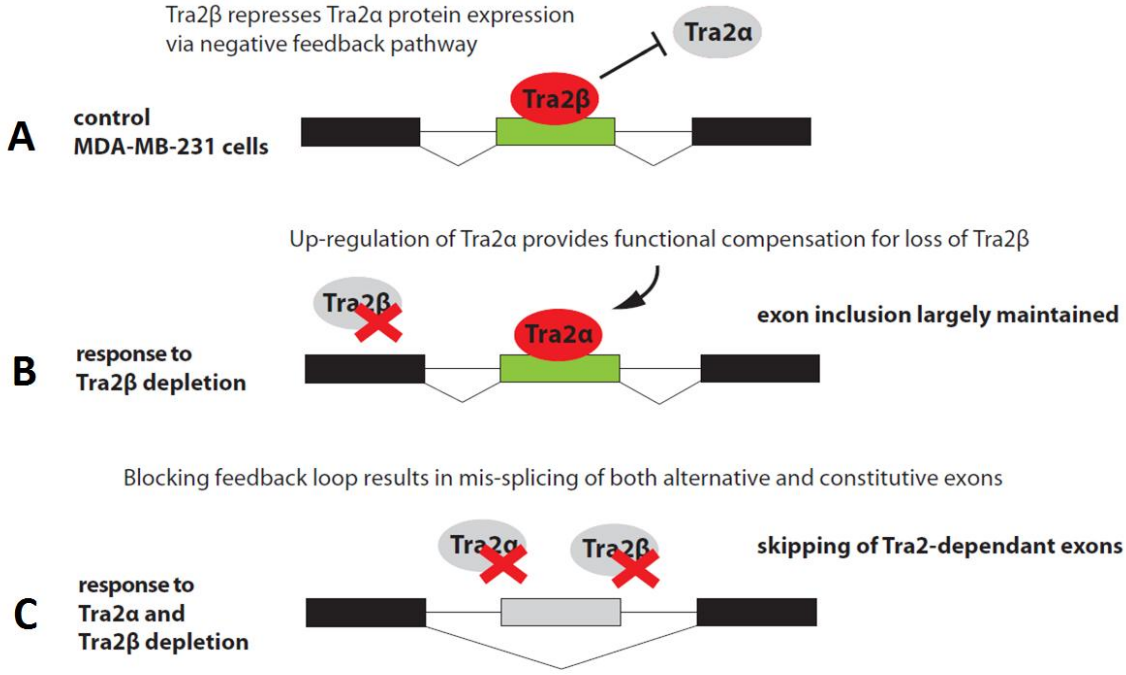


Figure 3.19 A model of paralog compensation between Tra2α and Tra2β, which largely maintains splicing inclusion of Tra2 target exons. (A) Tra2β efficiently represses expression of Tra2α protein in MDA-MB-231 cells, but is sufficient to maintain inclusion of target exons. As Tra2α expression is very low, depletion of Tra2α protein has minimal effect on inclusion of target exons. (B) Depletion of Tra2β significantly reduces inclusion of a poison exon within the *TRA2A* mRNA, increasing Tra2α protein expression. Up-regulation of Tra2α protein can functionally compensate for loss of Tra2β, largely maintaining splicing inclusion of target exons. (C) Joint depletion of both Tra2α and Tra2β is required to induce a substantial reduction in splicing inclusion of Tra2 target exons.

3.5.5 Intronic binding is associated with exon repression

Tra2β was previously reported to regulate alternative splicing of the *CD44* gene. Co-transfection with Tra2β was found to enhance inclusion of *CD44* variable exons v4 and v5, which were cloned into a minigene and transfected into HeLa cells (Watermann et al., 2006). To investigate alternative splicing regulation of endogenous *CD44*, I analysed iCLIP tags mapping to the *CD44* gene. Multiple clusters of Tra2β iCLIP tags mapped to the central variable region of *CD44*, which contains 10 consecutive alternatively spliced exons (Figure 3.17 B). To investigate splicing regulation of this central variable region, expression of each exon was quantified by qPCR after joint depletion of Tra2α and Tra2β compared to negative control siRNA treated cells. The largest cluster of Tra2β iCLIP tags was associated with exon v3, and joint depletion of Tra2α and Tra2β significantly reduced variable exon v3 expression as predicted. However, joint depletion of Tra2α and Tra2β unexpectedly led to a significant increase in expression of variables exons v4 and v5 (Figure 3.17 C). Joint depletion of Tra2α and Tra2β had no effect on expression of the control constitutive exon 2 and exon 5, suggesting the changes in expression of exons v3, v4 and v5 were due to alternative splicing, rather than changes to *CD44* gene expression. There was no significant change in any of the other variable exons. Unusually, the cluster of Tra2β iCLIP tags associated with exon v4 was entirely intronic and mapped within the first 50bp upstream of the exon (see Figure 3.20 A).

RNA splicing maps have revealed that many RNA-binding proteins share many common positional principles in splicing regulation (Darnell, 2010). RNA-binding proteins including Nova, hnRNP C, Fox, PTB and Mbnl1 have all been found to silence exon inclusion by binding upstream of exons, within close proximity to the branch point and 3' splice sites (Witten and Ule, 2011). Assuming this is a genuine site of protein-RNA interaction, Tra2 proteins may follow a similar principle. One possible model is that Tra2α and Tra2β may inhibit inclusion of variable exons v4 and v5 by preventing assembly of the U2 snRNP/U2AF complex which is essential for exon recognition (Figure 3.20 B).

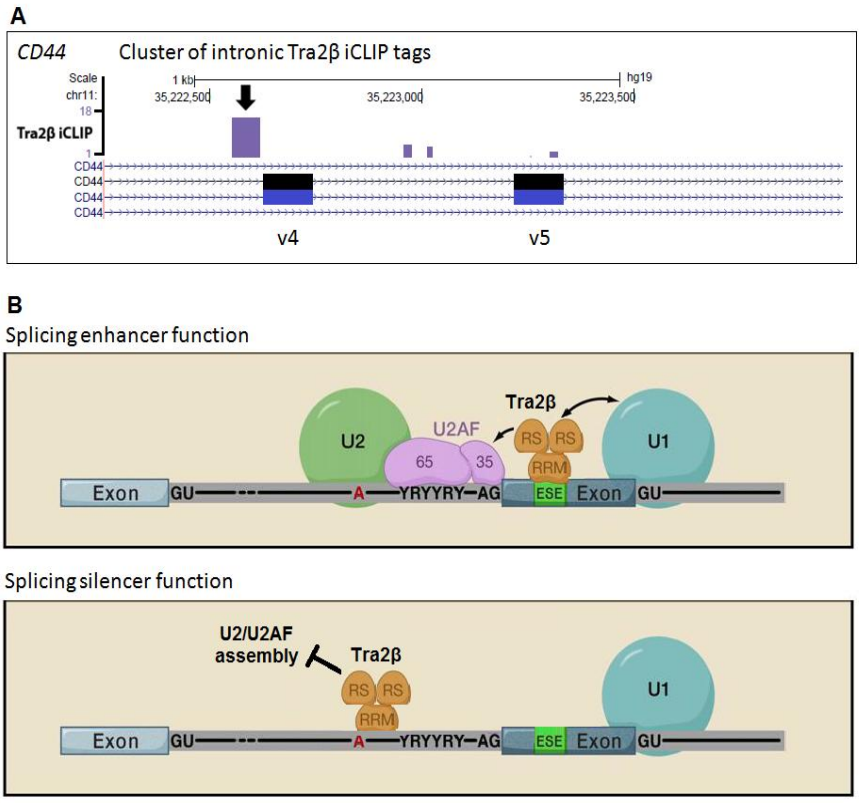


Figure 3.20 Model of Tra2β silencing of *CD44* exons v4 and v5. (A) Clusters of Tra2β iCLIP tags mapped immediately upstream of variable exon 4. (B) Model of Tra2β splicing regulation. Typically, exonic binding by SR proteins enhances exon inclusion by stabilising components of the spliceosome. Conversely, binding immediately upstream of the variable exon 4 may inhibit assembly of the U2 snRNP/U2AF complex, inhibiting exon inclusion. Figure 3.20 part B is adapted from (Wahl et al., 2009).

However, it is also possible that changes to v4 and v5 expression are indirect consequences of joint depletion of Tra2α and Tra2β. The precise mechanism of exon repression may be investigated in future work and may benefit from a more detailed dissection using minigenes. On the surface, the data from Watermann et al. and my own data appear to be contradictory, as over-expression of Tra2β was found to enhance exon inclusion in their model, whilst joint depletion of Tra2α and Tra2β also enhanced exon inclusion in mine. These inconsistencies may be due to the different nature of experimental procedures used. Watermann et al. use a minigene model containing a short region of *CD44* transfected in HeLa cells and over-expression of Tra2β. I investigated splicing of endogenous *CD44* in MDA-MB-231 cells, following depletion of endogenous Tra2α and Tra2β. For that reason, Tra2β mediated splicing regulation of the *CD44* exons v4 and v5 requires further investigation to resolve these differences.

3.5.6 Chapter Summary

In this chapter, I established that Tra2β efficiently suppresses Tra2α protein expression in MDA-MB-231 cells, most likely by promoting inclusion of a poison exon within the *TRA2A* mRNA. Cross-regulation of expression was found to be largely asymmetrical in MDA-MB-231 cells.

I also mapped the transcriptome-wide binding sites of Tra2β in MDA-MB-231 cells using iCLIP. This facilitated the identification of Tra2β target exons which I functionally validated by RT-PCR in Tra2α and Tra2β depleted cells. Joint Tra2 protein depletion (but not single depletion of either Tra2 protein) was required to observe robust changes in splicing. This demonstrated a large degree of functional redundancy between the Tra2 proteins and suggested up-regulation of Tra2α (following Tra2β depletion) could largely maintain Tra2 target exon inclusion, a phenomenon termed "paralog compensation".

Tra2β was largely found to activate splicing inclusion of the newly identified target exons through exonic binding. However, intronic binding was associated with skipping of two variable exons (v4 and v5) from *CD44*. Further study is required to establish whether this is a direct or indirect consequence of Tra2 protein depletion.

Chapter 4: Transcriptome-wide investigation of Tra2 protein dependent splicing using RNA-seq

4.1 Introduction

The data from chapter 3 suggested that due to a negative feedback pathway between Tra2 proteins in MDA-MB-231 cells, robust splicing changes to Tra2 protein dependent exons would require joint depletion of both Tra2α and Tra2β. Spellman et al. previously described similar asymmetric expression patterns and cross-regulation of expression between members of the PTB family (2007). Comprehensive identification of PTB target RNAs also required joint depletion of PTBP1 and PTBP2 (Spellman et al., 2007). I largely detected low amplitude splicing switches in the target exons identified using iCLIP in chapter 3 and some exons were completely unaffected by joint Tra2 protein depletion. This is consistent with previous findings that evidence of direct binding *per se* is often insufficient to accurately predict functionally responsive exons (Grellscheid et al., 2011b). Consequently in this chapter, I have used RNA-seq following joint Tra2 protein depletion, with the aim of identifying higher amplitude splicing switches to functionally responsive exons in MDA-MB-231 cells. By combining the RNA-seq data with the Tra2β iCLIP data from chapter 3, my objective was to identify target exons which were both directly bound by Tra2β and functionally responsive to joint Tra2 protein depletion (Figure 4.1).

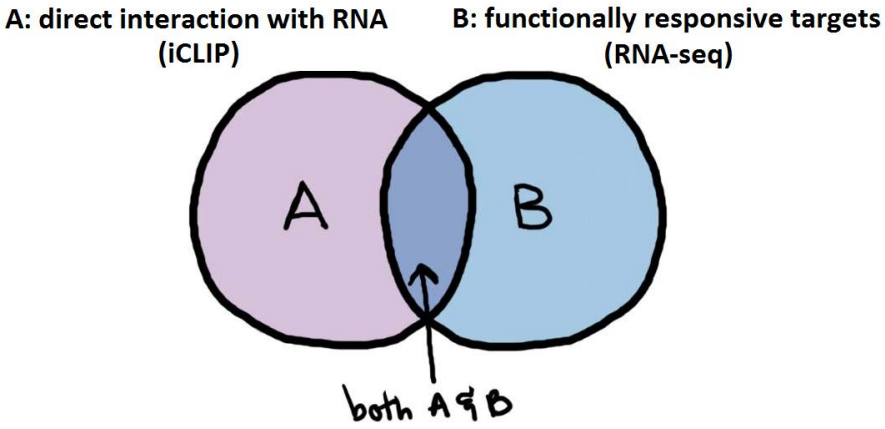


Figure 4.1 Experimental strategy used to identify Tra2 dependent target exons which were both directly bound by Tra2β (iCLIP) and functionally responsive to joint Tra2 protein depletion (RNA-seq). This image is adapted from (Hsu, 2008).

Tra2α and Tra2β share some degree of structural and functional similarity to the classical SR-protein family which regulate both constitutive and alternative splicing (Zhou and Fu, 2013). However Tra2α and Tra2β have been implicated exclusively in alternative splicing regulation and not constitutive splicing. More over, only classical SR-proteins can provide splicing activity to splicing deficient S100 extracts (Tacke *et al.*, 1998b). These data suggest there may be fundamental differences between the classical SR-protein family and Tra2 proteins in splicing activity. Hence in this chapter, I also use the RNA-seq data to investigate whether Tra2 proteins play a role in regulation of constitutively spliced target exons in addition to their known role in alternative splicing.

Some splicing factors have been found to regulate genes involved in functionally coherent biological processes. Recent examples include Nova (which regulates alternative splicing of genes encoding synaptic proteins and proteins involved in axon guidance) (Ule et al., 2005c), PTB (which regulates alternative splicing of genes encoding cytoskeletal proteins) (Boutz et al., 2007), T-STAR (which regulates alternative splicing of a family of genes encoding pre-synaptic proteins) (Ehrmann et al., 2013) and MBNL1/RBFOX2 (which regulate alternative splicing of genes involved in pluripotent stem cell differentiation) (Venables et al., 2013b). To examine whether Tra2 proteins similarly regulate splicing of genes involved in functionally related biological processes, I also use Gene Ontology (GO) enrichment analysis in collaboration with Dr. Katherine James (Newcastle University) to test for functional enrichment in Tra2 protein regulated genes.

In the final part of this chapter, I focus on characterising splicing regulation of a functionally important exon from *CHEK1*. *CHEK1* encodes the serine/threonine protein kinase CHK1, which plays a key role in the DNA damage response and cell cycle checkpoint control (Gagou *et al.*, 2010; Pabla *et al.*, 2012). I also investigate whether Tra2 proteins are required for expression of the full-length CHK1 protein, and the ensuing cell phenotype following joint Tra2 protein depletion, specifically focusing on DNA damage and cell viability.

4.2 Aims

The aims of this chapter were to:

1. Investigate changes to the MDA-MB-231 cell transcriptome following joint Tra2 protein depletion using RNA-seq
2. Investigate whether Tra2 proteins play a role in regulation of constitutively spliced target exons
3. Investigate whether Tra2 proteins regulate splicing of genes involved in functionally related biological processes
4. Characterise and investigate functional consequences of splicing regulation of the *CHEK1* gene by Tra2 proteins

4.3 Materials and Methods

4.3.1 RNA-seq

RNA was extracted from MDA-MB-231 cells using an RNeasy Plus Mini Kit (Qiagen) and re-suspended in nuclease-free water following the manufacturer's instructions. The RNA samples were DNase-treated using a DNA-free kit (Invitrogen) and stored at -80°C. The RNA quality was determined using a 2100 Agilent Bioanalyser to monitor RNA degradation. Only RNA samples with an RNA integrity score (RIN) greater than nine were selected for sequencing. RNA samples were sent on dry ice to the Bristol University Genomics Facility for sequencing, where mRNA libraries were prepared using a TruSeq mRNA library kit (Illumina). Paired-end sequencing was performed for six RNA samples (three biological replicates from negative control siRNA treated MDA-MB-231 cells and three biological replicates from TRA2A/B siRNA treated MDA-MB-231 cells) using an Illumina HiSeq 2000 sequencing system.

The RNA-seq data was processed and analyzed by Mr Yaobo Xu (Newcastle University). Mr Yaobo Xu analyzed the RNA-seq data to identify differentially expressed genes, differentially expressed exons and exons which have differential usages among transcripts of a gene. The quality of sequencing reads was first checked using FastQC (Andrews, 2014). Poly-N tails were trimmed from reads with an in house perl script. The 13bp on the left ends of all reads were clipped off with Seqtk (Li, 2014) to remove biased sequencing reads caused by random hexamer priming (Hansen *et al.*, 2010). Low quality bases (Q < 20) and standard Illumina (Illumina, Inc. California, U.S.) paired-end sequencing adaptors on 3' ends of reads were trimmed off using Trim-galore (Krueger) and only those that were at least 20bp in length after trimming were kept. The high quality reads were then mapped to the human reference genome hg19 with Tophat2 (Kim *et al.*, 2013). Reads aligned to genes and exons were counted with Bedtools (Quinlan and Hall, 2010). Differentially expressed genes and exons were then identified with the Bionconductor (Gentleman et al., 2004) package DESeq (Anders and Huber, 2010). Differentially used exons were identified with Bionconductor package DEXSeq (Anders *et al.*, 2012).

To search for candidate exons, the processed RNA-seq data was uploaded to the UCSC genome browser, so that the Tra2β iCLIP tracks and RNA-seq tracks could be viewed in

parallel. Genes containing differentially expressed exons that were statistically significant were ranked in order of the greatest fold-change in exon expression and filtered for a minimum level of gene expression. Differentially expressed exons containing Tra2β iCLIP tags were selected for downstream validation by RT-PCR and capillary gel electrophoresis.

4.3.2 Splicing analysis

A standard splicing assay is used throughout this chapter and was described previously in chapter 2, methods section 2.3.9. Briefly, the PSI of endogenous target exons was determined using RT-PCR and capillary gel electrophoresis. PCR primers were designed within flanking constitutive exons and the PSI of target exons was determined from three biological replicates per experiment. In this chapter, the PSI of target exons was analysed from negative control siRNA treated cells, TRA2A siRNA treated cells, TRA2B siRNA treated cells, and TRA2A and TRA2B (TRA2A/B) siRNA treated cells. The standard siRNA transfection protocol used in this chapter was also described previously in chapter 2, methods section 2.3.7. All primer sequences used for splicing assays in this chapter are provided in Appendix A.

4.3.3 Molecular cloning and mutagenesis of the *ANKRD1* minigene

The *ANKRD1* consitutive exon and approximately 200 nucleotides of upstream and downstream flanking intronic sequence was amplified from human genomic DNA using two cloning primers (Table 4.1). A restriction digest was performed on the gel purified PCR product using the restriction endonuclease *Eco*RI. The digested insert was cloned into an *Mfe*1 restriction site within the pXJ41 minigene vector. Single base mutations to Tra2β binding sites were introduced into the wildtype *ANKRD1* minigene using site directed mutagenesis (as described previously in chapter 2.3.4). The molecular cloning and mutagenesis of the *ANKRD1* minigene was carried out by Ms. Marina Danilenko (Newcastle University). The primers used for site directed mutagenesis of the wildtype *ANKRD1* minigene are provided in Table 4.1. Different versions of the *ANKRD1* minigene were co-transfected into HEK-293 cells for splicing analysis as previously described in chapter 2.3.2. Inclusion of the various *ANKRD1* minigene exons was determined by RT-PCR using a Qiagen One-step RT-PCR kit (Qiagen) and the pXJ41-specific primers pXJRTF and pXJB1 (Table 4.1).

Primer	Sequence
ANKRD1 cloning primer F	AAAAAAAAAGAATTCAAAATCTAAGACTTGCTTATGGCATT
ANKRD1 cloning primer R	AAAAAAAAAGAATTCAGCATGAGAGTTACCGTGAGC
ANKRD1 M1F	AGAACACATATCAAAGCTTGCACATTTATACGACCTTGAAA
ANKRD1 M1R	CAAGGTCGTATAAATGTGCAAGCTTTGATATGTGTTCTAG
ANKRD1 M2F	ATCATTCAACTGCAGCAACGGCAACAATACAGGCACACTAAAG
ANKRD1 M2R	GAACTTTAGTGTGCCTGTATTGTTGCCGTTGCTGCAGTTGAATG
pXJRTF	GCTCCGGATCGATCCTGAGAACT
pXJB1	GCTGCAATAAACAAGTTCTGCT

Table 4.1 Primers sequences used for molecular cloning and mutagenesis of the *ANKRD1* minigene.

4.3.4 High throughput RT-PCR screen

The PSI of four endogenous target exons was monitored using a custom high-throughput RT-PCR plate containing cDNAs from MDA-MB-231 cells in which 53 known splicing regulators had been individually and systematically depleted using RNAi (Huang et al., 2013). Six cDNA samples (three biological replicates from negative control siRNA treated cells and three biological replicates from TRA2A/B siRNA treated cells), together with splicing assay primers for *ANKRD1*, *CHEK1*, *GLYR1* and *SMC4* (Table 4.2) were sent to the laboratory of Professor Benoit Chabot (Université de Sherbrooke, Quebec, Canada) for analysis. The high-throughput RT-PCR screen was performed by Dr. Roscoe Klinck (Université de Sherbrooke, Canada). Only samples with a minimum of 50% mRNA depletion determined by qPCR were included in this panel (Venables et al., 2013b).

Primer	Sequence
ANKRD1 F	GGAAAGAAGAATGGCAATGG
ANKRD1 R	GCAGCCTTCAGAAACGTAGG
CHEK1 F	GACTGGGACTTGGTGCAAAC
CHEK1 R	TGCCATGAGTTGATGGAAGA
GLYR1 F	CTAGACTCCGGGATGGTGAG
GLYR1 R	GCCTGGATCAAAGTGGAACA
SMC4 F	TGGATGTAGCCCAGTCAGAA
SMC4 R	CCAGTGCATGACAACAGGAT

Table 4.2 Primers sequences used in a high throughput RT-PCR screen of four endogenous target exons after individual depletion of 53 known splicing regulators.

4.3.5 Gene Ontology (GO) enrichment analysis and interaction network

The Gene Ontology (GO) (R Foundation for Statistical Computing) enrichment analysis was performed entirely by Dr. Katherine James (Newcastle University) using the Bioconductor GOstats package version 2.24.0 ((Gentleman, 2004; Falcon and Gentleman, 2007; Hogan *et al.*, 2008; Langmead *et al.*, 2009). Enrichments of GO biological process terms were calculated using the conditional hypergeometric test with a significance cut-off of 0.001 and using a background of genes normally expressed in MDA-MB-231 cells. Annotations were taken from the Bioconductor *Homo sapiens* annotation package org.Hs.eg.db version 2.8.0 (Carlson, 2012). The analysis was run in the open source statistical package R version 3.0.1 (R Foundation for Statistical Computing).

The protein interaction network was also generated by Dr. Katherine James (Newcastle University). Interaction data for *Homo sapiens* was retrieved from the BioGRID database (version 110). These data were integrated into a network in which nodes represented genes or gene products, and edges represented any type of BioGRID interaction between the nodes. The network was visualised using the Cytoscape visualisation platform (Shannon et al., 2003) and was coloured based on annotations to top five enriched GO biological processes (as downloaded from QuickGO (Shannon et al., 2003). Where a protein was annotated to more than one term, the most specific annotation was chosen.

4.3.6 Quantitative real time PCR (qPCR)

Quantitative real-time PCR (qPCR) was performed as described in chapter 3, section 3.3.3. Gene expression was compared between three biological replicates of negative control siRNA treated MDA-MB-231 cells and three biological replicates from *TRA2A* and *TRA2B* siRNA treated MDA-MB-231 cells. Three technical replicates were performed per biological sample. A no template control (NTC) and a no reverse transcriptase control (-RT) were included for each set of primers. Gene expression was calculated relative to three housekeeping genes *ACTB*, *GAPDH* and *TUBB*. Ct values for each sample were calculated using SDS 2.4 software (Applied Biosystems) and relative mRNA expression was calculated using the $2^{-\Delta\Delta Ct}$ method. All primers used for quantitative real-time PCR (qPCR) in chapter 4 are provided in Table 4.3.

Primer	Sequence
TRA2A F	TCCAATGTCTAACCGGAGAAG
TRA2A R	CCAAACACTCCAAGGCAAGT
TRA2B F	CGCCAACACCAGGAATTTAC
TRA2A R	TCATAGCCCCGATCATATCC
SELPLG F	GGCTGGGACCTTGTCACTAA
SELPLG R	AACAGGAGGAGTTGCAGAGG
FOSB F	CTCAATATCTGTCTTCGGTGGA
FOSB R	GGTCCTGGCTGGTTGTGAT
FOS F	CCGGGGATAGCCTCTCTTAC
FOS R	GTGGGAATGAAGTTGGCACT
DDIT4 F	CCTGGACAGCAGCAACAGT
DDIT4 R	GAAGTCGGGCAACGACAC
EGR1 F	AGCCCTACGAGCACCTGAC
EGR1 R	GTCTCCACCAGCACCTTCTC
TTC14 F	CAATCCACACTTCCGTAGCC
TTC14 R	AATTTCTGTATCTCGATGTTGTCA
RP11-618G20.2 F	AGAATGGAGCCAAGACAAGG
RP11-618G20.2 R	GCAGAACATATCAGCAGAAGAA
RP5-1129J21 F	GGAGCCACCGTTTTGTGC
RP5-1129J21 R	GACTTATTTTCCTTTGCCTGGA
CTD-2037K23 F	CTTCCCCTCCGCTCTCAG
CTD-2037K23 R	TCTCCGTGATTTAGCAGCAA
ANKRD26 F	TCTGACCAATCACCTTTCCA
ANKRD26 R	CCAGTCTTTCCTTGCTTTGC
EIF5B F	GACAAGCAATGGGGAAGAAA
EIF5B R	GCCAAGGCATCAAGATCAAT
HIST2H2BE F	CCAGAAGAAAGACGGCAAGA
HIST2H2BE R	ATGGCCTTGGACGAGATG
RP11-221J22 F	CCCACATTAACTCATCCATCC
RP11-221J22 R	ATTCAGACAAGCCCTGGTTG
LINC00601 F	AGTAATGGGACACGCAGACC
LINC00601 R	TCAATATCTTCCTTCTTCCTCTTCA
ACTB F	CATCGAGCACGGCATCGTCA
ACTB R	TAGCACAGCCTGGATAGCAAC
GAPDH F	AACAGCGACACCCATCCTC
GAPDH R	CATACCAGGAAATGAGCTTGACAA
TUBB F	CTTCGGCCAGATCTTCAGAC
TUBB R	AGAGAGTGGGTCAGCTGGAA

Table 4.3 Primers used for quantitative real-time PCR (qPCR).

4.3.7 siRNA transfection

In this chapter, siRNA transfections were performed in MDA-MB-231, MCF-7, HEK-293, HeLa and PC3 cells following the standard siRNA transfection protocol described in chapter 2, methods section 2.3.7. CHK1 protein depletion was performed using a *CHEK1*-specific siRNA (Ambion Silencer Select siRNA ID: s503).

4.3.8 Electrophoretic Mobility Shift Assays (EMSAs)

Three short regions (68-69bp) from the human *CHEK1* gene were cloned in front of a T7 promoter within the pBluescript vector, to facilitate the *in vitro* transcription of three short *CHEK1* RNA probes. The primers used to clone the short sequences from *CHEK1* into the pBluescript vector are provided in Table 4.4. I performed the molecular cloning, whilst the Electrophoretic Mobility Shift Assays (EMSAs) were performed by Mrs. Caroline Dalgliesh (Newcastle University). The EMSAs were carried out using purified full-length Tra2β protein and *in vitro* transcribed, radio-labelled RNA probes generated from the *CHEK1* pBluescript constructs containing amplified regions of the human *CHEK1* gene. The EMSA experiments were performed as previously described (Grellscheid *et al.*, 2011a; Grellscheid *et al.*, 2011b).

Primer	Sequence
CHEK1 intron F	AAAAAAAAAGGTACCTGTGTACCTCTCCTTCACTACC
CHEK1 intron R	AAAAAAAAAGAATTCCTGTCCTAAGCTCCTATGGGG
CHEK1 exonA F	AAAAAAAAAGGTACCGTTCAACTTGCTGTGAATAGAGT
CHEK1 exonA R	AAAAAAAAAGAATTCGGCACGCTTCATATCTACAATCT
CHEK1 exonB F	AAAAAAAAAGGTACCAGTAAAATTCTATGGTCACAGGA
CHEK1 exonB R	AAAAAAAAAGAATTCCTCCACTACAGTACTCCAGAAAT

Table 4.4 Primer sequences used to clone short sequences from the *CHEK1* gene into the pBluescript vector.

4.3.9 Molecular cloning of the wildtype and mutant *CHEK1* minigenes

Exon 3 of the *CHEK1* gene and approximately 250 nucleotides of the upstream and downstream flanking intronic sequences was synthesised by gBlocks gene fragments (IDT UK) and subsequently cloned into the pXJ41 minigene vector. A mutated version of the *CHEK1* exon 3 was also synthesised and cloned into the pXJ41 vector, which contained single base mutations within the exonic Tra2β binding sites. The wildtype and mutant *CHEK1* exon 3 sequences are provided in Figure 4.19. Cloning of the

wildtype and mutant *CHEK1* exon 3 minigenes was carried out by Ms. Marina Danilenko (Newcastle University). The *CHEK1* exon 3 minigenes were co-transfected into HEK-293 cells for splicing analysis as described previously in chapter 2, methods section 2.3.2.

4.3.10 Centrifugal elutriation and cell cycle evaluation

Frozen KG-1 cell pellets fractionated by centrifugal elutriation were a gift from Professor Caroline Austin (Newcastle University). The KG-1 cells were subject to centrifugal elutriation and cell cycle evaluation, performed by Dr. Ian Cowell (Newcastle University) and Mr. Ka Cheong Lee (Newcastle University). Briefly, KG-1 cells were size fractionated by centrifugal elutriation, using flow rates of 10, 13, 17, 20, 24 and 28 ml/min (Ly, 2014). Cell cycle phase enrichment of cells was assessed using immunofluorescence staining for CENPF (late S, G_2, G_2/M) (Liao et al., 1995) and phosphorylated histone H3S10 (G_2/M, M) (Hendzel MJ1, 1997). Asynchronous and elutriated KG-1 cells were suspended in PBS, spotted onto poly-lysine-coated slides and processed for immunofluorescence. Images were captured and cells were scored for CENPF and phospho-H3S10 staining. RNA was purified by Trizol RNA extraction. cDNA was generated using a SuperScript VILO cDNA synthesis kit (Invitrogen) following the maufacturer's instructions. The PSI of endogenous *CHEK1* exon 3 was determined by RT-PCR and capillary gel electrophoresis. Primers used for the *CHEK1* exon 3 splicing analysis are provided in Table 4.5.

Primer	Sequence
CHEK1 F	GACTGGGACTTGGTGCAAAC
CHEK1 R	TGCCATGAGTTGATGGAAGA

Table 4.5 Primers sequences used for the *CHEK1* exon 3 splicing assay.

4.3.11 Breast cancer tumour biopsies

The splicing profile of endogenous *CHEK1* exon 3 was monitored by RT-PCR and capillary gel electrophoresis in a panel of ten breast cancer tumour biopsies. The cDNA samples from ten breast cancer tumour biopsies were a gift from Dr. Alison Tyson-Capper (Newcastle University) and were originally obtained with appropriate consent from the Breast Cancer Campaign Tissue Bank.

4.3.12 Western Immunoblotting

Protein expression was detected by Western Immunoblotting using the following primary antibodies and dilutions: Tra2α (Novus Biologicals, H00029896-B01P) (1:500 dilution), Tra2β (Abcam, ab31353) (1:2000 dilution), CHK1 (Proteintech, 10362-1-AP) (1:250 dilution), Histone H2AX (Santa Cruz Biotechnology, sc54-606) (1:500 dilution), phosphorylated Histone H2AX (Ser 139) (Santa Cruz Biotechnology, sc-101696) (1:500 dilution), FLAG (Sigma Aldrich, F3040) (1:2000 dilution) and α-Tubulin (Sigma-Aldrich, T5168) (1:2000 dilution).

4.3.13 Immunofluorescent detection of γH2AX

The expression and localisation of γH2AX within MDA-MB-231 cell nuclei was investigated using immunofluorescence. The expression of γH2AX was compared between negative control siRNA treated cells and TRA2A/B siRNA treated cells. MDA-MB-231 cells were seeded onto sterile cover slips and immediately transfected with siRNA, following the standard siRNA transfection protocol described in chapter 2.3.7. After 72 hours, the siRNA treated cells were fixed by removing the media, washing gently with PBS and incubating the cover slips with 0.5ml 4% PFA for 20 minutes at room temperature. The MDA-MB-231 cells were permeabilised by incubating the cover slips with 0.1% Triton X-100 in PBS for 10 minutes at room temperature. The cover slips were blocked with blocking solution (10% FBS in PBS) for 10 minutes at room temperature. Cover slips were incubated with a 1:100 dilution of primary antibody against γH2AX (Santa Cruz Biotechnology, sc-101696) in blocking solution for 2 hours at room temperature. Cover slips were washed gently three times with PBS to remove residual antibody. Cover slips were incubated with a 1:400 dilution of goat anti-rabbit IgG Alexa Fluor 488 fluorescent secondary antibody (Abcam, ab150077) in blocking solution for one hour, protected from light. Finally, cover slips were gently washed three times with PBS and mounted onto slides using VECTASHIELD Mounting Medium with DAPI (Vector Labs, H-1200). Slides were analysed using a ZEISS Axioplan2 fluorescent microscope with Axiovision software. An optimal exposure time was determined for detection of the Alexa Fluor 488 fluorescent secondary antibody and the same exposure time was used for image analysis of all slides.

4.3.14 MTT assay

An MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay was used to monitor viability of MDA-MB-231 and HEK-293 cells following siRNA transfection. The MTT assays were performed using an MTT Cell Proliferation Assay Kit (Cayman Chemical) following the manufacturer's instructions. A standard siRNA transfection was performed as described in chapter 2, methods section 2.3.7. A 100µl aliquot of cell suspension containing approximately $2X10^3$ cells in complete growth media was added to each well in a 96-well plate. The absorbance from the MTT assay was measured over 5 days at 24h, 48h, 72h, 96h and 120h following siRNA transfection. The relative density of cells was also observed by microscopy 120h after siRNA transfection.

4.3.15 Flow cytometry analysis of EdU incorporation

MDA-MB-231 cells were incubated with 10µM EdU for 4 hours, 96 hours after siRNA transfection. Cell fixation, permeabilization and EdU detection was performed using the Click-iT EdU Flow Cytometry Assay Kit (Life technologies) following the manufacturer's instructions. Data was collected and analysed using a BD LSR II flow cytometer using 488nm excitation and a 520/20 band pass for detection of EdU Alexa Fluor 488 azide and 355nm excitation and a 450/50 band pass for detection of DAPI. Experiments were performed using triplicate biological samples and 30,000 cells were analysed per sample. A negative control siRNA transfected no EdU control sample was used to inform the gating strategy used to calculate the proportion of EdU-positive cells.

4.3.16 Generation of a tetracycline inducible full-length CHK1-FLAG FLP-in HEK-293 cell line

Full-length CHK1-FLAG cDNA was amplified from the pcDNA4-Chk1-Flag plasmid (Addgene plasmid #22894) using the primers provided in Table 4.6. The insert was subsequently cloned into the Flp-In expression vector pcDNA5. To generate a tetracycline inducible cell line, the CHK1-FLAG-pcDNA5 vector was co-transfected with the Flp recombinase plasmid (pOG44) into Flp-In HEK-293 cells. Transfected cells were selected for Hygromycin B resistance by incubating cells with 50µg/ml Hygromycin B for a 2-3 weeks period until all untransfected control cells had died. Following

Hygromycin B selection, CHK1-FLAG expression was induced by treating cells with 2µg/ml tetracycline to induce CHK1-FLAG expression via a tetracycline inducible promoter. The molecular cloning of the CHK1-FLAG-pcDNA5 plasmid, plasmid transfections and antibiotic selection of Flp-in HEK-293 cells was performed by Dr. Mahsa Kheirollahi-Kouhestani (Newcastle University).

Efficient expression of the full-length CHK1-FLAG protein was detected by Western Immunoblotting using a FLAG-specific primary antibody (Sigma Aldrich, F3040) (1:2000 dilution). An MTT assay was used (as previously described in section 4.3.13) to monitor viability of the full-length CHK1-FLAG Flp-In HEK-293 cell line following siRNA transfection. Cell viability was monitored for cells treated with 2µg/ml tetracycline and mock-treated cells. Tetracycline was added every 24 hours, beginning 24 hours after the initial siRNA transfection.

Primer	Sequence
CHEK1 FLAG F	AAAAAAAAAGCGGCCGCATGGCAGTGCCCTTTGTGGAAGAC
CHEK1 FLAG R	AAAAAAAAAGTCGACTCATGTGGCAGGAAGCCAAATCTTC

Table 4.6 Primers sequences used to amplify full-length CHK1-FLAG cDNA from the pcDNA4-Chk1-Flag plasmid.

4.4 Results

4.4.1 The *TRA2A* and *TRA2B* genes are differentially expressed in MDA-MB-231 cells

When comparing the RNA-seq data from control MDA-MB-231 cells, I observed that the *TRA2B* gene is expressed at significantly higher levels than the *TRA2A* gene (see Figure 4.2, RNA-seq reads are shown in green and the Y-axis represents read depth which is indicative of relative gene expression). Noticeably, the only significant clusters of Tra2β iCLIP tags which mapped to the human *TRA2B* and *TRA2A* genes were located within their respective poison exons (Figure 4.2, iCLIP sequencing reads are shown in orange). Despite significantly lower expression of the *TRA2A* gene in the control MDA-MB-231 cells, the *TRA2A* poison exon had a very similar density of Tra2β iCLIP tags as the *TRA2B* poison exon. This suggested that the *TRA2A* poison exon might be a stronger physiological target for Tra2β than the *TRA2B* poison exon. Consistent with this hypothesis, the *TRA2A* poison exon does contain a higher density of Tra2β binding sites than the *TRA2B* posion exon (Grellscheid et al., 2011a). This provides a potential mechanism for the assymetrical cross-regulation of expression observed in chapter 3. If the *TRA2A* poison exon is a stronger physiological target, it could be more sensitive to Tra2 protein expression, possibly explaining why Tra2β efficiently suppresses expression of Tra2α, rather than vice versa, in MDA-MB-231 cells.

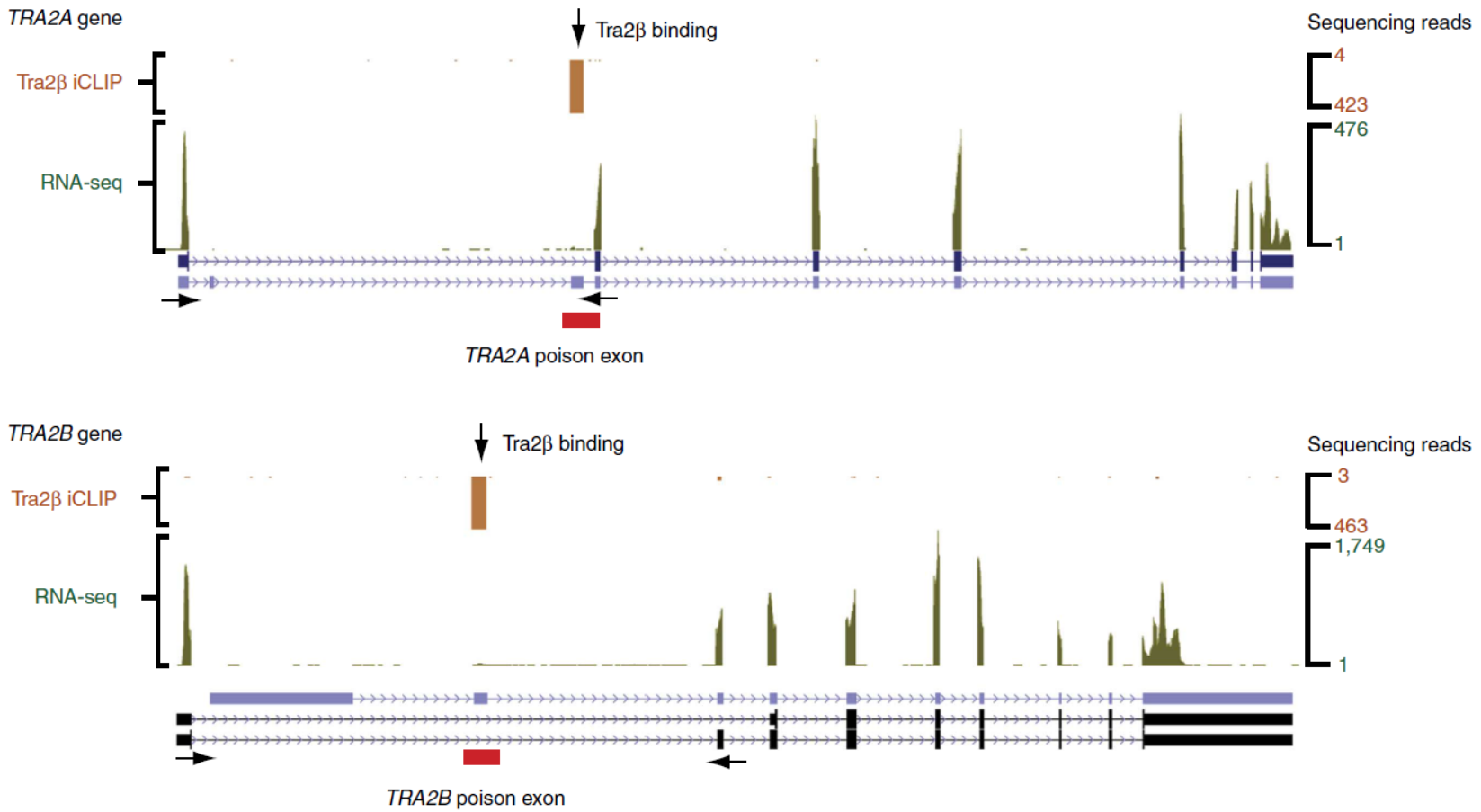


Figure 4.2 UCSC genome browser view of the *TRA2A* and *TRA2B* genes, showing the positions of Tra2β binding (orange clusters of significant cross-linking by Tra2β protein identified by iCLIP) and aligned RNA-seq reads from control MDA-MB-231 cells (green peaks over exons). A very similar density of Tra2β iCLIP tags mapped to both poison exons (see orange scale, right hand side), despite much higher expression of the *TRA2B* gene. This figure if taken from (Best et al., 2014b).

4.4.2 Validation of Tra2 protein dependent alternative splicing events from RNA-seq

To identify Tra2 protein dependent exons, I compared the splicing profile between MDA-MB-231 cells transfected with either a negative control siRNA or TRA2A and TRA2B (subsequently abbreviated to TRA2A/B) siRNA. The RNA-seq was carried out using three biological replicates per experiment. The RNA-seq data was processed and analysed by Mr Yaobo Xu (Newcastle University). Details of the RNA-seq analysis are provided in methods section 4.3.1.

By combining the Tra2β iCLIP data with the RNA-seq data, my aim was to identify exons which were both directly bound by Tra2β and functionally responsive to Tra2 protein depletion. An example of the combined iCLIP and RNA-seq data as viewed on the UCSC genome browser is shown in Figure 4.3, which shows a novel Tra2 protein dependent cassette exon in the *CHEK1* gene. Using this approach, I validated a further 21 additonal Tra2 protein dependent exons by RT-PCR and capillary gel electrophoresis. Although this was not exhaustive and additional Tra2 dependent exons remain to be identified and validated from the RNA-seq data, I validated targets in order of the highest amplitude change (i.e. the highest amplitude splicing changes were validated first). An initial panel of 12 Tra2 dependent alternative exons validated by RT-PCR and capillary gel electrophoresis are shown in Figure 4.4. All RT-PCRs used for the splicing assays were performed using triplicate biological samples to confirm reproducibility and statistical significance.

The mean change in PSI following joint Tra2 protein depletion was far greater for exons identified from the RNA-seq data when compared to exons identified using the Tra2β iCLIP data alone. The mean change in PSI for the 32 Tra2 dependent exons identified in chapter 3 using the Tra2β iCLIP data was -15.9%. In comparison, the mean change in PSI for the 12 Tra2 dependent exons identified from the RNA-seq data shown in Figure 4.4 was -51.6%. The vast difference in responsiveness between the two sets of exons is likely due to key differences in the experimental approaches. The quantity of iCLIP tags mapping to an exon is influenced by the level of gene expression; therefore an exon containing a large number of iCLIP tags may be from a highly expressed gene and isn't necessarily a highly responsive functional target. By contrast, the differentially expressed exons identified from the RNA-seq data were specifically ranked in order of the greatest fold change in exon expression.

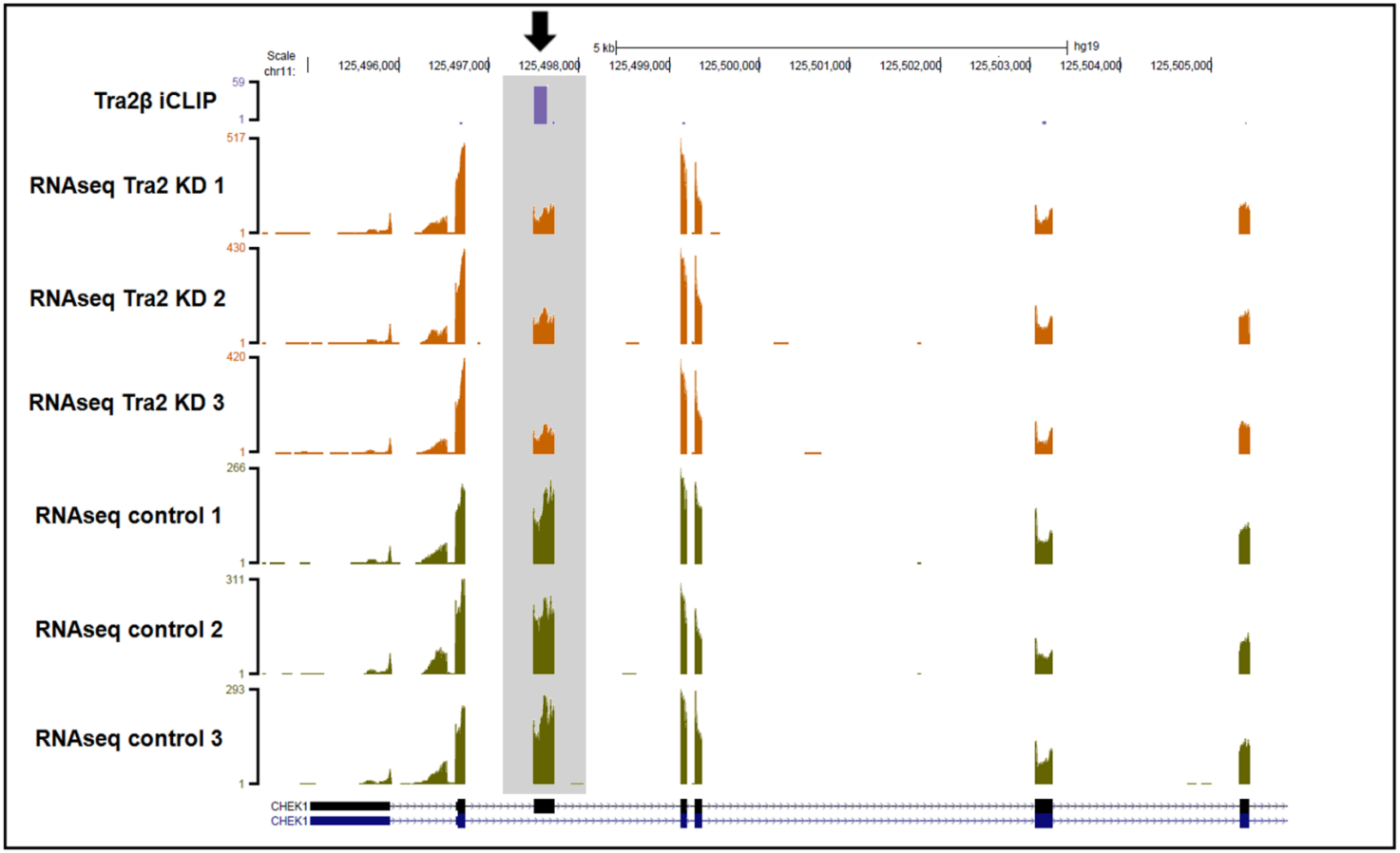


Figure 4.3 Identification of a cassette exon from *CHEK1* which was highly responsive to joint Tra2 protein depletion. A highly specific cluster of Tra2β iCLIP tags mapped directly to the exon (top track, purple). Exon expression was significantly reduced following depletion of Tra2α and Tra2β (orange RNA-seq reads) compared to negative control siRNA treated cells (green RNA-seq reads) in three biological replicates.

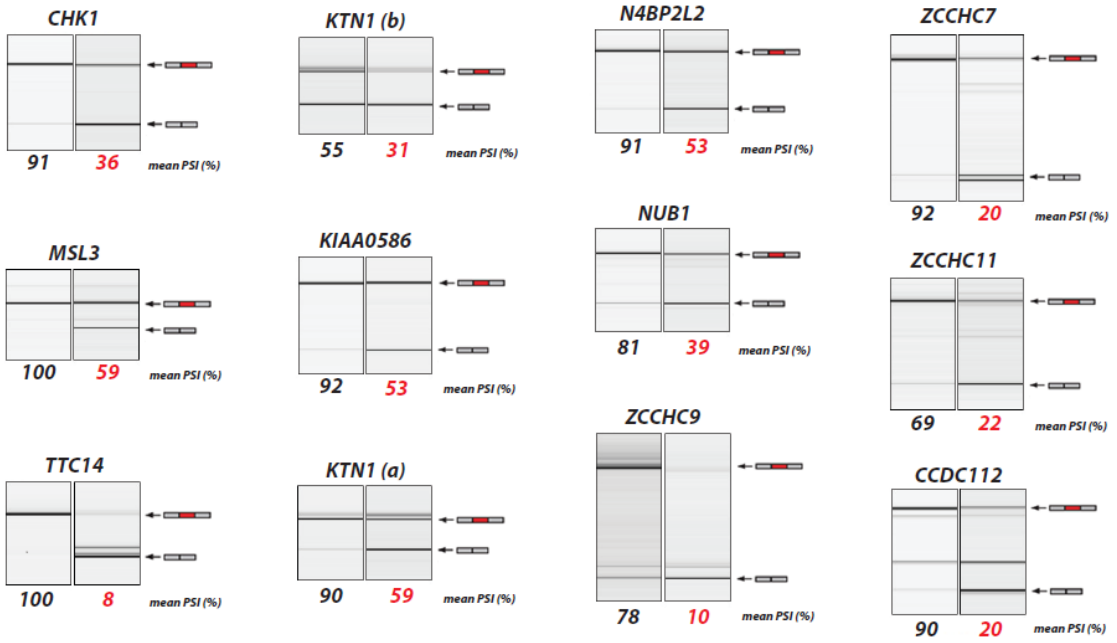


Figure 4.4 Validation of 12 Tra2 protein dependent exons identified from the combined iCLIP and RNA-seq data following joint Tra2 protein depletion. The mean percentage splicing inclusion (PSI) of Tra2 dependent exons in negative control siRNA treated cells (left hand side, black) and joint Tra2 protein depleted cells (right hand side, red), as determined by RT-PCR and capillary gel electrophoresis.

4.4.3 Tra2 proteins regulate alternative 3' ends of the *TTC14* and *BTBD7* mRNAs

In addition to the identification of regular cassette exons, I also identified Tra2 dependent alternative splicing at the 3' end of the *TTC14* and *BTBD7* mRNAs. Tra2 protein expression was essential for inclusion of an 111bp exon within an alternative 3' end of the *TTC14* gene (Figure 4.5A). To date, the Tra2 dependent exon within *TTC14* is the most highly responsive exon that I have validated; switching from 100% inclusion in negative control siRNA treated cells to just 8% inclusion in cells depleted of Tra2α and Tra2β. I also identified an exon encoding an alternative 3'UTR within the *BTBD7* gene which significantly responsded to Tra2 protein depletion (Figure 4.5B). A cluster of Tra2β iCLIP tags mapped directly to the alternative 3'UTR within *BTBD7*, suggesting that the alternative 3'UTR exon is likely a direct target of Tra2β. Tra2 protein depletion significantly reduced selection of the early 3'UTR in *BTBD7*. Selection of the early 3'UTR would produce an mRNA isoform of *BTBD7* encoding a substantially shorter protein than the full-length mRNA. Inclusion of alternative events from *TTC14* and *BTBD7* was validated by RT-PCR and capillary gel electrophoresis (Figure 4.5, right hand side).

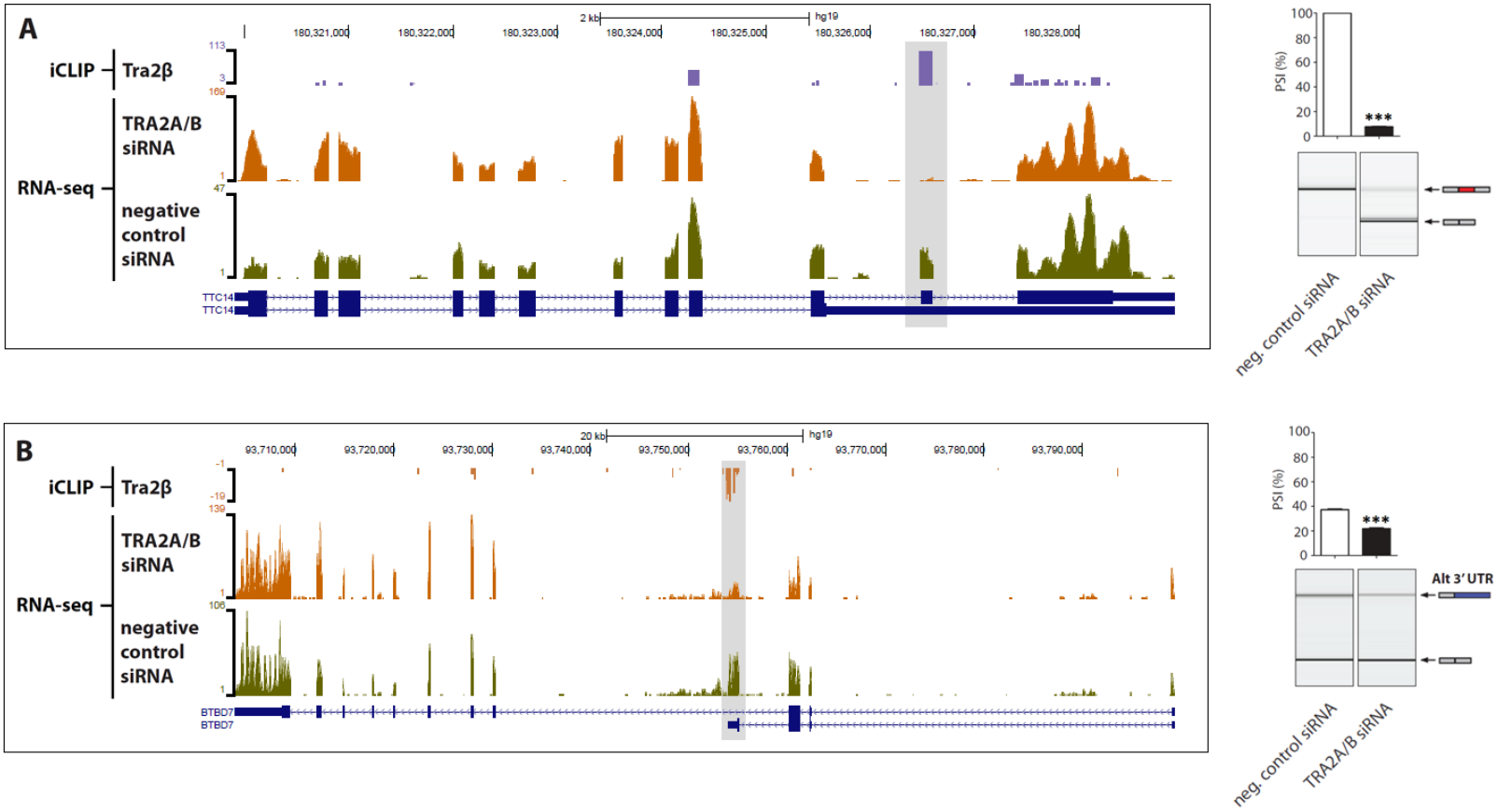


Figure 4.5 Tra2 protein dependent splicing regulation of alternative 3' ends of the *TTC14* and *BTBD7* genes. The alternative splicing events were identified using the combined iCLIP and RNA-seq data and were validated by RT-PCR and capillary gel electrophoresis. (A) Tra2 proteins are essential for inclusion of an 111bp exon close to the 3' end of the *TTC14* gene. (B) Tra2 protein depletion reduced selection of an early 3'UTR in the *BTBD7* gene. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001.

4.4.4 Human Tra2 proteins regulate inclusion of constitutively spliced target exons

Tra2 proteins have been exclusively implicated in alternative splicing regulation. However, from the initial panel of 32 Tra2 protein dependent alternative exons that I validated in chapter 3, 7 of the 32 were 100% included, whilst a further 17 of the 32 had a PSI greater than or equal to 95% in the negative control siRNA treated MDA-MB-231 cells (highlighted in green, Figure 4.6). The mean PSI of all 53 Tra2 protein dependent target exons identified in this study was 85% in negative control siRNA treated MDA-MB-231 cells (Figure 4.6, left hand column). As a large proportion of the Tra2 protein dependent exons were included at very high percentages, this suggested that either Tra2 proteins may be important for the inclusion of some constitutively spliced target exons (in addition to their expected function in alternative splicing) or alternatively that Tra2 protein expression in MDA-MB-231 cells is sufficient to induce 100% inclusion of some alternatively spliced target exons. To differentiate between these two possibilities, I used the RNA-seq data to search for changes in expression of constitutive exons following joint Tra2 protein depletion. To the best of my knowledge, these constitutive exons have never previously been annotated as alternatively spliced in human cells based on EST databases (Dreszer *et al.*, 2012; Meyer *et al.*, 2013).

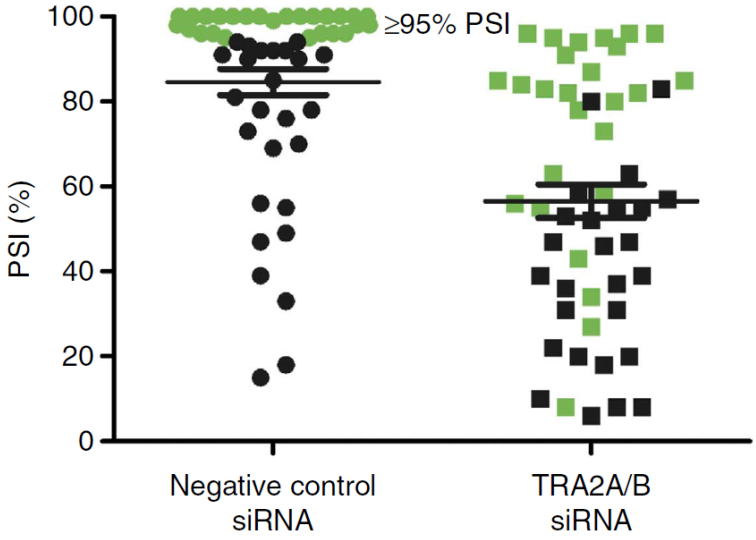


Figure 4.6 Comparison of percentage splicing inclusion of Tra2 protein dependent exons. Comparison of PSI between negative control siRNA treated cells (left hand side) and joint Tra2 protein depleted cells (right hand side). Many Tra2 dependent exons were included at very high levels under control conditions (negative control siRNA). Exons included at equal to or greater than 95% PSI in control conditions are highlighted in green. This figure is taken from (Best et al., 2014b).

I identified 9 candidate constitutive exons which appeared to change expression in the RNA-seq data following joint Tra2 protein depletion. Each of the candidate constitutive exons were annotated as constitutive on the UCSC genome browser and I could find no previous evidence of exon skipping based on the UCSC expressed sequence tag (EST) database. The 9 exons all contained Tra2β iCLIP tags mapping directly to the exon, suggesting each exon was a likely direct target of Tra2β. An example of a candidate constitutive exon in the *SMC4* gene is shown below in Figure 4.7.

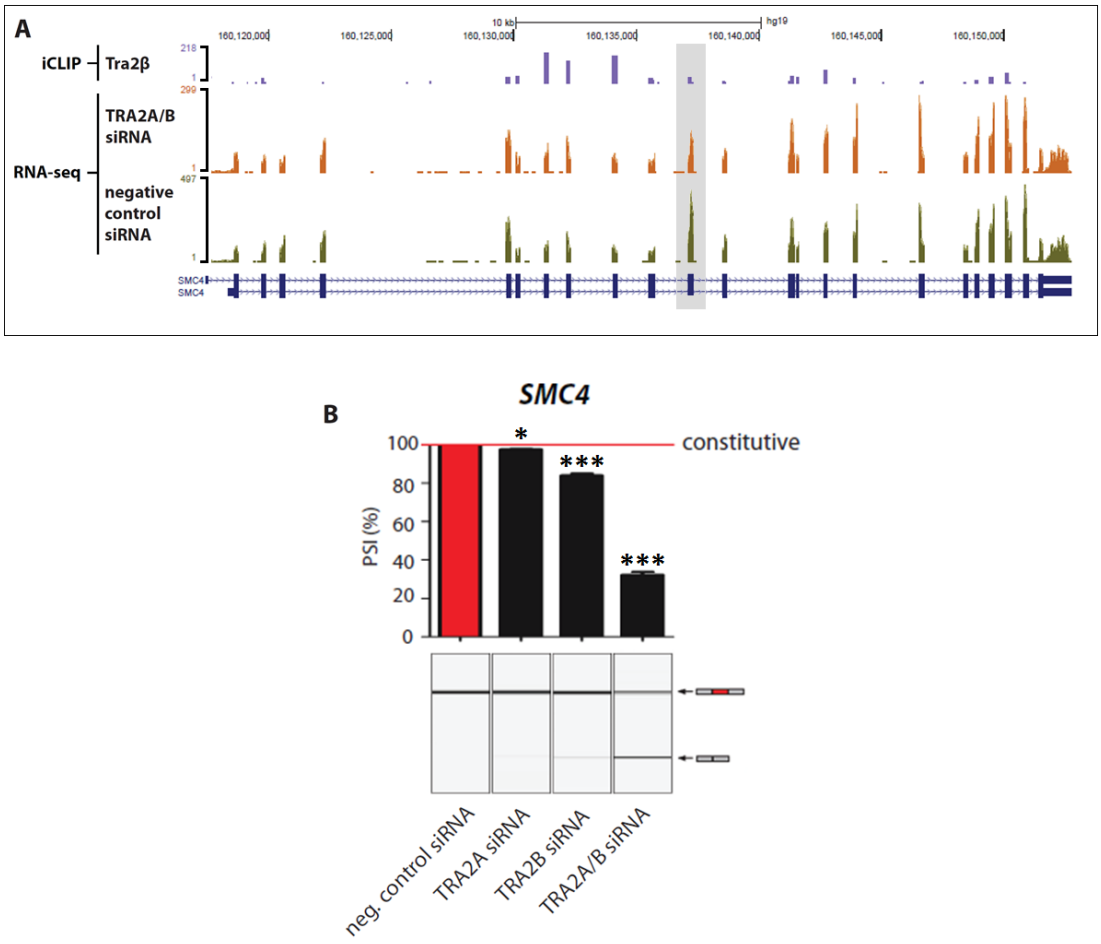


Figure 4.7 Tra2 proteins are essential for inclusion of a constitutively spliced exon in the *SMC4* gene. (A) Combined iCLIP and RNA-seq data, shopwing expression of a constitutive exon within *SMC4* is significantly reduced following joint Tra2 protein depletion compared to negative control siRNA treated cells (exon highlighted in grey). (B) The *SMC4* exon is constitutively spliced (100% included) in negative control siRNA treated MDA-MB-231 cells (red). Single depletion of either Tra2α or Tra2β slightly reduced exon inclusion, whilst joint depletion of both Tra2α and Tra2β substantially reduced inclusion to just 30% PSI (validation by RT-PCR and capillary gel electrophoresis). Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001.

Similar Tra2 protein dependent constitutive exons were validated from *ANKRD1*, *NFXL2*, *NIPBL* and *PDCD6IP*. Each of the five exons were 100% included in negative control siRNA treated MDA-MB-231 cells (they were constitutively spliced) but were skipped to different degrees following joint Tra2 protein depletion (Figure 4.8).

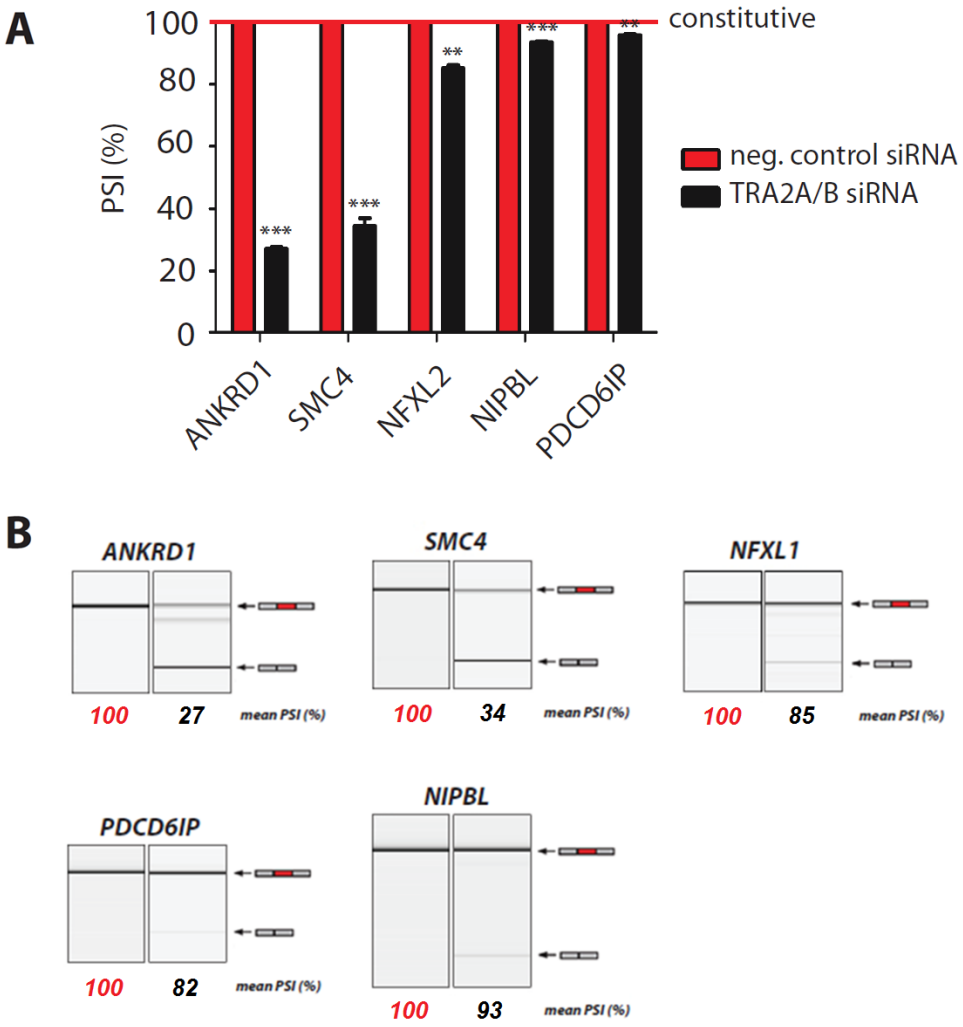


Figure 4.8 Identification of five constitutively spliced exons from *ANKRD1, SMC4, NFXL2, NIPBL* and *PDCD6IP* which were dependent on Tra2 protein expression for 100% splicing inclusion in MDA-MB-231 cells. (A) The five exons were included 100% in negative control siRNA treated MDA-MB-231 cells (red). Inclusion of each exon was significantly reduced following joint depletion of Tra2α and Tra2β (black). (B) Representative RT-PCR and capillary gel electrophoresis images which show inclusion of the constitutive exons was significantly reduced following joint Tra2 protein depletion. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is adapted from (Best et al., 2014b).

The remaining four candidate exons from *NEXN*, *ATXN2*, *TBCD1D12* and *MPHOSPH10* were highly responsive to Tra2 protein depletion, but were included less than 100% in negative control siRNA treated MDA-MB-231 cells (ranging from 94-98% inclusion) (Figure 4.9). This data suggests that the UCSC annotation of alternative events is not fully comprehensive.

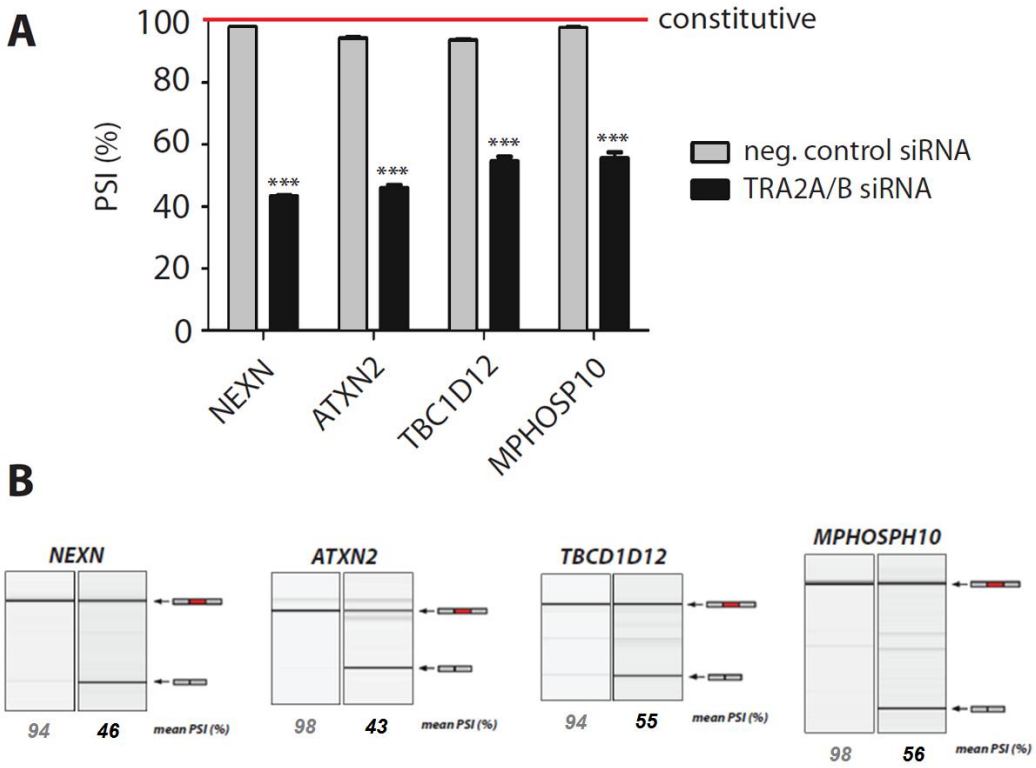


Figure 4.9 Four candidate constitutive exons from *NEXN*, *ATXN2*, *TBC1D12* and *MPHOSPH10* that were highly responsive to Tra2 protein depletion, but were included less than 100% in negative control siRNA treated MDA-MB-231 cells (i.e. not constitutively included). (A) The four exons ranged from 94-98% inclusion in negative control siRNA treated MDA-MB-231 cells (white). Inclusion of each exon was significantly reduced after joint depletion of Tra2α and Tra2β (black). (B) Representative RT-PCR and capillary gel electrophoresis images showing inclusion of each exon was significantly reduced following joint depletion of Tra2α and Tra2β (black, right hand side) compared to negative control siRNA treated cells (grey, left hand side). Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is adapted from (Best et al., 2014b).

4.4.5 Tra2β binding sites are important for inclusion of the *ANKRD1* constitutive exon

The most highly responsive, bona fide constitutive exon following joint Tra2 protein depletion in the MDA-MB-231 cells was an exon from *ANKRD1* (combined iCLIP and RNA-seq data for the *ANKRD1* gene is shown in Figure 4.10 A). Remarkably, single depletion of either Tra2α or Tra2β had no significant effect on inclusion of the *ANKRD1* constitutive exon in MDA-MB-231 cells (Figure 4.10 C, columns 2 and 3), yet joint Tra2 protein depletion substantially reduced exon inclusion (a -73 point PSI change) (Figure 4.10 C, compare columns 1 and 4). To investigate whether exonic Tra2β binding sites were required for inclusion of this *ANKRD1* constitutive exon, Ms. Marina Danilenko (Newcastle University) cloned the *ANKRD1* constitutive exon and approximately 300bp of the flanking intronic sequences into a pXJ41 minigene for splicing analysis. Subsequently, two clusters of Tra2β binding sites were mutated to analyse their impact on inclusion of this constitutive exon. The *ANKRD1* minigene mutagenesis was also performed by Ms. Marina Danilenko. I performed the transfection of the minigene constructs into HEK-293 cells and the RNA analysis.

The wildtype *ANKRD1* exon expressed from the minigene had a PSI of just 72% when co-transfected into HEK-293 cells along with a GFP-only control (Figure 4.10 D, column 2). In a parallel experiment, the endogenous *ANKRD1* exon was 100% included in the same HEK-293 cells (Figure 4.10 D, column 1). This suggests that the minigene may lack some important flanking sequences required for 100% inclusion or that the shorter introns created in the minigene construct may reduce the level of splicing inclusion compared to the endogenous exon.

Analysis of the *ANKRD1* exonic sequence uncovered two clusters of GAA-rich sequences; annotated as Tra2β binding site clusters 1 and 2 (Figure 4.10 B). To determine whether either of the two binding site clusters was important for inclusion of the *ANKRD1* minigene exon, site-directed mutagenesis was used to insert single base mutations and disrupt the Tra2β binding sites (single base mutations are highlighted in red, Figure 4.10 B). Consistent with the exonic binding sites being important for splicing inclusion, mutation of either Tra2β binding site cluster significantly reduced splicing inclusion on the *ANKRD1* exon when compared to the wildtype minigene exon. Strikingly, mutation of Tra2β binding site cluster 1 (labelled as Mutation 1, which is closer to the 3' splice site) completely abolished splicing inclusion

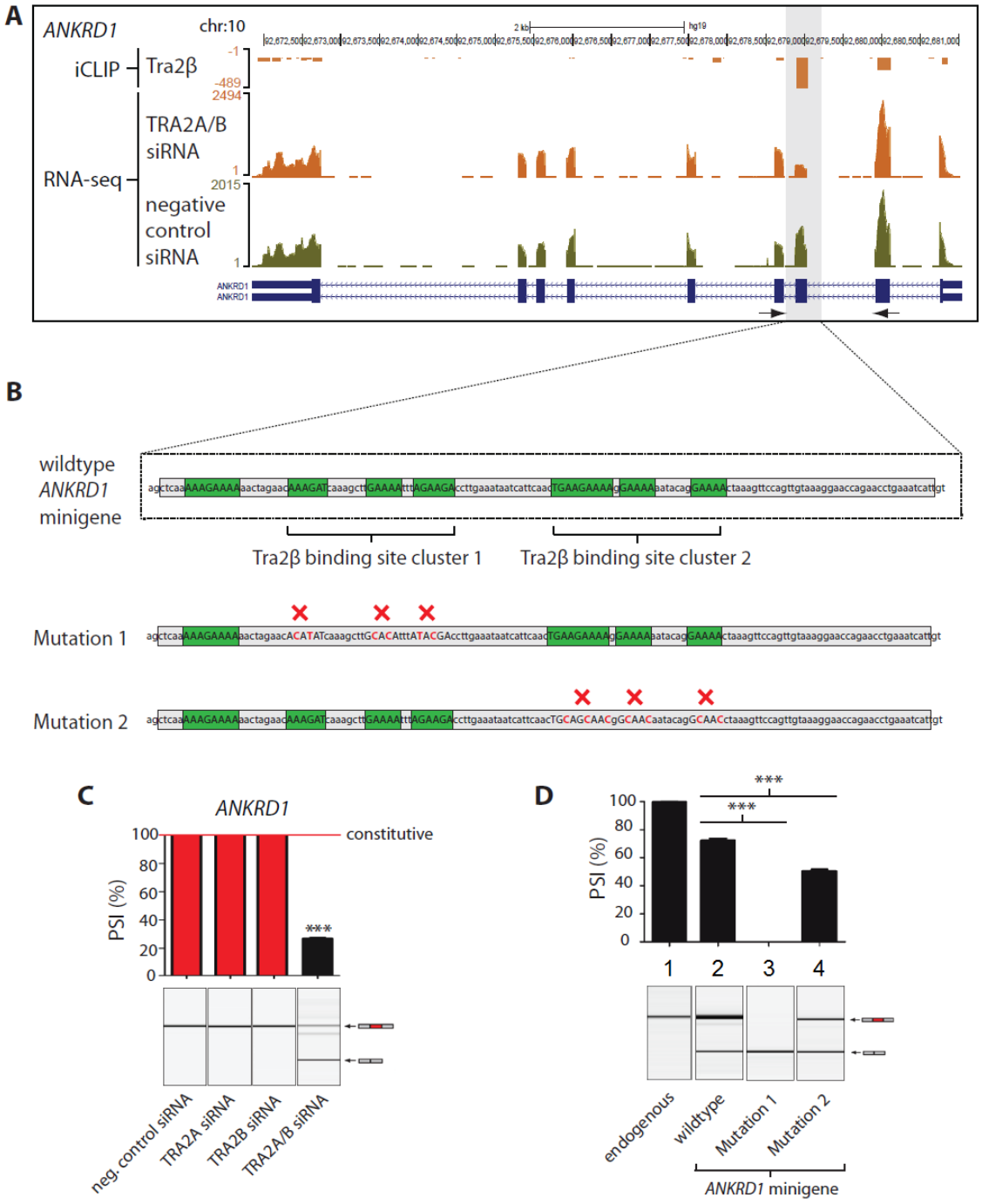


Figure 4.10 Exonic Tra2β binding sites are required for inclusion of a constitutive exon in *ANKRD1*. (A) Combined iCLIP and RNA-seq data showing a Tra2 dependant constitutive exon from the *ANKRD1* gene (dashed box) (B) The wildtype *ANKRD1* minigene contained two clusters of Tra2β binding sites, which were independently mutated. (C) Splicing inclusion of the endogenous *ANKRD1* exon following single or joint Tra2 protein depletion in MDA-MB-231 cells. (D) Inclusion of the endogenous *ANKRD1* exon and various *ANKRD1* minigenes transfected into HEK-293 cells. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is adapted from (Best et al., 2014b).

of the *ANKRD1* exon, from 72% inclusion to 0% inclusion (Figure 4.10 D, compare column 2 with column 3). This suggests the exonic sequences within cluster 1 are particularly important for inclusion of this exon. Mutation of Tra2β binding site cluster 2 (labelled as Mutation 2) also significantly reduced inclusion of the *ANKRD1* exon when compared to the wildtype minigene exon, but to a lesser extent than Mutation 1 (Figure 4.10 D, compare column 2 with column 4).

4.4.6 Tra2 dependent exons have an intrinsic sensitivity to Tra2 protein expression

I validated an additional 21 Tra2 protein dependent exons from the RNA-seq data. Therefore in total I validated 53 Tra2 protein dependent exons by RT-PCR and capillary gel electrophoresis (from both the iCLIP and RNA-seq datasets). I observed a broad range in the amplitude of response to joint Tra2 protein depletion. This varied considerably, ranging from just a -4% change in splicing inclusion, up to a -92% change in splicing inclusion (Figure 4.11). This indicates that individual Tra2 protein dependent exons have different intrinsic sensitivities to Tra2 protein expression. The most highly responsive exons (which I define as showing a greater change in PSI than the mean) are labelled and highlighted in red (Figure 4.11). The length of Tra2 protein dependent exons also ranged considerably, from just 64 nucleotides at the shortest (a short cassette exon within the *SMYD2* gene, which showed a -5% reduction in PSI in response to Tra2 protein depletion), up to 5916 nucleotides at the longest (a particularly large cassette exon within the *SON* gene, which showed a -8% reduction in PSI in response to Tra2 protein depletion).

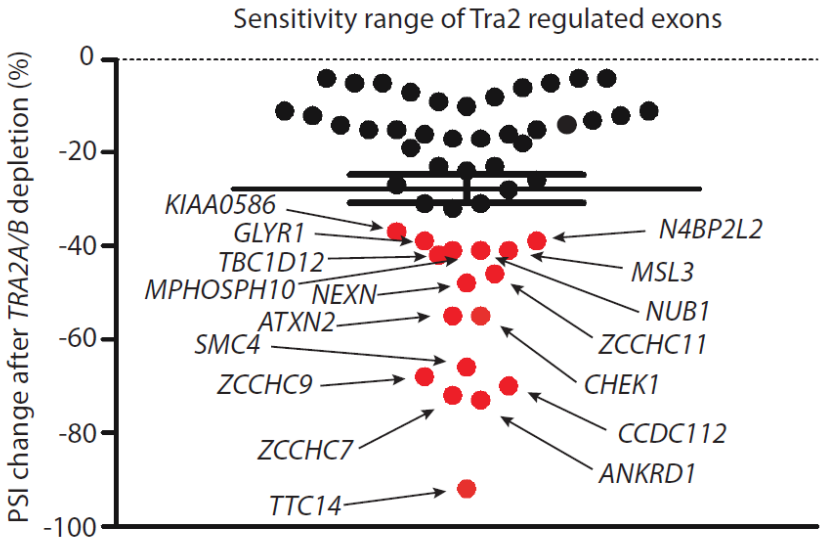


Figure 4.11 Scatterplot showing the change in PSI for all 53 Tra2 dependent exons following joint depletion of endogenous Tra2α and Tra2β. The change in exon inclusion ranged from a 4% reduction in PSI up to a 92 reduction in PSI following joint Tra2 protein depletion, indicating individual exons have intrinsic sensitivities to Tra2 protein expression. Scatterplot shows the mean change in PSI of 53 exons ±s.e.m. The highest amplitude individual exons (which I define as showing a greater change in PSI than the mean ±s.e.m.) are labelled and highlighted in red. This figure is taken from (Best et al., 2014b).

4.4.7 Functionally responsive exons have a higher density of Tra2β binding sites

In an attempt to discern why some exons containing significant numbers of Tra2β iCLIP tags responded to Tra2 protein depletion and other exons did not, I compared various parameters between the responsive and non-responsive exons. The parameters I compared were the exon length (bp), the density of Tra2β binding sites within the exon, the 5' splice site strength, the 3' splice site strength, the combined 5' and 3' splice site strength and the single weakest splice site score. Complete exon comparison data is available as supplementary data in the following publication (Best et al., 2014b). The density of Tra2β binding sites within the exon was calculated as the percentage of bases within the exon that matched the top 10 pentamers from the Tra2β iCLIP experiment (Figure 3.9). Splice site scores were calculated using the Splicing Regulation Online Graphical Engine (SROOGLE) (Schwartz et al., 2009) and is based on the Max entropy model of splicing (Burge., 2004). As some exons responded but very weakly, I classified responsive exons as those which showed a greater than 15% change in mean PSI% following joint Tra2 depletion.

The only parameter which was significantly different between the responsive (>15%) and non-responsive exons was the density of Tra2β binding sites within the exon, which was significantly higher in Tra2 responsive exons (>15%) than non-responsive exons (Figure 4.12). Interestingly, the two Tra2 poison exons had a significantly higher density of Tra2β binding sites compared to the average Tra2 responsive exon (>15%), as well as the non-responsive exons (Figure 4.12). This suggests that the negative feedback loop between Tra2 proteins is highly sensitive to Tra2 protein concentration due to a particularly high density of Tra2 binding sites within the poison exons.

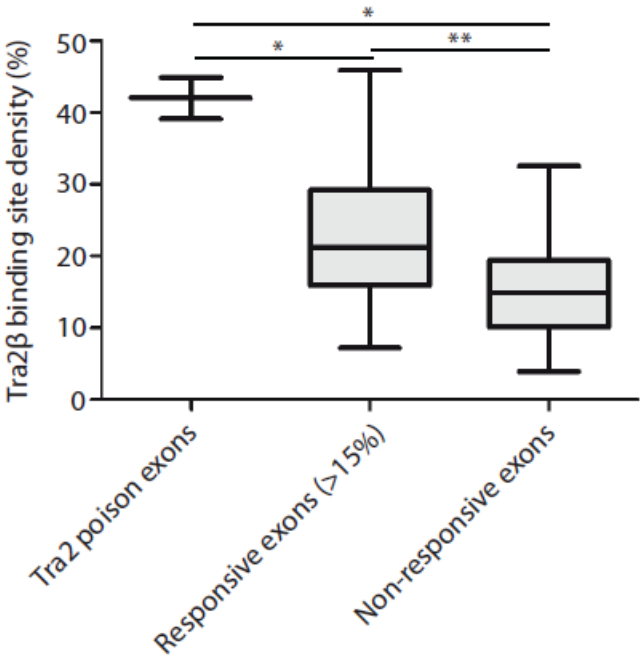


Figure 4.12 Comparison of Tra2β binding site density between responsive and non-repsonsive exons. Tra2β binding site density was significantly higher in Tra2 responsive exons (>15 PSI change) than in non-responsive exons. Tra2β binding site density was also significantly higher in the two Tra2 poison exons than the average Tra2 responsive exons (>15 PSI change) and the non-responsive exons. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This image is taken from (Best et al., 2014b).

4.4.8 Tra2 proteins are highly specific splicing regulators of certain target exons

To investigate whether Tra2 protein dependent exons are under the combinatorial control of numerous RNA-binding proteins, or are specifically dependent on Tra2 proteins for splicing inclusion, the percentage splicing inclusion of four Tra2 protein dependent exons was monitored after knockdown of a large panel of splicing factors. Exon inclusion was monitored using a custom high-throughput RT-PCR plate containing

cDNAs from MDA-MB-231 cells in which 53 known splicing regulators had been individually and systematically depleted using RNAi (we also included cDNA from Tra2 protein depleted MDA-MB-231 cells for comparison) (Huang et al., 2013) (Figure 4.13). The high-throughput RT-PCR screen was performed by the team of Dr. Roscoe Klinck (Université de Sherbrooke, Canada). Only knockdowns with a minimum efficiency of 50% mRNA depletion (previously determined via qPCR) were included in this panel (Venables et al., 2013b). Splicing inclusion was monitored for two constitutive exons (from *ANKRD1* and *SMC4*) and two alternative exons (from *CHEK1* and *GLYR1*).

Remarkably, of all 53 knockdowns tested, only joint Tra2 protein depletion affected inclusion of the *ANKRD1* constitutive exon, suggesting regulation of this constitutive exon was highly specific to Tra2 proteins (TRA2A/B siRNA treated cells are highlighted in red, Figure 4.13 A). Joint Tra2 protein depletion also had the largest effect on inclusion of the *GLYR1* alternative exon; though this exon did respond minimally (<10% PSI change) to depletion of numerous other splicing factors (Figure 4.13 D). Inclusion of the *SMC4* constitutive exon was significantly affected by knockdown of *SNRP70* (which encodes U170K) and SRPK1, though joint Tra2 protein depletion also induced the strongest effect on this exon (Figure 4.13 B). Splicing inclusion of the *CHEK1* alternative exon was strongly affected by joint Tra2 proteins depletion, but was also significantly shifted by depletion of multiple RNA-binding proteins including three core U2snRNP components (SF3A1, SF3B4 and SF3A3), SFRS2, hnRNPK, hnRNPC2, KHSRP and CDC5L (Figure 4.13 C). This is consistent with a much broader combinatorial control of *CHEK1* exon 3 compared to the other three exons.

As this panel of splicing factor depletions is not completely exhaustive, it is not possible to exclude other RNA-binding proteins from playing a role in control of these exons. However, this data is consistent with Tra2 proteins being amongst the most quantitatively important splicing regulators for these particular target exons.

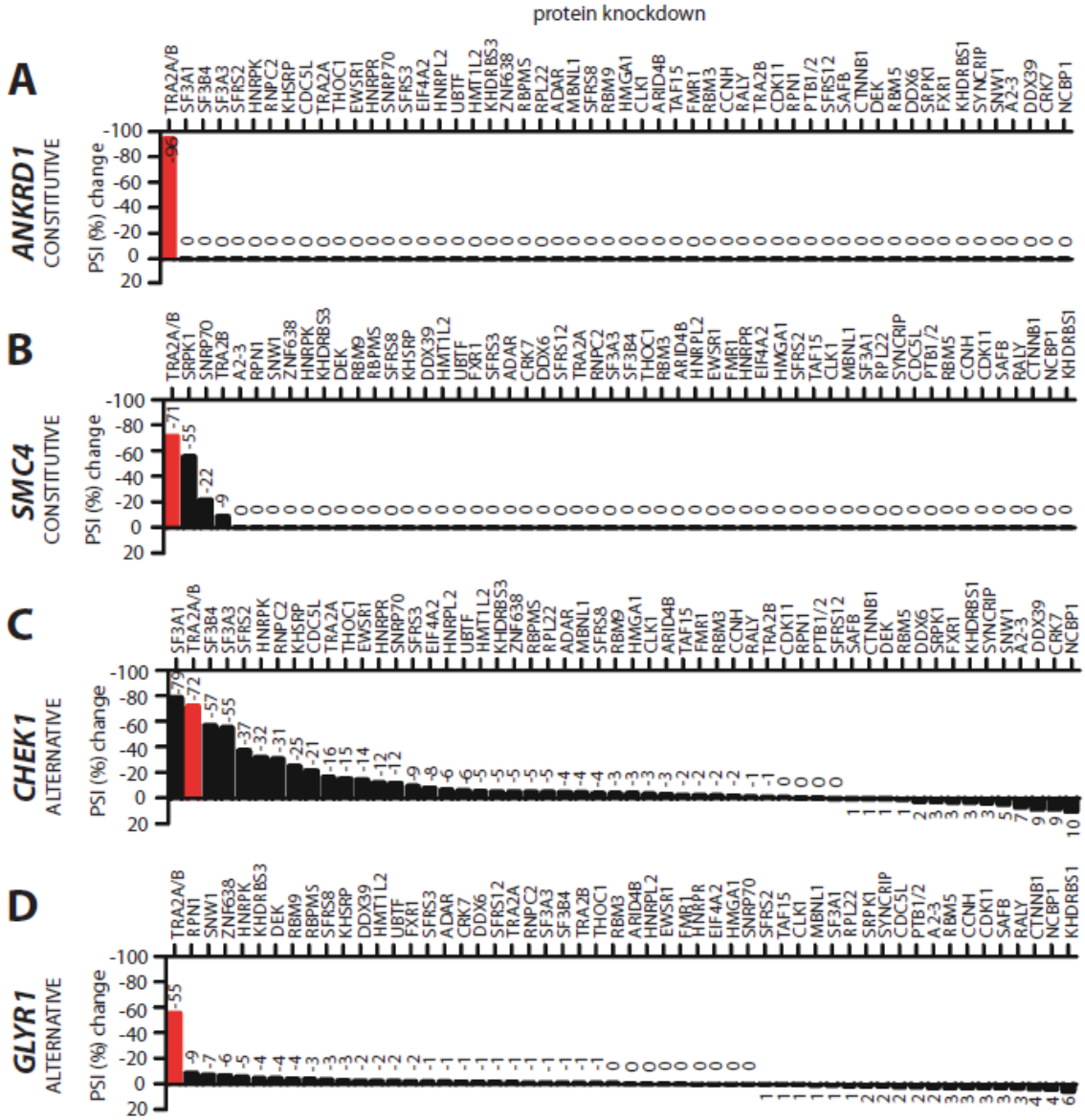


Figure 4.13 Change in percentage splicing inclusion of four Tra2 dependent exons following knockdown of 53 individual RNA binding proteins in MDA-MB-231 cells. The change in percentage splicing inclusion was monitored for two constitutively spliced exons (A) *ANKRD1* and (B) *SMC4*, as well as two alternatively spliced exons (C) *CHEK1* and (D) *GLYR1*. Joint Tra2 protein depleted cells are highlighted in red. This figure is adapted from (Best et al., 2014b).

4.4.9 Tra2 target exons are enriched within genes associated with chromosome biology

To investigate whether Tra2 proteins regulate alternative splicing of genes involved in functionally related biological processes, a Gene Ontology (GO) enrichment analysis was performed using my data by Dr. Katherine James (Newcastle University). The GO enrichment analysis determined whether the 53 genes containing Tra2 protein dependent exons were enriched for particular biological processes (GO terms) relative to all genes expressed in the MDA-MB-231 cell line. Further details regarding the methodology of the GO enrichment analysis are provided in methods section 4.3.

The GO enrichment analysis identified five biological processes (GO terms) which were significantly enriched in genes containing Tra2 protein dependent exons compared to all genes expressed in MDA-MB-231 cells (Figure 4.14 A). Eleven of the 53 Tra2-regulated genes we identified (20.8%) were annotated to one or more of the five enriched GO terms, with ten of those eleven genes annotated to the single, broader term “chromosome organisation” (Figure 4.14 B). Some genes annotated to “chromosome organisation” were also annotated to the conceptually related terms “histone modification” and “chromatin modification” (Figure 4.14 B).

To investigate whether the genes containing Tra2 protein dependent exons which were not associated with one of the five enriched GO terms had any known interactions with other proteins associated with these terms, an interaction network analysis was also performed by Dr. Katherine James (Newcastle University). Briefly, the BioGRID database (Stark et al., 2006) was used to retrieve a network of known functional interaction data involving the genes containing Tra2 protein dependent exons and other proteins associated with any of the five enriched GO terms. In addition to the eleven genes directly annotated to the five enriched GO terms (Figure 4.14 A), an additional 23 of the 53 genes containing Tra2 dependent exons (43.4%) have known functional interactions with genes annotated to one or more of these five terms (Figure 4.15). Suprisingly, although indirectly connected within the network via these annotated genes, none of the 53 genes containing validated Tra2 dependent exons have any known functional interactions with one another in the BioGRID database (Figure 4.15). The GO enrichment analysis and protein interaction network data is summarised in Figure 4.14 C.

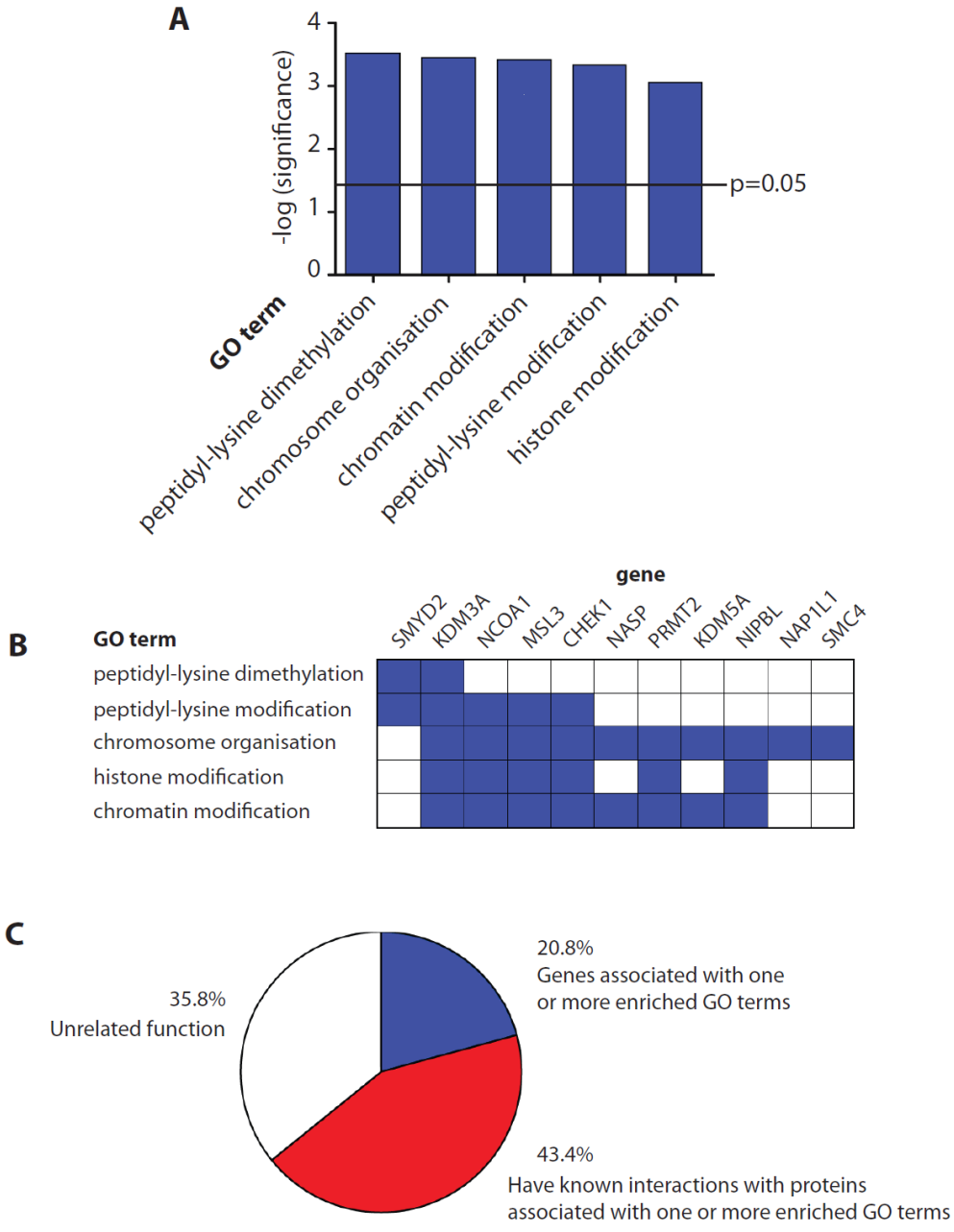


Figure 4.14 Tra2 protein dependent exons are enriched within genes associated with chromosome biology. (A) A Gene Ontology (GO) enrichment analysis identified enrichment of five specific GO terms in genes containing Tra2 dependent exons, which include multiple aspects of chromosome biology. (B) The eleven genes containing Tra2 dependent exons that are associated with one or more of the enriched GO terms. Ten of the eleven genes are associated with the single broader term "chromosome organisation". A number of the Tra2-regulated genes are annotated to multiple conceptually related GO terms. (C) Summary of Tra2-regulated genes: genes that were directly annotated to an enriched biological process (20.8%, blue), genes that have a known functional interaction with a protein that is annotated to one of enriched processes (43.4%, red), and those which have unknown or unrelated functions (35.8%, white). This figure is adapted from (Best et al., 2014b).

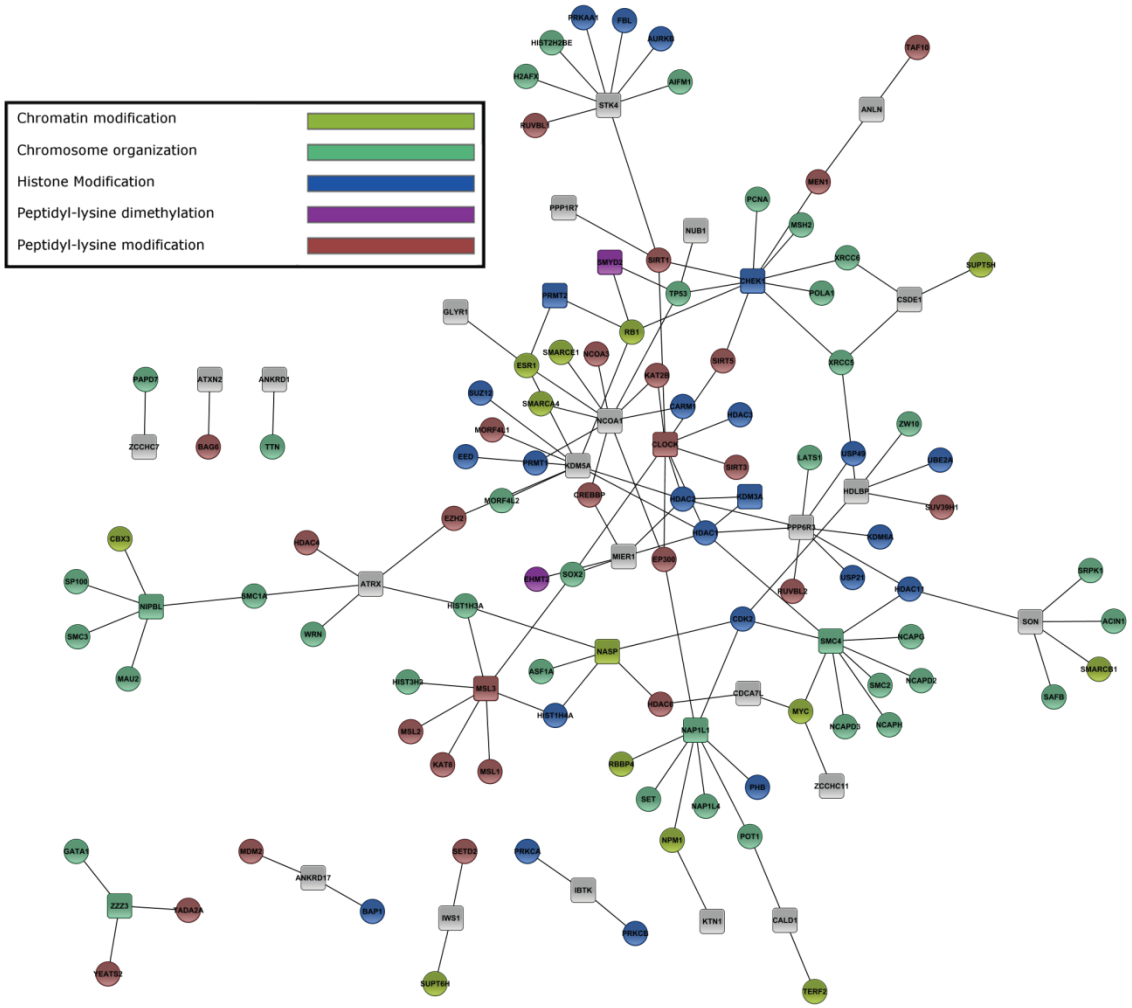


Figure 4.15 Network representation of genes containing Tra2 protein regulated exons (square nodes) and their known interactions with other proteins annotated to one of the five enriched GO terms (circular nodes). Where a protein was annotated to more than one term, the more specific child term was chosen for colouring. However, it should be noted that many of the genes in the network are annotated to multiple enriched terms (as shown in Figure 4.12B). This network interaction was produced entirely by Dr. Katherine James (Newcastle University), using my data. A higher resolution version of this figure is available as supplementary information in the following publication (Best et al., 2014b).

4.4.10 Differentially expressed genes (following joint Tra2 protein depletion) were enriched for cell cycle-related functions

In addition to the GO enrichment analysis of genes containing Tra2 protein dependent exons, Dr. Katherine James (Newcastle University) also performed an independent GO enrichment analysis on the differentially expressed genes (identified from RNA-seq data) following joint Tra2 protein depletion. The GO enrichment analysis of differentially expressed genes indentified ten GO terms that were significantly enriched ($p<0.05$) in the differentially expressed genes following joint Tra2 protein

depletion compared to all genes expressed in MDA-MB-231 cells (Figure 4.16). Remarkably, 7 of the 10 enriched GO terms were directly linked to cell cycle related functions (highlighted in green, Figure 4.16.) The enriched biological processes also included the terms "response to radiation", "endoplasmic reticulum unfolded" and "DNA repair". The effect of joint Tra2 protein depletion on cell growth and the cell cycle is investigated later in the chapter.

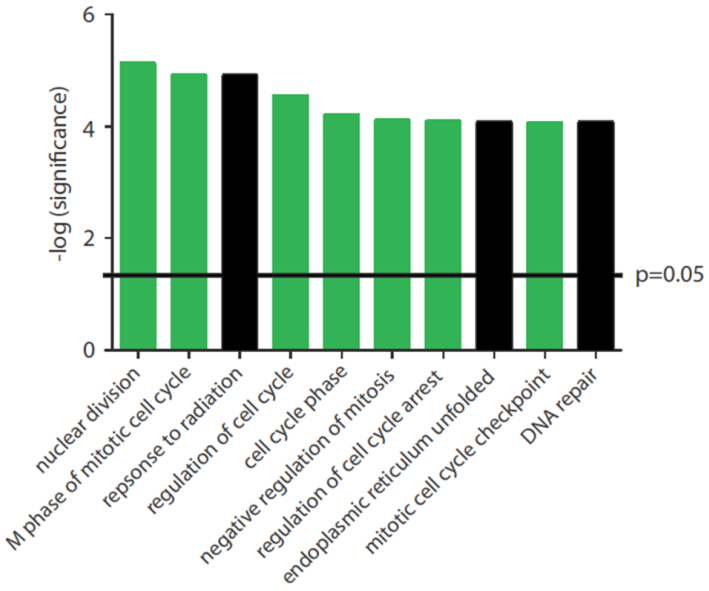


Figure 4.16 Gene Ontology (GO) enrichment analysis of differentially expressed genes following joint Tra2 protein depletion. Differentially expressed genes were highly enriched in GO terms associated with the cell cycle (green).

4.4.11 Validation of differentially expressed changes (following joint Tra2 protein depletion) by qPCR

I validated a panel of the differentially expressed genes identified from the RNA-seq data by qPCR (Figure 4.17). Gene expression was compared between negative control siRNA treated MDA-MB-231 cells and *TRA2A*/*B* siRNA treated MDA-MB-21 cells, 72 hours after siRNA transfection. Gene expression changes were confirmed from Tra2 protein depleted cells using 2 independent sets of siRNA, targeting different regions of the *TRA2A* and *TRA2B* mRNA. In total, statistically significant changes in gene expression were validated for 14 genes. A significant increase in gene expression was confirmed for the protein coding genes *SELPLG*, *FOSB*, *FOS*, *DDIT4*, *EGR1* and *TTC14* (Figure 4.17 A, left hand side). A significant increase in gene expression was also confirmed for the lncRNA genes *RP11-618G20.2*, *RP-1129J21* and *CTD-2037K23* (Figure 4.17 A, right hand side). A significant reduction in expression was confirmed for the

protein coding genes *ANKRD26*, *EIF5B* and *HIST2H2BE*, as well as the lncRNA genes *RP11-221J22* and *LIN00601* (Figure 4.17 B.) Tra2 proteins have not previously been associated with any roles in transcriptional regulation, and the differentially expressed genes were not enriched with Tra2β iCLIP tags. It was considered that the changes in gene expression were possibly an indirect consequence of Tra2 protein depletion (i.e. downstream of changes to cell growth or the cell cycle), therefore they were not investigated further in this chapter. Alternatively, changes in alternative splicing of genes encoding proteins involved in chromosomal regulation (i.e. transcriptional factors in the broadest sense) may have a direct effect on the transcription of genes associated with cell cycle regulation.

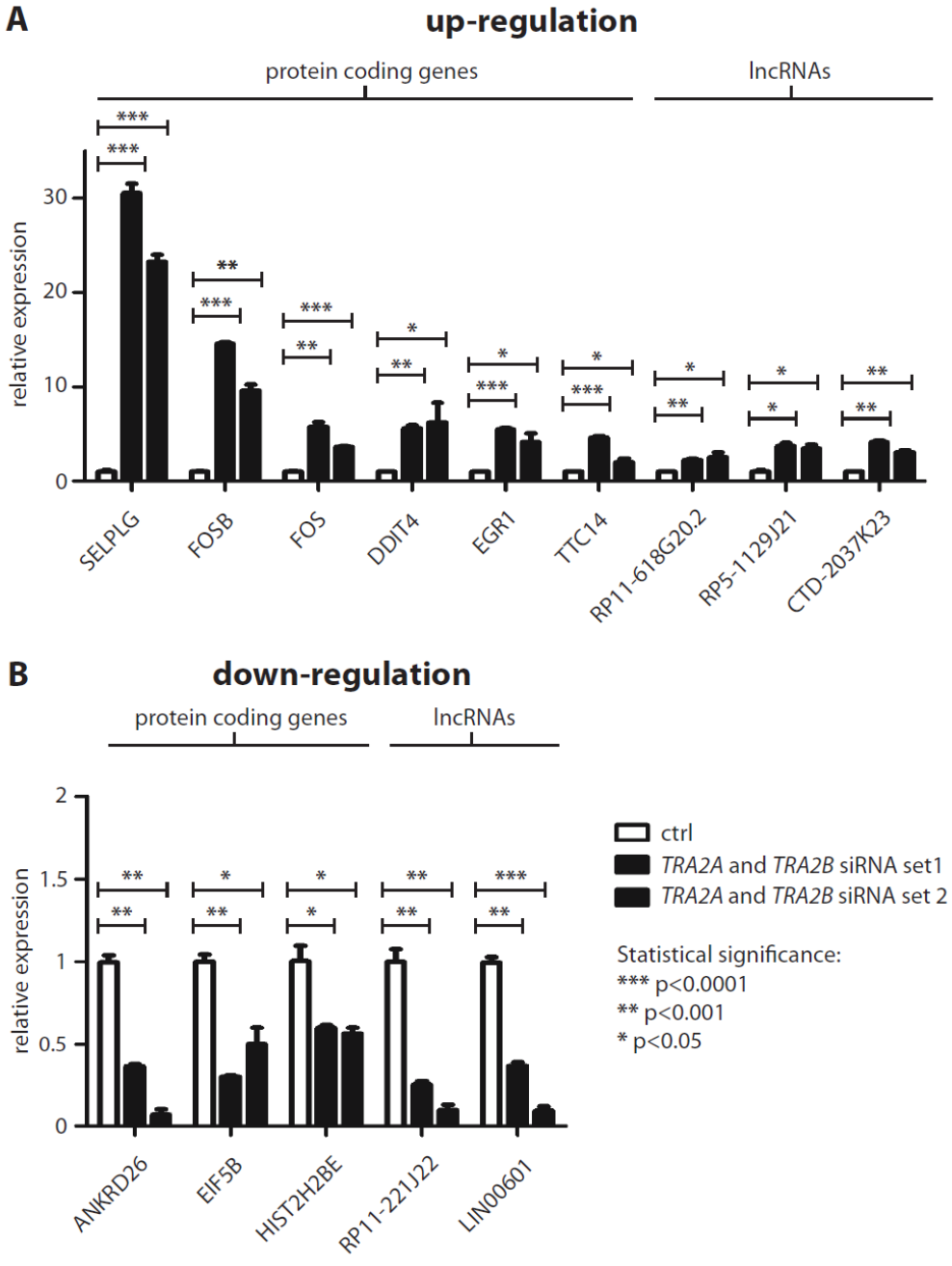


Figure 4.17 Validation of gene expression changes following joint Tra2 protein depletion by qPCR. (A) Validation of a panel of protein coding genes and lncRNAs with increased expression following joint Tra2 protein depletion. (B) Validation of a panel of protein coding genes and lncRNAs with decreased expression following joint Tra2 protein depletion. Data represents the mean of three biological replicates ±s.e.m. Gene expression changes were validated using two indepedant set of siRNA, targeting different regions of the *TRA2A* and *TRA2B* mRNA. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001.

4.4.12 Human Tra2 proteins are required for splicing inclusion of *CHEK1* exon 3

One of the ten Tra2 protein dependent exons from genes involved in chromosome organisation was exon 3 of the *CHEK1* gene. *CHEK1* encodes the serine/threonine protein kinase CHK1, which plays a key role in the DNA damage response and cell cycle checkpoint control (McNeely et al., 2014). A significant cluster of Tra2β iCLIP tags mapped directly to this exon and the RNA-seq data showed inclusion of this exon was significantly reduced following joint Tra2 protein depletion (Figure 4.18 A). Validation of the RNA-seq data by RT-PCR and capillary gel electrophoresis showed a 55% reduction in splicing inclusion of *CHEK1* exon 3 following joint Tra2 protein depletion in MDA-MB-231 cells (Figure 4.18 B). To determine whether *CHEK1* exon 3 was regulated by Tra2 proteins in other cells types as well as MDA-MB-231 cells, I transfected three additional cell types (MCF-7, PC3 and HeLa) with either a negative control siRNA or joint TRA2A/B siRNAs and monitored splicing inclusion of *CHEK1* exon 3 (Figure 4.18 B). The siRNA transfections were performed using three biological replicates per experiment for each cell line. Inclusion of *CHEK1* exon 3 was significantly reduced in all four cell lines tested following joint Tra2 protein depletion. This data suggests that Tra2 protein expression is important for inclusion of *CHEK1* exon 3 in multiple different cell lines.

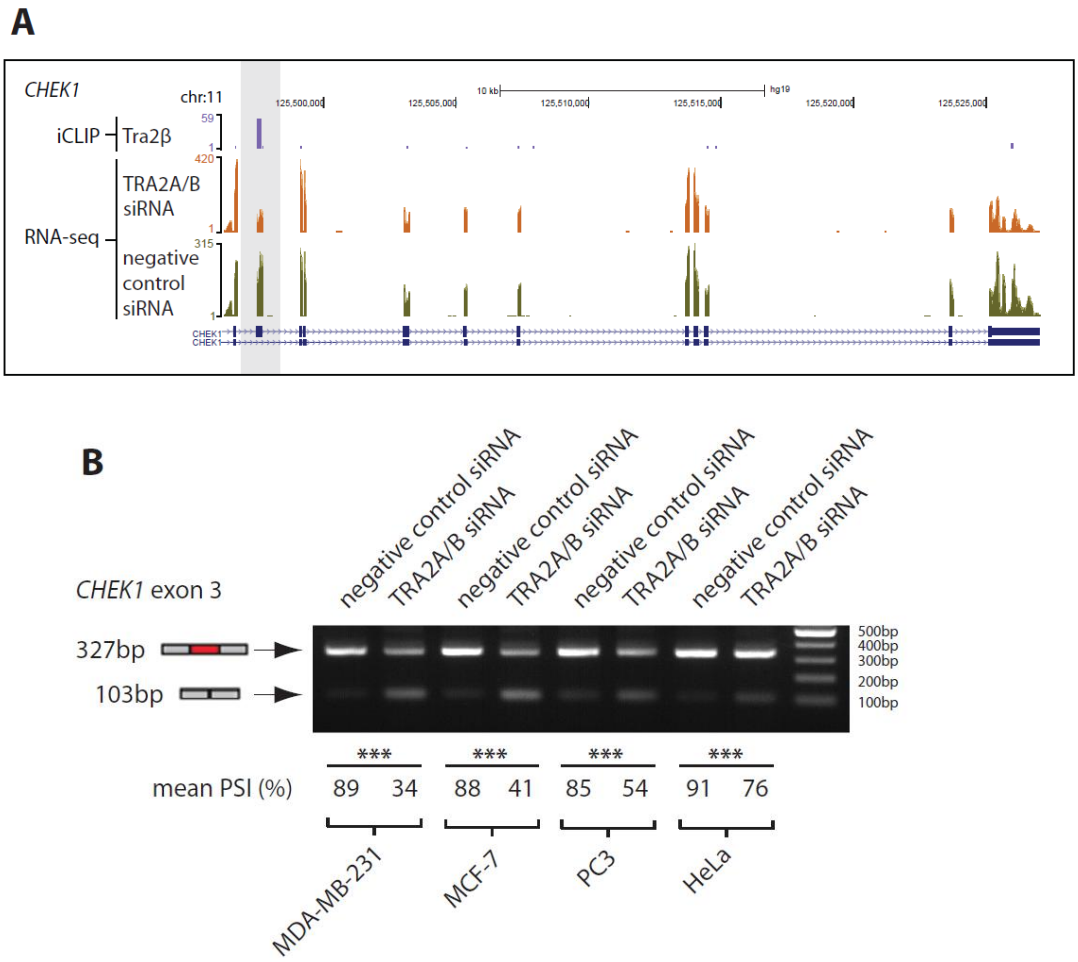


Figure 4.18 Inclusion of *CHEK1* exon 3 is significantly reduced following joint depletion of Tra2α and Tra2β. (A) Tra2β iCLIP tags mapped directly to exon 3 of the *CHEK1* gene (top track, purple). Expression of this exon was also significantly reduced following joint knockdown of Tra2α and Tra2β (orange RNA-seq reads) compared to the negative control cells (green RNA-seq reads). (B) Splicing inclusion of *CHEK1* exon 3 was significantly reduced in MDA-MB-231, MCF-7, PC3 and HeLa cells following TRA2A/B siRNA transfection compared to negative control siRNA treated cells. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is taken from (Best et al., 2014b).

4.4.13 Confirming direct interaction between Tra2β and the *CHEK1* exon 3 RNA using Electrophoretic Mobility Shift Assays (EMSAs)

To confirm that *CHEK1* exon 3 was a direct target for Tra2β binding *in vitro,* Electrophoretic Mobility Shift Assays (EMSAs) were performed using radio-labelled RNA probes corresponding to different regions of the *CHEK1* gene. I carried out the molecular cloning of different regions of the *CHEK1* gene into the pBluescript vector used for *in vitro* transcription. The remaining *CHEK1* exon 3 EMSA experiment was performed entirely by Mrs. Caroline Dalgliesh (Newcastle University).

Analysis of the iCLIP data at higher resolution showed that the Tra2β iCLIP tags mapped throughout exon 3, but were particularly enriched towards the 3' splice site of the exon (Figure 4.15A). RNA probe A corresponds to a region close to the 3' splice site of *CHEK1* exon 3. This sequence contained the most Tra2β iCLIP tags and also contained the most predicted binding sites for Tra2β (Tra2β binding sites are shaded green in Figure 4.19 B, right hand side). RNA probe A was very efficiently shifted by very low concentrations of purified Tra2β protein (as low as 25ng). RNA probe B corresponds to a region of the exon closer to the 5' splice site. This sequence contained fewer Tra2β iCLIP tags and just one predicted Tra2β binding site. RNA probe B was also shifted by purified Tra2β protein, but not as efficiently as RNA probe A. A control intronic sequence which contained no Tra2β iCLIP tags or predicted binding sites for Tra2β was not shifted by any concentration of purified Tra2β protein. Importantly, all RNA probes were of a similar length (either 68 or 69 nucleotides).

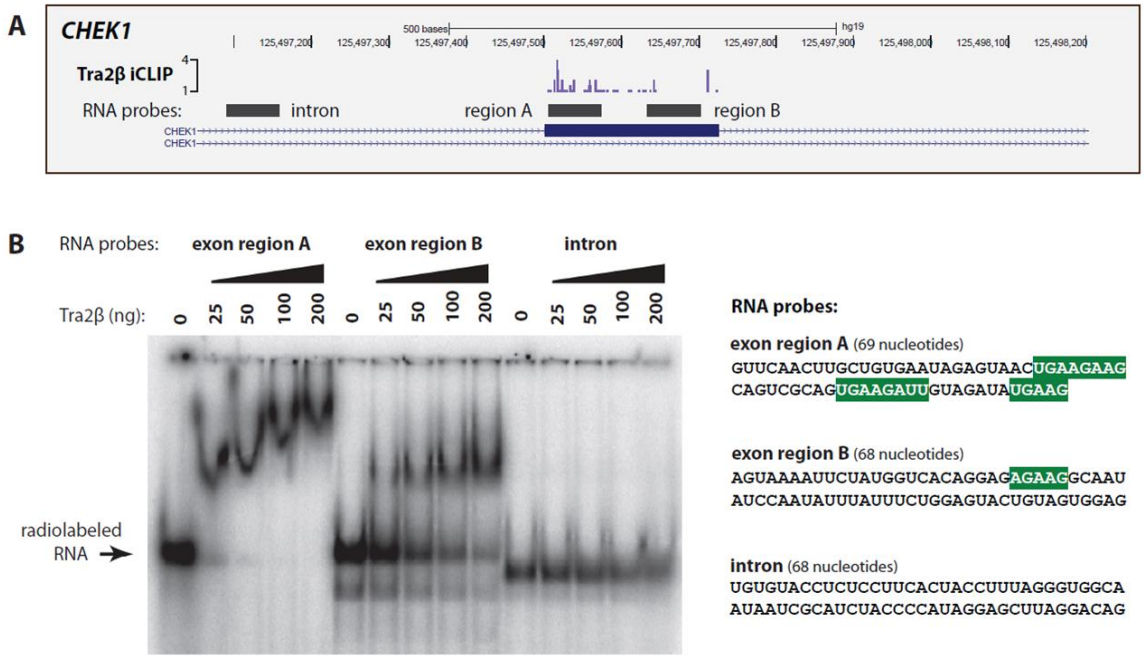


Figure 4.19 Confirming direct interactions between Tra2β and *CHEK1* exon 3 by EMSA. (A) High magnification view of Tra2β iCLIP tags mapping directly to exon 3 of the *CHEK1* gene. Three short regions (exon region A, exon region B and intron) were cloned into a Bluescript vector to create RNA probes for the electrophoretic mobility shift assay (EMSA). (B) EMSA showing radio labelled RNA probes shifting with increasing concentrations of purified Tra2β protein. The RNA probe corresponding to exon region A was shifted efficiently by increasing concentrations of purified Tra2β protein. Similarly, the RNA probe corresponding to exon region B was also shifted by purified Tra2β protein, but less efficiently than region A. The RNA probe corresponding

to the intron was not shifted by any concentration of purified Tra2β protein. This figure is taken from (Best et al., 2014b).

4.4.14 Tra2α and Tra2β directly activate splicing of a *CHEK1* exon 3 minigene

To confirm that *CHEK1* exon 3 is a direct target for splicing regulation by Tra2 proteins, *CHEK1* exon 3 and approximately 300bp of flanking intronic sequences was cloned into a pXJ41 minigene for splicing analysis. Molecular cloning of the two *CHEK1* minigenes was carried out by Ms. Marina Danilenko (Newcastle University), whilst I performed the subsequent minigene transfections, RNA extraction, RT-PCRs and analysis.

Unexpectedly, the wildtype *CHEK1* exon 3 minigene exon was not included (0% inclusion) when co-tranfected into HEK-293 cells with a GFP-only control (Figure 4.20 A, column 1). However, inclusion of the exon was strongly induced in response to co-transfection with either Tra2β-GFP or Tra2α-GFP (Figure 4.20 A, columns 2 and 3). Co-transfection with Tra2βΔRRM-GFP (which lacks the RNA recognition motif) failed to activate exon inclusion (Figure 4.20 A, column 4), suggesting Tra2β activates *CHEK1* exon 3 inclusion through direct RNA binding. Next, the predicted Tra2β binding sites within the exon were mutated to assess their impact on splicing inclusion of the exon (point mutations are highlighted in red, Figure 4.20 B). Remarkably, point mutations within the *CHEK1* exon completely abolished splicing activation in response to co-transfection with either Tra2β-GFP or Tra2α-GFP (Figure 4.20 B, columns 2 and 3), suggesting that both Tra2α and Tra2β directly activate splicing inclusion of *CHEK1* exon 3 through through exonic binding.

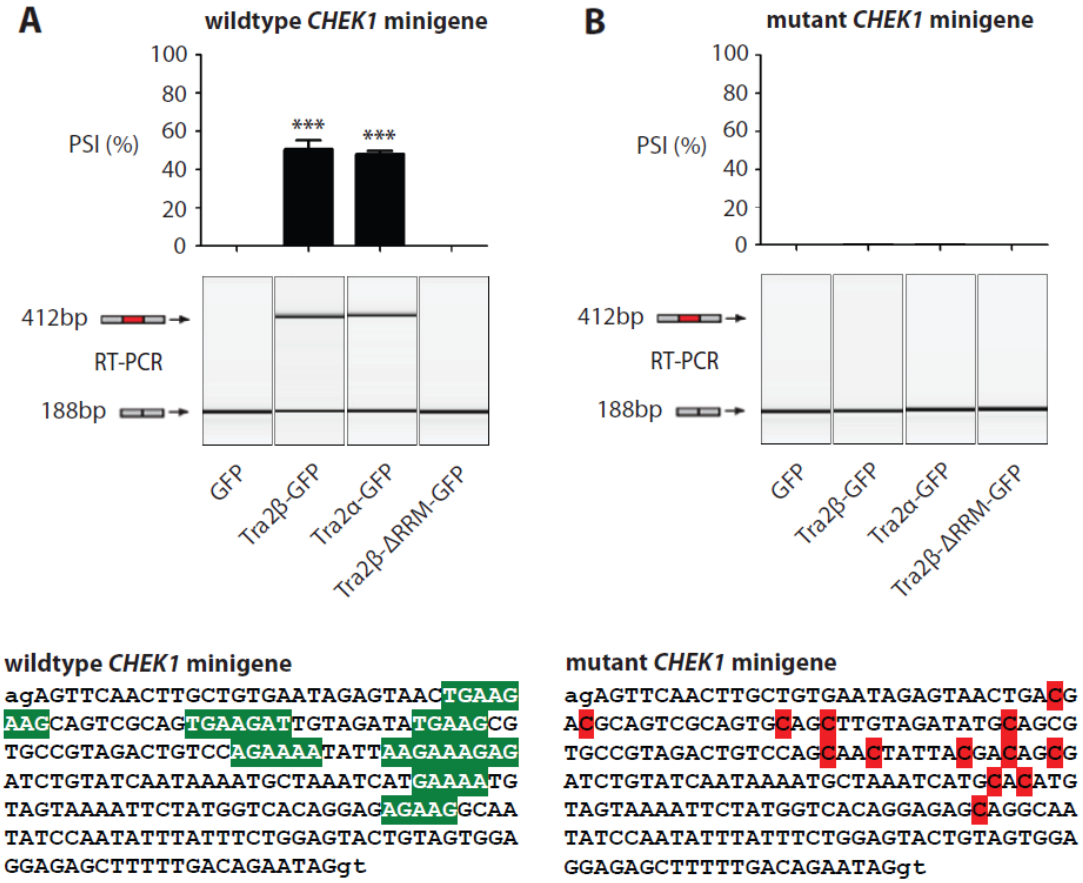


Figure 4.20 Splicing patterns of two minigenes containing either a wildtype or mutated version of *CHEK1* exon 3. (A) The wildtype exon was not included when co-transfected with GFP alone, but was strongly activated when co-transfected with either Tra2α-GFP or Tra2β-GFP. Co-transfection with Tra2βΔRRM-GFP did not activate splicing inclusion of the wildtype exon. (B) Disrupting the exonic Tra2β binding sites using single base mutations completely abolished splicing activation of the *CHEK1* exon by Tra2α-GFP or Tra2β-GFP. Data represents the mean of three biological replicates ±s.e.m. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is adapted from (Best et al., 2014b).

4.4.15 Splicing inclusion of endogenous *CHEK1* exon 3 remained constant in KG1 cell populations containing different cell cycle profiles

Full-length CHK1 protein (54kDa) is produced when exon 3 is included in the *CHEK1* mRNA. A shorter isoform of the CHK1 protein termed CHK1-S (produced when exon 3 is skipped) has previously been reported to be cell cycle regulated. Expression of the CHK1-S protein isoform was reported to significantly increase during S-phase and reach maximal expression during G2/M (Pabla et al., 2012). Therefore to investigate whether splicing inclusion of *CHEK1* exon 3 was cell cycle regulated, I analysed inclusion of the exon in a panel of KG1 cell populations with different cell cycle profiles.

KG1 cells were fractionated by centrifugal elutriation and assessed for expression of two cell cycle markers (CENPF and H3Ser10) to estimate the percentage of the cell population in various stages of the cell cycle (Figure 4.21, upper panel). The frozen pellets of fractionated cells were a gift from Professor Caroline Austin (Newcastle University). The elutriation and cell cycle analysis was performed entirely by Dr. Ian Cowell and Ka Cheong Lee (Newcastle University). I performed the RNA analysis. However, splicing inclusion of *CHEK1* exon 3 remained constant in all cell fractions tested by RT-PCR and capillary gel electrophoresis, despite containing different cell cycle profiles (Figure 4.21, lower panel).

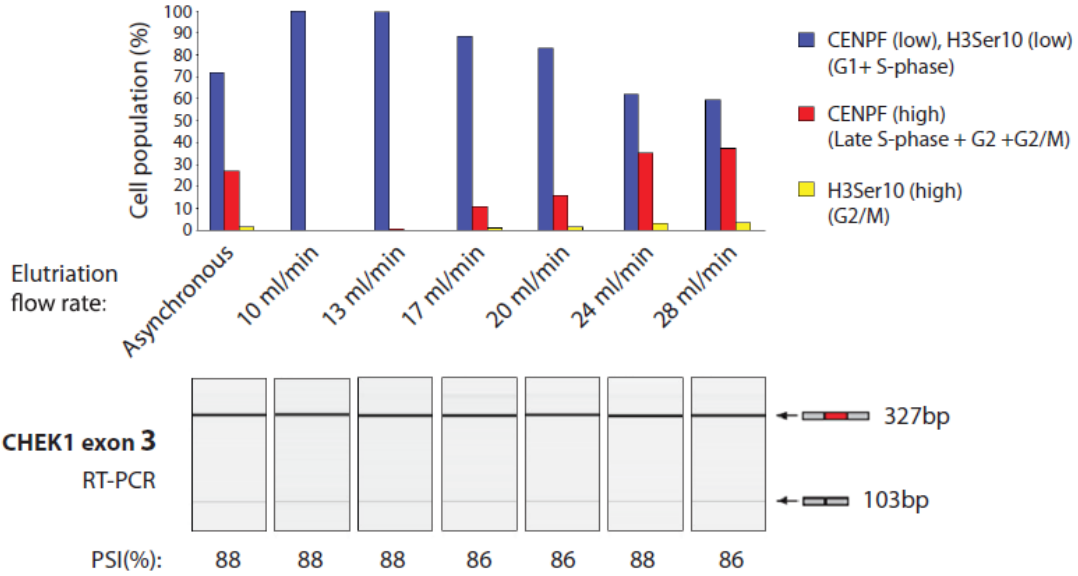


Figure 4.21 Splicing inclusion of *CHEK1* exon 3 across a panel of KG1 cell populations containing different cell cycle profiles. Inclusion of endogenous *CHEK1* exon 3 was monitored in the different populations of KG1 cells by RT-PCR and capillary gel electrophoresis. Inclusion of *CHEK1* exon 3 remained constant in all cell fractions tested. The data shown is generated from single biological samples. This figure is taken from (Best et al., 2014b).

4.4.16 Splicing profile of *CHEK1* exon 3 in a panel of breast tumour biopsies

Expression of CHK1-S has been reported to be up-regulated in various cancer tissues. CHK1-S mRNA and protein expression was reported to be up-regulated in testicular carcinomas, whilst CHK1-S protein expression was also found to be up-regulated in late stage ovarian cancer (Pabla et al., 2012). Therefore I also investigated the level of splicing inclusion of *CHEK1* exon 3 in a panel of breast cancer biospsies by RT-PCR and capillary gel electrophoresis. I detected high levels of both *CHEK1* splice isoforms in

RNA purified from a small panel of breast cancer tissues, although I did not observe any obvious enrichment of either isoform in any particular tumour classification at the RNA level (Figure 4.22).

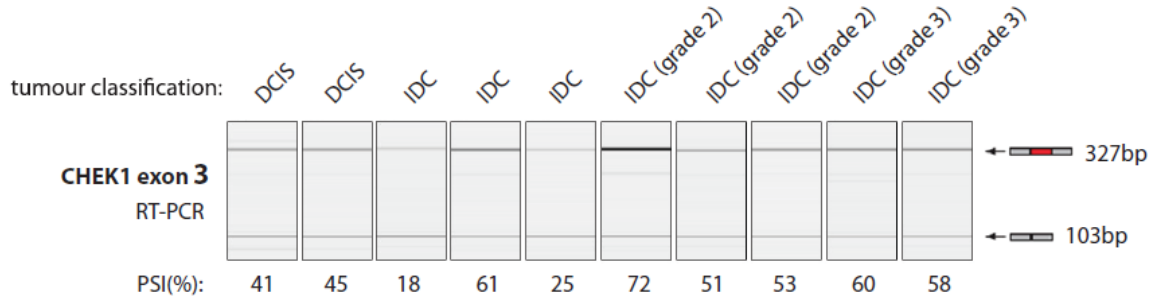


Figure 4.22 Splicing inclusion of *CHEK1* exon 3 in a panel of ten breast tumour biopsies. Inclusion of the endogenous exon was monitored by RT-PCR and capillary gel electrophoresis, using RNA extracted from frozen breast tumour sections obtained from the Breast Cancer Campaign Tissue Bank. Tumour classifications: ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC), either assigned no specific grade or classified as grade 2 or 3. This figure is taken from (Best et al., 2014b).

4.4.17 Tra2 protein depletion led to reduced expression of the full-length CHK1 protein

To investigate whether Tra2 proteins were required for expression of the full-length CHK1 protein, I monitored CHK1 protein expression by Western Immunoblotting. A single major CHK1 protein isoform was detected in MDA-MB-231 cells, corresponding to the expected size of full-length CHK1 protein (54KDa) (Figure 4.23, lanes 1-3). This band was considerably reduced following transfection with a *CHEK1*-specific siRNA, suggesting this band was specific to the CHK1 protein (Figure 4.23, lanes 4-6). Consistent with Tra2 proteins maintaining expression of the full-length CHK1 protein, expression of this band was also substantially reduced following joint Tra2 protein depletion, in three biological replicates (Figure 4.23, lanes 7-9).

A shorter isoform of CHK1 (termed CHK1-S, predicted size 47kDa) has previously been reported to be translated from an alternative downstream translational initiation site when exon 3 is excluded from the *CHEK1* mRNA (Pabla et al., 2012). In my experiments, although joint Tra2 protein depletion substantially reduced splicing inclusion of *CHEK1* exon 3 and expression of the full-length CHK1 protein, it did not lead to any detectable increase in any shorter isoform of the CHK1 protein (Figure 4.23, lanes 7-9). N.B. the

CHK1 antibody used in this study was validated in previous publications as capable of recognising the shorter isoform (CHK1-S) of this protein (Pabla *et al.*, 2012). I also observed a smaller reduction in expression of the full-length CHK1 protein following joint Tra2 protein depletion in MCF-7, PC3 and HeLa cells (Figure 4.24). However, this data is generated from single biological samples. The reduction in full-length CHK1 protein expression across the four cell lines was proportional to the change in PSI of *CHEK1* exon 3 at the RNA level for each cell line. Overall, the above data is most consistent with Tra2α and Tra2β activity being essential for expression of full-length CHK1 protein, rather than inducing expression of any shorter protein isoform of CHK1.

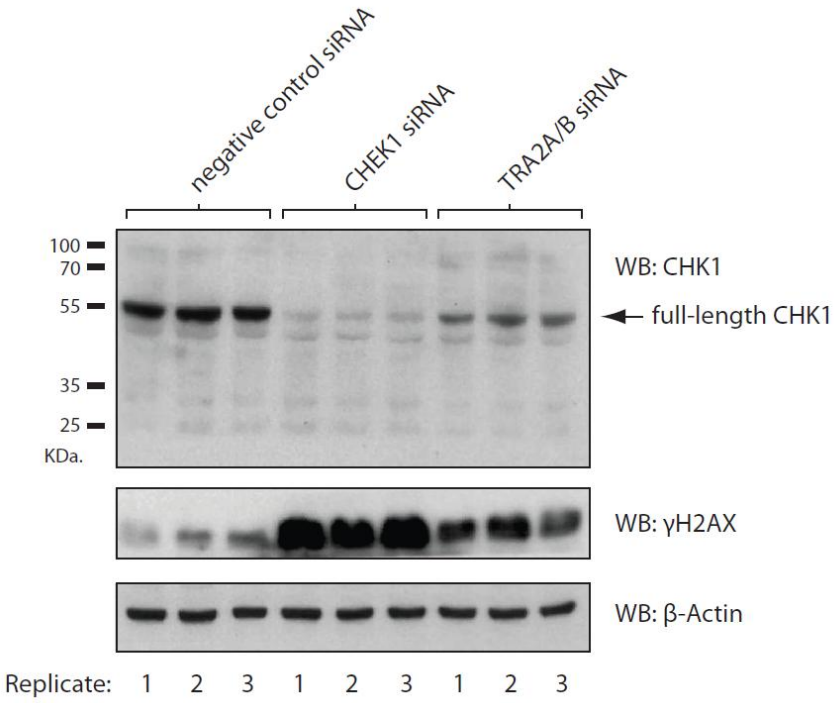


Figure 4.23 Human Tra2 proteins are essential for expression of full-length CHK1 protein. CHK1 protein expression was detected by Western Immunoblotting. Full-length CHK1 (54kDa) protein expression was reduced following transfection with either a CHEK1-specific siRNA or with TRA2A and TRA2B specific siRNAs in MDA-MB-231 cells. Loss of full length CHK1 expression was associated with increased levels of phosphorylated γH2AX in both CHEK1 siRNA and TRA2A/B siRNA treated cells. Expression of β-Actin was detected as a loading control. Three biological replicates are shown for each experiment. This figure is taken from (Best et al., 2014b).

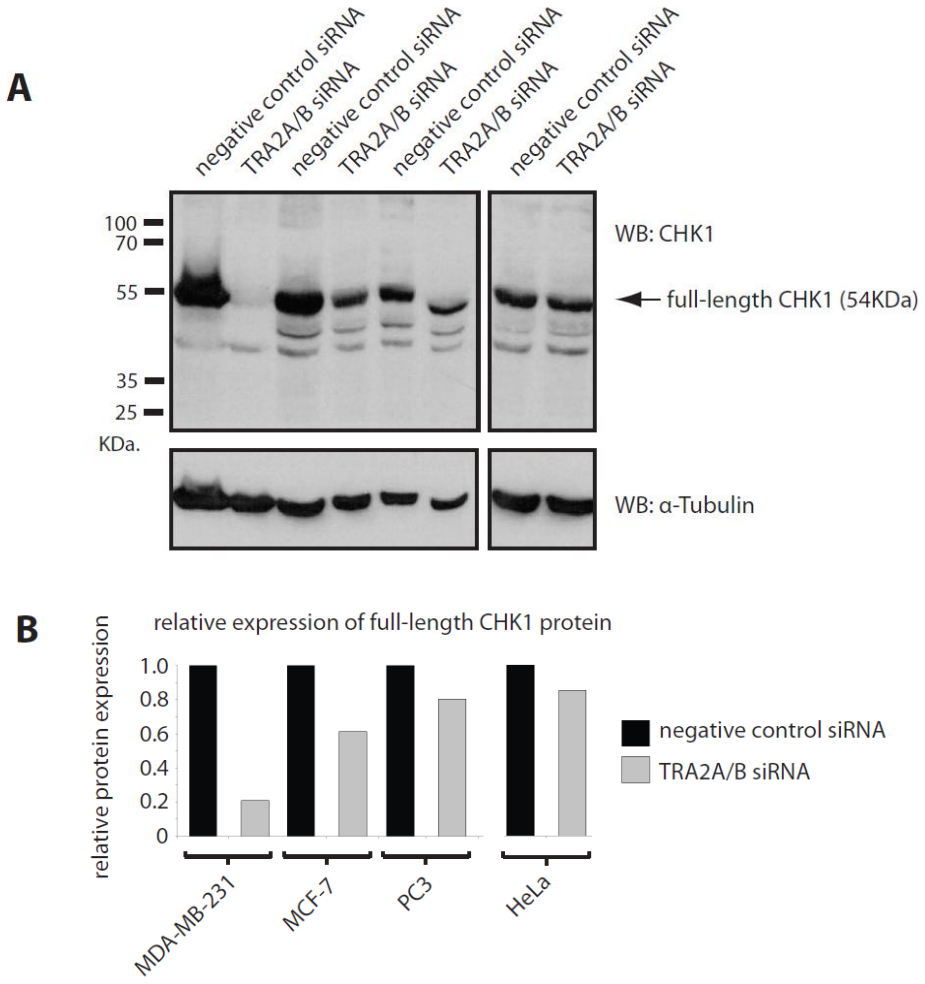


Figure 4.24 Tra2 protein depletion reduced full-length CHK1 protein expression (to varying degrees) in different cell lines. (A) Western Immunoblot showing detection of CHK1 protein in negative control siRNA and TRA2A/B siRNA treated cells across four different cell lines (MDA-MB-231, MCF-7, PC3 and HeLa cells). (B) Expression of full-length CHK1 protein (54kDa band) was quantified relative to expression of the α-Tubulin loading control, using ImageJ software. Data represents single biological samples. This figure is adapted from (Best et al., 2014b).

4.4.18 Tra2 protein depletion is associated with accumulation of the DNA damage marker phosphorylated Histone H2AX (γH2AX)

Previous studies have reported that depletion of full-length CHK1 protein is associated with an increase in replication stress and the accumulation of DNA damage, assessed by the presence of the DNA damage marker phosphorylated Histone H2AX (subsequently referred to as γH2AX) (Höglund et al., 2011). γH2AX is not a direct target for CHK1-mediated phosphorylation. However, it is is a well established marker of DNA damage which can occur following CHK1 protein depletion. Similar to previous

observations (Westendorf et al., 1998), I observed significantly increased levels of the DNA damage marker γH2AX following depletion of CHK1 protein by siRNA compared with cells treated with a negative control siRNA (Figure 4.23 middle panel, compare lanes 1-3 with lanes 4-6). Increased γH2AX levels were also observed after joint depletion of Tra2α and Tra2β (Figure 4.23 middle panel, compare lanes 1-3 with lanes 7-9). Consistent with reduced expression of full-length CHK1 protein following joint Tra2 protein depletion, the increase in the levels of γH2AX observed after CHEK1 or TRA2A/B siRNA transfection appear proportional to the reduction in full-length CHK1 protein expression.

To confirm that Tra2 protein depletion is associated with an increase in phosphorylation of Histone H2AX and not an increase in Histone H2AX expression, I monitored expression of phosphorylated γH2AX and total H2AX protein expression in Tra2 protein depleted cells by Western Immunoblot. Consistent with joint Tra2 protein depletion leading to phosphorylation of Histone H2AX, γH2AX expression strongly increased relative to total H2AX protein and α-Tubulin after Tra2 protein depletion in both MDA-MB-231 and MCF-7 cells (Figure 4.25).

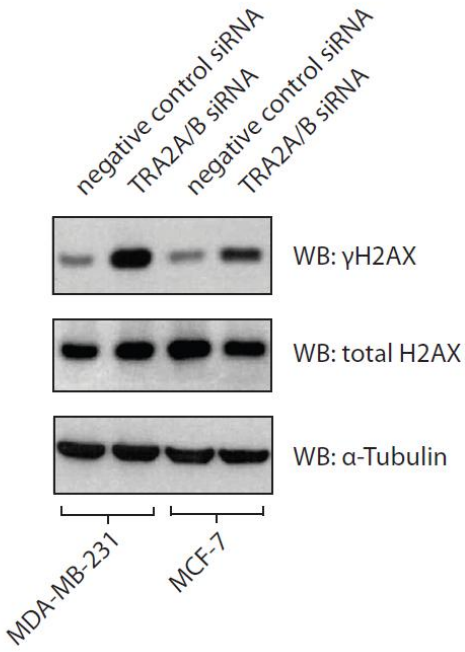


Figure 4.25 Expression of phosphorylated histone H2AX (γH2AX) and total histone H2AX protein after joint Tra2 protein depletion in MDA-MB-231 cells and MCF7 cells. The level of γH2AX significantly increased relative to total H2AX and α-Tubulin expression after joint Tra2 protein depletion in both cell lines. This figure is adapted from (Best et al., 2014b).

4.4.19 Joint Tra2 protein depletion is associated with an accumulation of nuclear γH2AX foci

Phosphorylation of H2AX plays an essential role in the recognition and repair of DNA double strand breaks (DSBs) (Pilch et al., 2003). H2AX molecules are rapidly phosphorylated at DNA break sites, facilitating the recruitment of DNA repair factors to nuclear foci to initiate DNA repair (Paull et al., 2000). A significant increase in phosphorylation of H2AX was observed by Western Immunoblot following joint Tra2 protein depletion. Next I monitored for the presence of γH2AX foci within the nucleus of Tra2 protein depleted cells using immunofluoresence. Expression of γH2AX was detected by immunofluoresence in MDA-MB-231 cells, 72 hours after siRNA transfection with either a negative control siRNA or TRA2A/B siRNA. Very weak expression of γH2AX could be detected in negative control siRNA treated cells (Figure 4.26, left hand panel). In contrast, and consistent with the increase in H2AX phosphorylation observed by Westen Immunoblotting, discrete γH2AX foci were clearly visible within the nucleus of TRA2A/B siRNA treated cells (Figure 4.26, right hand panel). Strong γH2AX foci expression was observed in approximately 50% of TRA2A/B siRNA treated cells.

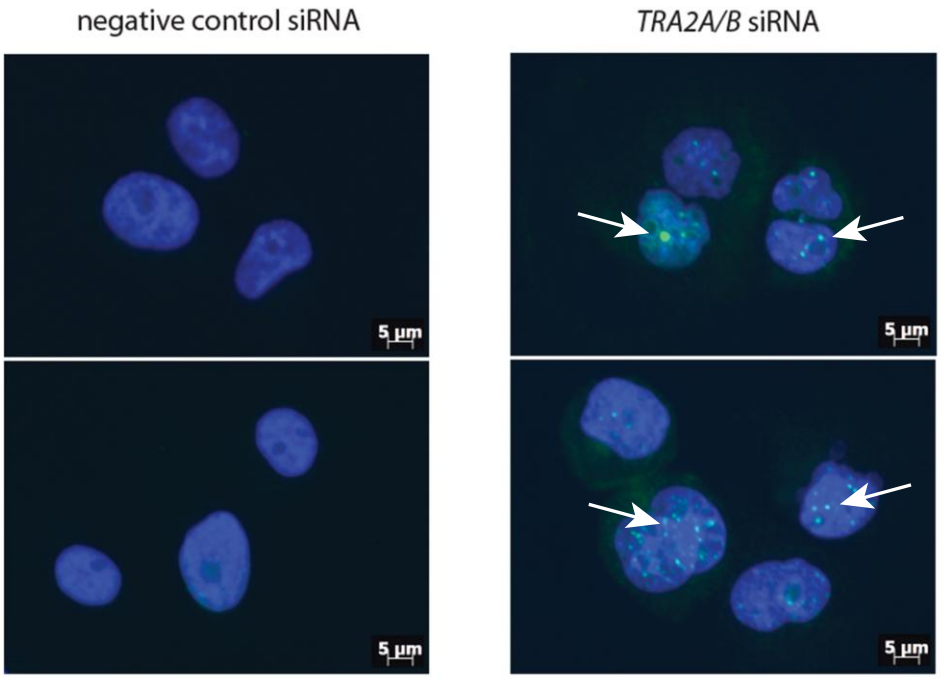


Figure 4.26 Joint Tra2 protein depletion is associated with an accumulation of γH2AX foci within the nucleus. Immunofluorescent detection of γH2AX (green) with DAPI staining (blue) in negative control siRNA treated cells and TRA2A/B siRNA treated MDA-MB-231 cells.

4.4.20 Joint Tra2 protein depletion is associated with abnormal nuclear morphology

Whilst observing the DAPI-stained nuclei of cells following siRNA transfection, I observed a significant number of cells with severely abnormal nuclear morphology 96 hours following TRA2A/B siRNA transfection. Typical circular nuclei observed in untreated or negative control siRNA treated MDA-MB-231 cells stained with DAPI are shown in Figure 4.27 A. Following joint Tra2 protein depletion, approximately 20% of cells contained severely abnormally shaped nuclei which were multi-lobed in appearance (Figure 4.27 B). Unlike the circular nuclei which contained discrete γH2AX foci after TRA2A/B siRNA transfection (Figure 4.26, right hand panel), some of the multi-lobed nuclei showed much stronger, more diffuse γH2AX staining (example shown in Figure 4.27 C). The abnormal, multi-lobed nuclei are consistent with severe biological defects following joint Tra2 protein depletion.

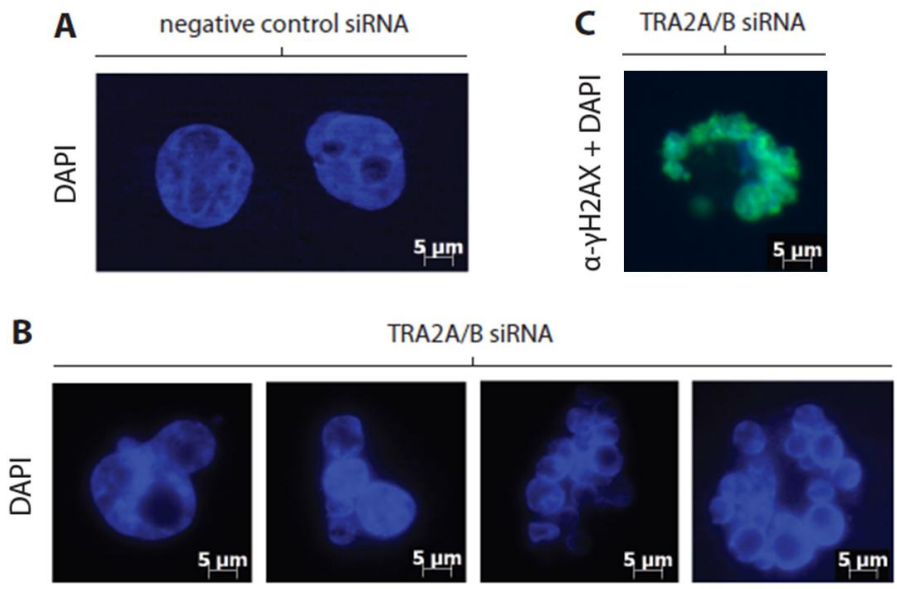


Figure 4.27 Joint Tra2 protein depletion is associated with abnormal nuclear morphology. (A) The typical circular nuclear morphology of negative control siRNA treated cells stained with DAPI. (B) Abnormal nuclear morphology with a multi-lobed appearance was observed in approximately 20% of TRA2A/B siRNA treated cells stained with DAPI. (C) Unlike the circular nuclei which contained discrete γH2AX foci after TRA2A/B siRNA transfection, the mutli-lobed nuclei showed a stronger, more diffuse staining specific for γH2AX (green). This figure is adapted from (Best et al., 2014b).

4.4.21 Joint Tra2 protein depletion reduced cell viability in MDA-MB-231 cells

I next assessed the effect of joint Tra2 protein depletion on viability of MDA-MB-231 cells using MTT assays. Cell viability was monitored by MTT assay every 24 hours following siRNA transfection for a period of 120 hours. Single depletion of either Tra2α or Tra2β protein had minimal effect on MDA-MB-231 cell viability over the 5 day time course. Conversely, joint Tra2 protein depletion strongly reduced cell viability (Figure 4.28). I observed similar results using two independent sets of siRNA, targeting different regions of the TRA2A and TRA2B mRNA. The reduction in cell viability observed after jointly depleting Tra2α and Tra2β, compared with the negligible effects of depleting either protein alone, suggest that Tra2α and Tra2β may also be functionally interchangeable in maintaining cell viability, as well as in splicing control in MDA-MB-231 cells.

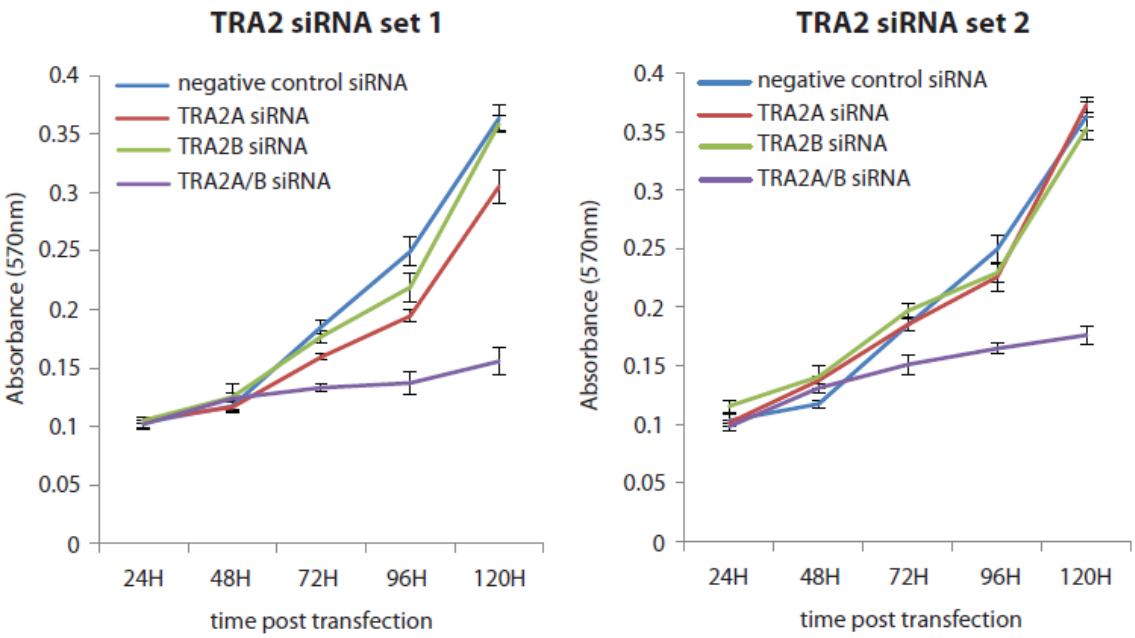


Figure 4.28 Joint (but not single) depletion of Tra2α and Tra2β reduced cell viability in MDA-MB-231 cells. Cell viability was monitored by MTT assay for a period of 120 hours following siRNA transfection. Data represents the mean of eight experimental replicates per time point ±s.e.m. A reduction in cell viability was confirmed using two independent sets of siRNA, targeting different regions of the TRA2A and TRA2B mRNAs. This figure is adapted from (Best et al., 2014b).

4.4.22 CHK1 protein depletion alone is sufficient to reduce cell viability in MDA-MB-231 cells

I then assessed the effect of CHK1 protein depletion on the viability of MDA-MB-231 cells using the MTT assay (Figure 4.29A), as well as by microscopy (Figure 4.29B). Depletion of CHK1 also reduced the viability of MDA-MB-231 cells to a similar extent as joint Tra2 protein depletion (Figure 4.29). This was also visible when comparing the density of cells by microscopy 120 hours after siRNA transfection (representative images shown in Figure 4.29 B). This data suggested that loss of full-length CHK1 protein expression would likely be sufficient by itself to contribute to the loss of cell viability observed after joint Tra2α and Tra2β protein depletion.

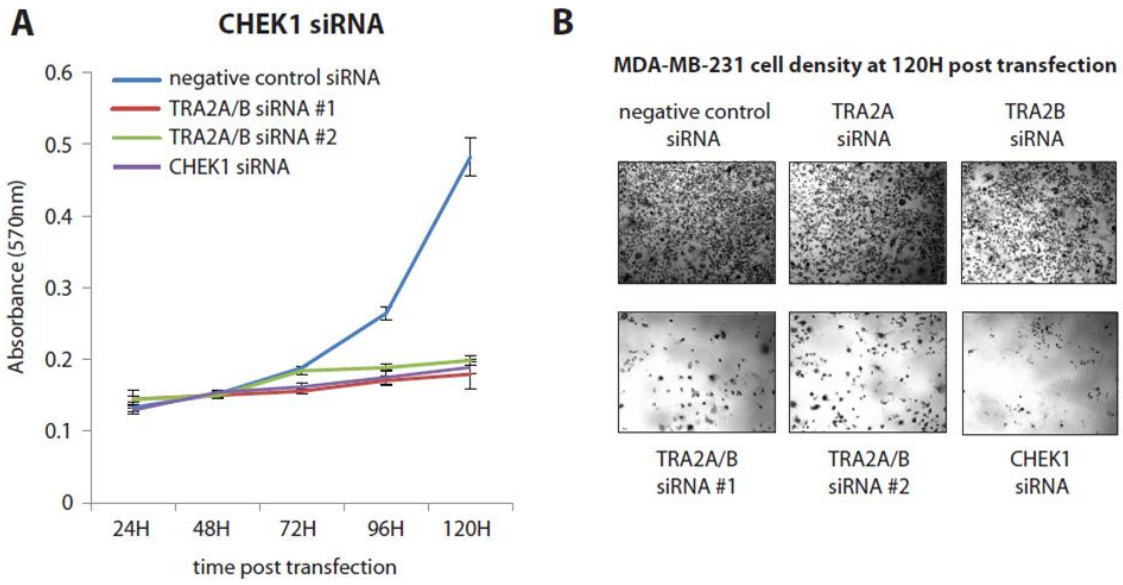


Figure 4.29 CHK1 protein depletion alone is sufficient to reduce cell viability in MDA-MB-231 cells. (A) Depletion of CHK1 protein reduced cell viability to a similar extent as joint Tra2 protein depletion in MDA-MB-231 cells. Cell viability was monitored by MTT assay for 120 hours following siRNA transfection. Data represents the mean of eight experimental replicates per time point ±s.e.m. (B) MDA-MB-231 cell density was observed by microscopy 120 hours after siRNA transfection. This figure is adapted from (Best et al., 2014b).

4.4.23 Joint Tra2 protein depletion significantly reduced cell proliferation in MDA-MB-231 cells

To determine whether the loss of cell viability observed after joint Tra2 protein depletion or CHK1 protein depletion was associated with reduced cell proliferation, I monitored the incorporation of the thymidine analogue EdU into DNA after siRNA

transfection, as a direct measurement of the number of cells initiating DNA replication. Cells were incubated with 10µM EdU for 4 hours (96 hours after siRNA transfection), and then immediately analysed using flow cytometry. Joint Tra2 protein depletion significantly reduced the proportion of EdU-positive cells when compared to negative control siRNA treated cells (the mean reduction was 8.4%, p=0.02) (Figure 4.30, compare top right panel to lower right panel). CHK1 protein depletion had no significant effect on the proportion of EdU-positive cells when compared to negative control siRNA treated cells (Figure 4.30, compare top right panel to lower left panel).

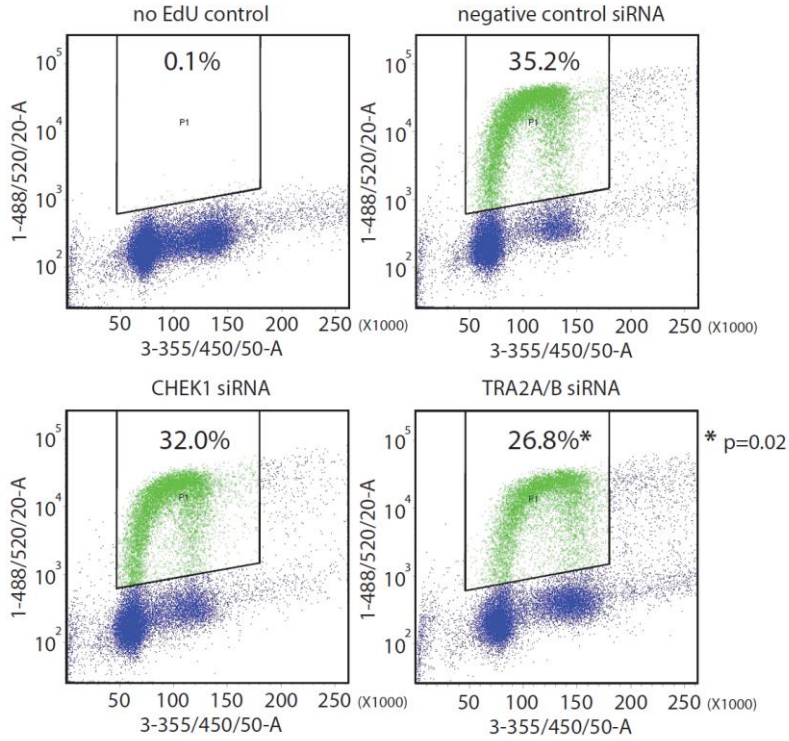


Figure 4.30 Joint Tra2 protein depletion significantly reduced cell proliferation in MDA-MB-231 cells. Cell proliferation was determined by monitoring incorporation of the thymidine analogue EdU over 4 hours, 96 hours after siRNA transfection. The proportion of EdU-positive cells was determined by FACS analysis, counting 30,000 cells per sample. Clockwise from top left, negative control siRNA treated cells incubated without EdU, negative control siRNA treated cells incubated with 10µM EdU, TRA2A/B siRNA treated cells incubated with 10µM EdU, CHEK1 siRNA treated cells incubated with 10µM EdU. The no EdU control was used to inform the gating strategy for EdU-positive cells. TRA2A/B siRNA transfection significantly reduced the proportion of EdU-positive cells when compared to negative control siRNA treated cells. Single CHK1 depletion had no significant effect compared to negative control siRNA treated cells. Data represents the mean of three biological replicates. Statistical significance was calculated using an independent two-sample t-test, where *p<0.05, **p<0.01, ***p<0.0001. This figure is adapted from (Best et al., 2014b).

4.4.24 Generation of a tetracycline-inducible full-length CHK1-FLAG protein expressing FLP-in HEK-293 cell line

The next step was to determine whether re-introduction of full-length CHK1 protein would rescue cell viability following joint Tra2 protein depletion. To do this, a FLP-in HEK-293 cell line was created in which a full-length FLAG-tagged CHK1 protein could be induced under control of a tetracycline promoter. This stable CHK1-FLAG FLP-in cell line was created entirely by Dr. Mahsa Kheirolahi-Kouhestani (Newcastle University). I confirmed efficient expression of the full-length CHK1-FLAG protein by Western Immunoblotting using a FLAG-specific antibody (Figure 4.31). Efficient expression of full-length CHK1-FLAG protein was observed 24 hours after tetracycline induction in the CHK1-FLAG FLP-in HEK-293 cells (Figure 4.31, lanes 1-4), but not in FLP-in HEK-293 cells transfected with an empty pcDNA5 vector (Figure 4.31, lanes 6-9). Minimal expression of the full-length CHK1-FLAG protein was detected in CHK1-FLAG FLP-in HEK-293 cells without tetracycline (Figure 4.31, compare lanes 4 and 5).

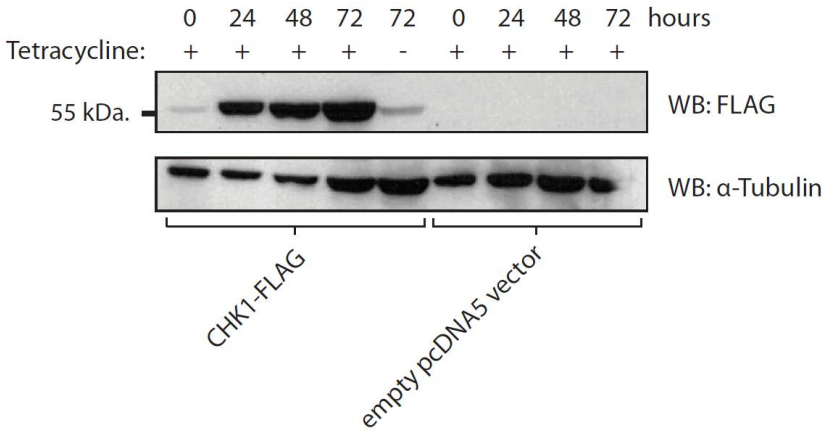


Figure 4.31 Generation of a tetracycline-inducible full-length CHK1-FLAG protein expressing cells using a FLP-in HEK-293 cell line. Efficient expression of the full-length CHK1-FLAG protein was confirmed by Western Immunoblotting using a FLAG-specific antibody. This figure is adapted from (Best et al., 2014b).

4.4.25 Re-introduction of full-length CHK1 protein expression was insufficient to rescue cell viability in joint Tra2 protein depleted HEK-293 cells

I then went on to assess the effect of joint Tra2 protein depletion on cell viability in the CHK1-FLAG FLP-in HEK-293 cell line, both with and without tetracycline, using MTT assays. Cells were treated with tetracycline or a mock treatment every 24 hours following siRNA transfection. Cell viability was monitored by MTT assay every 24 hours

for 120 hours following siRNA transfection. Similar to previous results, joint Tra2 protein depletion also reduced cell viability of the FLP-in HEK-293 cells, though to a lesser extent than that observed in the MDA-MB-231 cells (Figure 4.32 A). Although the full-length FLAG-tagged CHK1 protein was efficiently induced by tetracycline, it was not sufficient to rescue cell viability after joint Tra2 protein depletion (Figure 4.32 B). I confirmed knockdown of the Tra2β protein and induction of the full-length CHK1-FLAG protein in parallel experiments by Western Immunoblotting (Figure 4.32 C). This result is consistent with there being multiple essential exons controlled by Tra2 proteins (in addition to *CHEK1* exon 3) that are important for cell viability. However, it is impossible to conclusively rule out that induction of the full-length CHK1-FLAG protein failed to rescue viability for another reason (e.g. due to the timing of tetracycline addition, or differences in expression levels of the induced CHK1-FLAG protein and the endogenous CHK1 protein). Alternatively, the addition of the FLAG-tag may have compromised the function of the CHK1 protein.

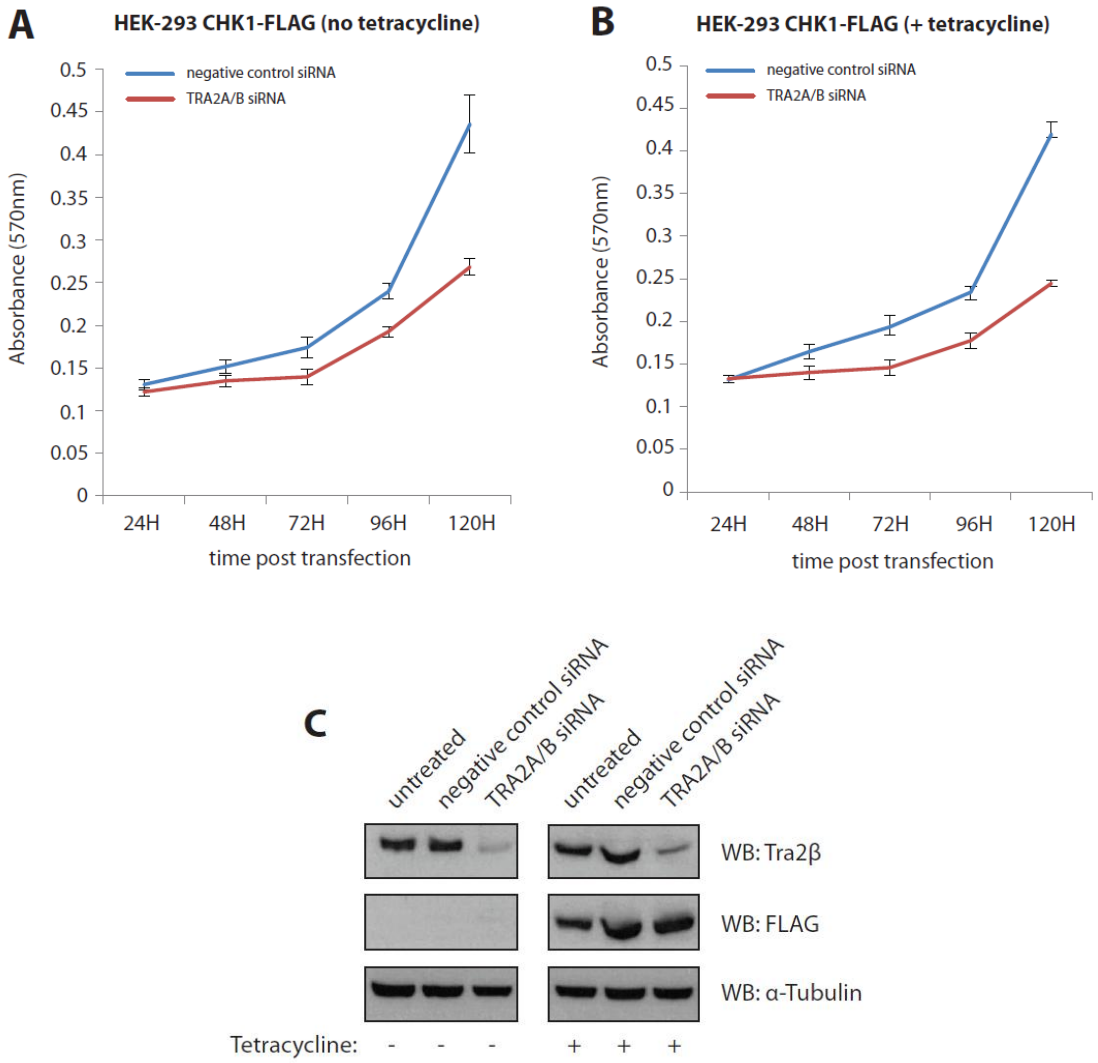


Figure 4.32 Induction of full-length CHK1-FLAG protein expression was insufficient to rescue cell viability following joint Tra2 protein depletion in FLP-in HEK-293 cells. (A) Joint Tra2 protein depletion reduced cell viability in control FLP-in HEK-293 cells. (B) Induction of full-length CHK1-FLAG protein expression with tetracycline was insufficient to rescue cell viability following joint Tra2 protein depletion in FLP-in HEK-293 cells. (C) In a parallel experiment, efficient expression of the full-length CHK1-FLAG protein and knockdown of Tra2β was confirmed by Western Immunoblotting. This figure is adapted from (Best et al., 2014b).

4.5 Discussion

In this chapter, RNA-seq was used to perform a global analysis of human Tra2α and Tra2β dependent exons in MDA-MB-231 cells. My data from chapter 3 revealed that Tra2β efficiently represses expression of Tra2α in MDA-MB-231 cells, and following Tra2β depletion, up-regulation of Tra2α provides an effective functional replacement in splicing regulation. Therefore prior to RNA-seq, joint Tra2 protein depletion was required to induce substantial changes in splicing of Tra2 protein dependent exons.

4.5.1 The *TRA2A* and *TRA2B* genes are differentially expressed in MDA-MB-231 cells

In chapter 3, I observed that whilst both Tra2 proteins were able to cross-regulate one another, this cross-regulation was largely asymmetrical in MDA-MB-231 cells. The *TRA2B* mRNA and Tra2β protein are expressed at much high levels than the *TRA2A* mRNA and Tra2α protein in MDA-MB-231 cells, as determined by RNA-seq and Western Immunoblotting respectively. Differential expression of the *TRA2A* and *TRA2B* genes may in part be established at the transcriptional level. However, increased *TRA2B* gene expression could equally repress both Tra2α and Tra2β protein expression, if both poison exons were equally responsive to the Tra2β protein. The combined iCLIP and RNA-seq data suggested that the *TRA2A* poison exon may be a stronger physiological target for Tra2β than the *TRA2B* poison exon, as a very similar quantity of iCLIP tags mapped to both poison exons despite much higher expression of *TRA2B*. Consequently, the *TRA2A* poison exon may be more sensitive to Tra2 protein expression, skewing cross-regulation in favour of Tra2β expression. Consistent with this hypothesis, the *TRA2A* poison exon has a higher density of Tra2β binding sites than the *TRA2B* poison exon (Grellscheid et al., 2011a). It is possible that a combination of differential gene expression at the transcriptional level (higher *TRA2B* expression) and a higher affinity of Tra2 proteins for the *TRA2A* poison exon than the *TRA2B* poison exon contribute to the repression of Tra2α in MDA-MB-231 cells.

4.5.2 Tra2 proteins regulate constitutively spliced target exons in addition to their known role in alternative splicing regulation

Previous studies have implicated Tra2α and Tra2β exclusively in alternative splicing regulation. However, my RNA-seq data uncovered several constitutively spliced exons

which were highly dependent on Tra2 protein expression for 100% splicing inclusion in MDA-MB-231 cells. It has been established that low level alternative splicing of constitutive exons can occur as a result of error prone exon recognition by the spliceosome (Melamud and Moult, 2009). However, consistent with the *ANKRD1* and *SMC4* exons being bona fide constitutive exons, both exons are not been annotated as alternatively spliced in any tissue, yet show very substantial splicing changes upon joint Tra2 protein depletion (-73 and -66 point PSI changes respectively). Interestingly, Tra2β is considered to be ubiquitously expressed, and is fairly evenly expressed across different mouse tissues, consistent with a more general role in constitutive splicing (Elliott et al., 2012). However, it remains possible that *Tra2b* is not expressed in specific cell types. Tra2 proteins may therefore share a greater degree of functionally similarity to the classical group of SR-proteins than previously thought. On the other hand, I only identified a relatively small number of Tra2 protein dependent constitutive exons, and the vast majority of constitutive exons were not affected by joint Tra2 depletion. Overall, this data is most consistent with Tra2 proteins regulating a relatively small number of constitutively spliced target exons. With the exception of the two *CD44* variable exons described earlier in chapter 3, inclusion of all other target exons was activated by Tra2 proteins (and not repressed) in this study. Although exon repression by Tra2 proteins has previously been described (McGlincy and Smith, 2008; Kaehler *et al.*, 2012), such exons might have either eluded the search criteria or occur less frequently.

4.5.3 Gene expression changes following joint Tra2 protein depletion

In addition to the alternative splicing changes observed in the transcriptome following joint Tra2 protein depletion, some statistically significant changes in gene expression were also identified and validated by qPCR. The GO enrichment analysis performed by Dr. Katherine James (Newcastle University) indicated that the differentially expressed genes were significantly enriched in cell cycle-related functions. The differentially expressed genes include increased expression of two members of the FOS gene family; *FOS* and *FOSB*. The FOS gene family encode leucine zipper proteins which dimersie with Jun proteins to form the AP-1 transcription factor complex (Milde-Langosch, 2005). AP-1 has been linked with a broad spectrum of cellular processes including cell transformation, proliferation and the induction of apoptosis (Ameyar et al., 2003).

Other gene expression changes include downregulation of *HIST2H2BE*, which encodes histone H2B which is ubiquitlyated during the DNA damage response (Shiloh et al., 2011). Gene expression changes were also validated for a number of lncRNAs which are not well characterised. As Tra2 proteins have never previously been linked with direct roles in transcriptional regulation and the associated changes in gene expression were not associated changes in alternative splicing of the gene, the gene expression changes may be caused through an indirect affect of joint Tra2 protein depletion (e.g. changes to cell growth). Subsequent investigations into the phenotypic consequences of joint Tra2 protein depletion revealed significant effects on cell growth and cell proliferation.

4.5.4 Tra2 proteins regulate splicing of genes associated with chromosome biology

The Gene Ontology (GO) enrichment analysis indicated that genes containing Tra2 protein dependent exons were enriched in processes involved in chromosome biology. However, a significant limitation of this analysis is that Gene Ontology annotations are known to be incomplete and can differ in accuracy (Keene, 2007). The GO analysis included all 53 validated targets; however validation of Tra2 dependent targets was not exhaustive in my study. Although the enriched terms were statistically significant, a more comprehensive GO analysis could benefit from a larger set of Tra2 dependent exons. Overall, this data lends further support to the association of particular splicing regulators with the regulation of coherent processes.

4.5.5 Highly responsive Tra2 protein dependent exons

Of the exons which showed the highest amplitude response to Tra2 protein depletion, 6 of the 17 exons were from genes involved in either chromosome structure or epigenetic regulation. Some of the highest amplitude targets include *MSL3* (41% reduction in PSI), the human ortholog of the *Drosophila melanogaster MSL3* gene which regulates chromatin remodelling during sex determination (Smith et al., 2005). *SMC4* (66% reduction in PSI) encodes a protein involved in DNA repair and chromosome condensation (Huang et al., 2013). *ANKRD1* (73% reduction in PSI) encodes a transcription factor which is a negative regulator of cardiac genes (Owen et al., 2011). *GLYR1* (41% reduction in PSI) encodes a protein cofactor involved in histone demethylation (Fang et al., 2013). *ZCCHC9* (68% reduction in PSI) encodes a zinc finger

protein which targets histone mRNAs for degradation (Schmidt et al., 2011). *ZCCHC7* (72% reduction in PSI) and *ZCCHC11* (46% reduction in PSI) also encode zinc finger proteins homologous to *ZCCHC9*, but with distinct roles in ncRNA metabolism (Hagan *et al.*, 2009; Fasken *et al.*, 2011; Schmidt *et al.*, 2011). The *MPHOSPH10* gene (42% reduction in PSI) encodes a protein involved in rRNA processing in interphase, and is associated with chromosomes during mitosis (Westendorf et al., 1998). The transcriptome-wide nature of the iCLIP and RNA-seq experiments uncovered multiple Tra2 protein dependent target exons which may be of interest for further study and characterisation. However due to time constraints, the exons described above were not investigated further in this study and I focused on characterisation of an exon from the *CHEK1* gene.

4.5.6 Tra2 proteins regulate *CHEK1* exon 3 and are for required for expression of full-length CHK1 protein

CHEK1 exon 3 was amongst the eleven Tra2 protein dependent exons from genes involved in chromosome biology. CHK1 plays a key role in the DNA damage response, as well as regulating cell cycle checkpoints in response to genotoxic stress (Xiao *et al.*, 2004; Llorian *et al.*, 2010; Pabla and Dong, 2012). The skipping of exon 3 has previously been reported to produce a truncated CHK1 protein isoform, termed CHK1-S (Pabla et al., 2012). Exon 3 of the *CHEK1* gene is 224 nucleotides long; hence omission of exon 3 produces a downstream shift in the reading frame of the *CHEK1* mRNA. An analysis of the *CHEK1* coding sequence suggested that if exon 3 were skipped, the downstream frameshift would create multiple premature termination codons (PTCs) within the *CHEK1* mRNA, beginning in exon 4 (see Figure 4.33).

Figure 4.33 Predicted coding sequence of the *CHEK1* gene when exon 3 is included (left hand side) or skipped (right hand side). Skipping of exon 3 is predicted to create a shift in the downstream reading frame, creating multiple premature termination codons (PTCs) within the *CHEK1* mRNA is the same translation initiation site is used. The translation initiation site is highlighted in green, exon 3 is highlighted in yellow and termination codons are highlighted in red.

Transcripts containing premature termination codons (PTCs) are targeted for nonsense-mediated decay (NMD) in the RNA degradation pathway. PTCs are recognised when located further than 50 nucleotides upstream of a splice site. Once a ribosome reaches a termination codon, UPF1 is recruited to the ribosome. Consequently, when a termination codon is upstream of a splice site, UPF1 interacts with the exon junction complex (EJC), resulting in degradation of the mRNA by exoRNases (Peebles and Finkbeiner, 2007). It is possible therefore that when the exon is skipped, PTCs created in exon 4 would target the *CHEK1* mRNA for NMD, rather than producing a functional transcript.

Pabla et al. suggest that CHK1-S is created when a downstream translation initiation site (TIS) is selected from exon 2 and exon 3 is skipped, producing a shorter in-frame mRNA encoding CHK1-S. CHK1-S was found to be an endogenous inhibitor of full-length CHK1 and during the unperturbed cell cycle, CHK1-S expression can inhibit full-length CHK1 allowing cell cycle progression from S-phase into G2/M (Pabla et al., 2012).

The combined iCLIP and RNA-seq data identified *CHEK1* exon 3 as a stong candidate for Tra2 protein dependent splicing regulation, and significant skipping of this exon was observed following joint Tra2 protein depletion in four different cell lines. Further evidence from EMSAs and minigene constructs confirmed *CHEK1* exon 3 as a direct target for splicing regulation by Tra2α and Tra2β. By Western Immunoblotting, a single major isoform of CHK1 was detected in MDA-MB-231 cells, corresponding to full-length CHK1 (CHK1-L) (54kDa) (Figure 4.23). The CHEK1 antibody used in this study (Proteintech 10362-1-AP) specifically recognises the C-terminus of the CHK1 protein and should therefore recognise both CHK1 protein isoforms (CHK1-L and CHK1-S). Upon longer chemiluminescent exposure, additional bands corresponding to shorter isoforms of the CHK1 protein could be detected (approximately 50kDa and 47kDa), though their relative expression is weak (Figure 4.24). Loss of full-length CHK1 protein expression following joint Tra2 protein depletion was not associated with an increase in expression of any shorter isoform of CHK1. Therefore overall, my data is most consistent with Tra2 proteins being required for expression of full-length CHK1 (CHK1-L), rather than inducing expression of a shorter CHK1 isoform (CHK1-S). The most likely explanation of this result may be that following joint Tra2 protein depletion, the same translation initiation site is used which producing a downstream frameshift when exon 3 is skipped, resulting in production of a non-coding transcript. Under different physiological conditions, skipping of exon 3 may coincide with selection of the downstream translation initiation site from exon 2, allowing the shorter CHK1 isoform (CHK1-S) to be produced (proposed model shown in Figure 4.34).

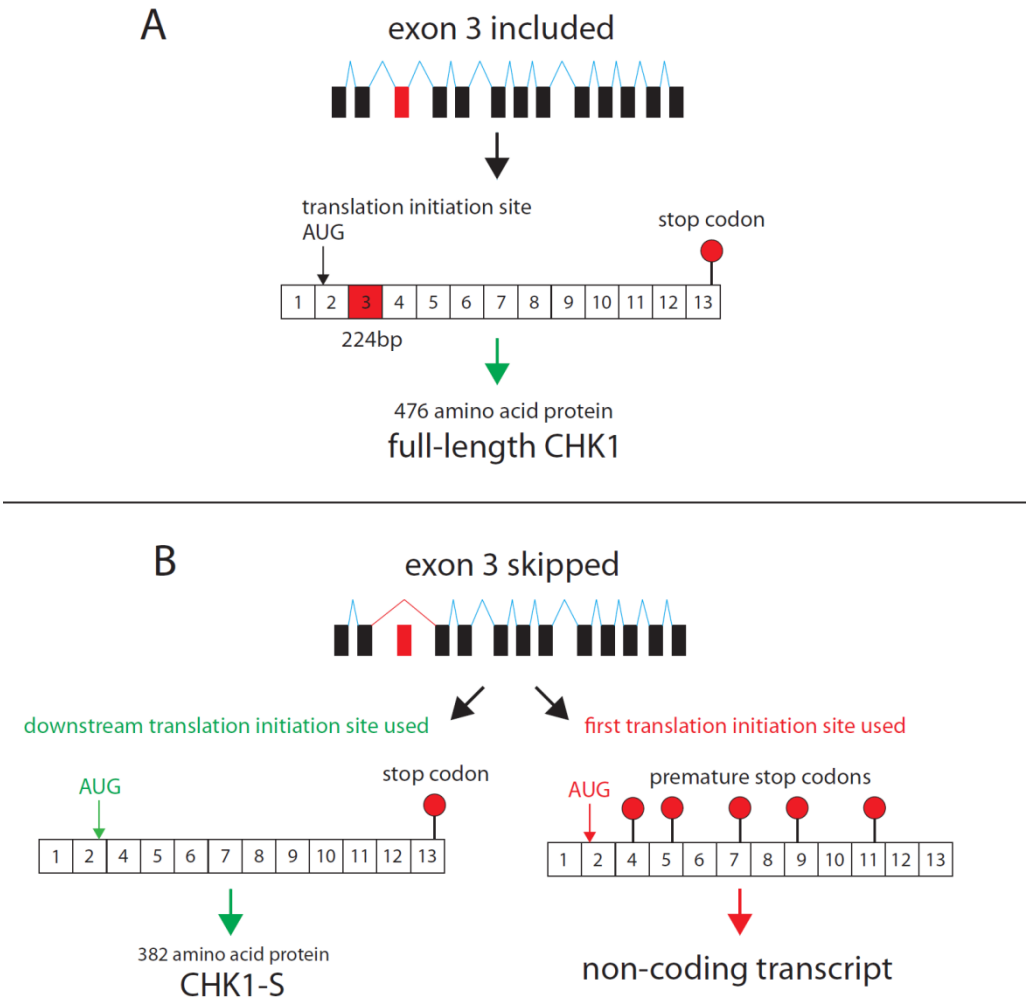


Figure 4.34 Proposed model of *CHEK1* alternative splicing. Skipping of exon 3 in the CHEK1 mRNA may create a shorter protein isoform or a non-coding transcript, depending on which translation initiation site is used. (A) Inclusion of exon 3 produces an in-frame mRNA encoding the full-length CHK1 protein. (B) If the initial translation initiation site is used, skipping of exon 3 would create a downstream frameshift leading to production of multiple PTCs within the *CHEK1* mRNA. Selection of an alternative downstream translation initiation site from exon 2 would produce a second in-frame transcript when exon 3 is skipped, encoding the shorter protein isoform of CHK1-S. This figure is adapted from (Pabla et al., 2012).

4.5.7 Tra2 protein depletion is associated with H2AX phosphorylation and reduced cell viabilty

Consistent with previous observations (Syljuåsen *et al.*, 2005; Gagou *et al.*, 2010), loss of full-length CHK1 protein expression (either via transfection with a CHEK1-specific siRNA or TRA2A/B siRNAs) led to a significant increase in phosphorylation of the DNA damage marker γH2AX. As no exogenous DNA damage insult was used in these

experiments, accumulation of γH2AX may result from failure to repair endogenous DNA damage caused by replication stress.

Replication stress is defined as the stalling of replication fork progression and/or DNA sysnthesis (Zeman and Cimprich, 2014). Replication stress may be exacerbated in cancer cells by factors including genomic instability and activation of oncogenes which promote DNA replication initiation or origin firing (Jones *et al.*, 2013; Srinivasan *et al.*, 2013). Stalled replication forks are associated with rapid phosphorylation of H2AX at DNA break sites, facilitating the recruitment of DNA repair factors to the point of DNA damage (Paull et al., 2000). Once a replication fork has stalled, the replication stress response is activated, leading to the recruitment of the replication stress response protein kinases ataxia-telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3 related (ATR) (Zeman and Cimprich, 2014). ATR directly phosphorylates CHK1 (Ser345), leading to stalled cell cycle progression and replication fork stabilisation (Nam and Cortez, 2011). Consequently, loss of CHK1 protein expression is associated with failure to repair replication stress-induced DNA damage leading to the accumulation of γH2AX foci (Gagou et al., 2010).

Joint depletion of Tra2 proteins was also associated with a reduction in MDA-MB-231 cell viabilty. Given CHK1 protein depletion also induced a similar reduction in cell viability; loss of productive *CHEK1* splicing would likely contribute to the loss of cell viabilty observed following joint Tra2 protein depeltion. However, re-introduction of full-length CHK1 protein failed to rescue viabilty of HEK-293 cells depleted of Tra2 proteins, suggesting Tra2 proteins may regulate other essential exons (as well as *CHEK1* exon 3) important for cell viability. Joint Tra2 protein depletion (but not CHK1 depletion alone) was also associated with reduced cell proliferation, consistent with Tra2 proteins regulating genes (other than *CHEK1*) important for cell proliferation. Consistent with an important role in maintaining cell viabilty, it was recently shown that Tra2β protein expression is essential for neuronal cell viabilty during murine development (Storbeck et al., 2014). *Tra2b* deficient mice showed severe abnormalities in cortical development caused by massive apoptotic events (Storbeck et al., 2014). Joint Tra2 protein depletion was associated with severely abnormally shaped cell nuclei, which were either lobed, multi-lobed or disintegrated in

appearance, consistent with major biological defects occurring following loss of Tra2 protein expression.

4.5.8 Chapter Summary

In this chapter, I used RNA-seq following joint Tra2α and Tra2β depletion to identify Tra2 dependent exons in MDA-MB-231 cells. Combining the RNA-seq data with Tra2β iCLIP data from chapter 3 facilitated the identification of exons which were both directly bound by Tra2β and functionally responsive to Tra2 protein depletion. Tra2 protein dependent exons included both alternative and constitutively spliced exons, and genes containing Tra2 protein dependent exons were enriched for functions associated with chromosome biology. Tra2 proteins were found to jointly regulate exon 3 of the *CHEK1* gene, which encodes the key DNA damage response protein CHK1. Joint depletion of Tra2 proteins reduced expression of full-length CHK1 protein, increased phosphorylation of the DNA damage marker γH2AX and reduced cell viability. Together, this data suggests Tra2 proteins jointly regulate both alternative and constitutive splicing patterns via paralog compensation which are important to maintain cell viability.

Chapter 5 Concluding remarks and future work

During the process of gene expression, RNAs undergo multiple processing and regulatory steps which are coordinated by RNA binding proteins (RBPs). By mapping RBP-RNA interactions and investigating their functional effect, the functions of many RBPs are beginning to be elucidated. In my PhD study, I aimed to identify novel RNA targets of the RBPs Tra2α and Tra2β, in an attempt to further understand the biological functions of these proteins in normal physiology and in disease.

This thesis includes data from multiple transcriptome-wide approaches, including data from HITS-CLIP, iCLIP and RNA-seq. The field of "ribonomics", in which genome-wide approaches are used to study how protein-RNA interactions regulate gene expression, has seen tremendous progress in recent years (König et al., 2012). Early ribonomic approaches such as RNA immunoprecipitation and microarray analyses (RIP-chip) were prone to detecting non-specific interactions and had low resolution (Mili and Steitz, 2004). However, such techniques have largely been superseded by a variety of UV crosslinking and immunoprecipitation (CLIP) techniques which provide greater resolution and more stringent purification. More recently, CLIP has been coupled to high-throughput sequencing (HITS-CLIP) to facilitate detection of protein-RNA interaction on a genome-wide scale (Darnell, 2010). Furthermore, with modifications to the original CLIP protocol, such as iCLIP (Konig et al., 2011) and PAR-CLIP (Hafner et al., 2010), genome-wide detection of protein-RNA interactions is now advancing towards individual nucleotide resolution. Excitingly, CLIP can be utilised to study the interaction of RBPs with a diverse range of transcripts, which include mRNAs (Licatalosi et al., 2008), miRNAs (Chi et al., 2009), snoRNAs and rRNAs (Granneman et al., 2009). However, a significant limitation of CLIP techniques and therefore of my study is that the number of sequencing reads or "CLIP tags" is strongly influenced by the expression level of the corresponding transcript. In the future, it is likely that normalising CLIP sequencing to gene expression profiles from RNA-seq will facilitate the development of "quantitative CLIP", which may prove a powerful tool for studying RBP-RNA interactions (König et al., 2012).

Regulated alternative splicing is particulary prevelant in the testis (Elliott and Grellscheid, 2006) and global splicing switches are reprogrammed during male meiosis

(Schmid et al., 2013). Tra2β is implicated in male germ cell development and is up-regulated during the progression from mitotic spermatagonia to meiotic spermatocytes (Schmid et al., 2013). To investigate Tra2β-regulated splicing during male germ cell development, a previous PhD student performed a Tra2β HITS-CLIP experiment to identify direct RNA targets of Tra2β (Grellscheid et al., 2011a). I used this dataset in the initial stages of my project to validate and characterise Tra2β-regulated exons from the mouse testis. Tra2β was found to regulate a meiotic isoform of the histone chaperone protein Nasp (termed tNASP). Interestingly, the tNASP protein isoform localises on the synaptonemal complex of meiotic chromosomes, where it is involved in cell cycle progression during meiosis (Alekseev et al., 2009). Hence up-regulation of Tra2β during male germ cell development may have a key physiological role in generating alternatively spliced isoforms important in male meiosis. Future studies in the lab aim to investigate the role of Tra2β in male germ cell development further and a testis-specific *Tra2b* knockout mice has recently been generated which may yield further insights into the role of Tra2β in male germ cell physiology.

The Tra2β HITS-CLIP study in the mouse testis also identified a poison exon within *Tra2a* which was a direct target of Tra2β. This data suggested that Tra2β may regulate expression of its paralogous protein Tra2α, by promoting inclusion of the poison exon within the *Tra2a* mRNA. Detection of endogenous Tra2α and Tra2β by Western blot confirmed this hypothesis; depletion of Tra2β led to a significant increase in Tra2α protein expression. Remarkably, up-regulation of Tra2α could largely compensate for loss of Tra2β in both splicing regulation and maintaining cell viability, a phenomenon termed “paralog compensation”. To date, the majority of studies on vertebrate Tra2 proteins have largely focused on Tra2β expression. However, this data suggests that future studies may benefit by considering the expression levels of both Tra2α and Tra2β in splicing regulation.

During my investigation into Tra2β-responsive minigenes, I unexpectedly found that co-transfection with a Tra2β protein isoform lacking the RS1 domain (Tra2βΔRS1) not only failed to activate splicing inclusion, but actively repressed splicing inclusion of exons normally activated by full-length Tra2β. Intruigingly, this data suggests that the endogenous Tra2β-1 and Tra2β-3 protein isoforms may be functionally distinct in

splicing regulation and may act in a competitive, antagonistic manner to either promote or repress inclusion of target exons respectively. Further investigation is required to determine whether this observation applies to endogenous Tra2β isoforms as well as over-expressed proteins.

Tra2β is up-regulated in several human cancers (reviewed by (Best et al., 2013)) and is important for cell growth in several cancer cell lines (Takeo *et al.*, 2009; Kajita *et al.*, 2013). Tra2β was found to be specifically up-regulated in invasive breast cancer and it has been postulated that induction of Tra2β expression in breast cancer cells may produce alternative splicing isoforms of the *CD44* gene associated with tumour progression and metastasis (Watermann et al., 2006). Therefore for the majority of this project, I aimed to identify RNA targets of Tra2β in the invasive breast cancer cell line MDA-MB-231, to explore the potential roles of Tra2β in cancer cell biology.

To identify Tra2β target exons in MDA-MB-231 cells, I used iCLIP to map the transcriptome-wide binding sites of Tra2β, and RNA-seq to detect changes in the transcriptome following joint Tra2 protein depletion. The development of next-generation sequencing technologies and whole transcriptome sequencing in particular has expanded the scale and speed in which the transcriptome can be analysed, which would not have been possible just a few years ago. One of the unexpected findings from the RNA-seq data was that Tra2-dependent exons include constitutively spliced target exons in addition to alternative exons.

Another finding from this study was that Tra2-dependent exons were enriched in genes involved in functionally related processes, specifically chromosome biology. This is consistent with other observations that RNA-binding proteins may regulate coherent cellular functions. For example, another recent study from the Elliott lab identified that the tissue-specific RBP T-STAR regulates alternative splicing of genes encoding synaptic proteins in the brain (Ehrmann et al., 2013). Similary, MBNL1 and RBFOX2 have been found to coordinate alternative splicing of genes involved in pluripotent stem cell differentiation (Venables et al., 2013c). Tra2β is not expressed in a tissue-specific manner (it is ubiquitously expressed), however it is possible that up-regulation of Tra2β protein expression during male germ cell development may alter the splicing profile of genes with coherent functions (e.g. chromosome biology) in a developmental-stage-specific manner.

One of the most interesting Tra2 target exons identified in this study was a functionally important exon within *CHEK1*, which encodes the key DNA damage response protein CHK1 (Chen and Sanchez, 2004). Joint Tra2 protein depletion resulted in skipping of *CHEK1* exon 3 and reduced expression of full-length CHK1 protein. Joint Tra2 protein depletion was also associated with accumulation of the DNA damage marker γH2AX and a significant loss of cell viability. As no exogenous DNA damage insult was used in these experiments, the accumulation of DNA damage may be the result of replication stress, caused by high rates of DNA replication in rapidly dividing cells. Replication stress refers to the accumulation of DNA damage caused by stalled DNA replication forks, which can occur due to deficines in the substrates required for replication (Zeman and Cimprich, 2014). Some proto-oncogenes including Myc (which is amplified in both MDA-MB-231 and MCF-7 cells) can initiate replication stress by driving rapid cell proliferation (Halazonetis *et al.*, 2008; Campaner and Amati, 2012; Schoppy *et al.*, 2012). In cancer cells, activation of the replication stress-signalling pathway involving the ATR and CHK1 kinases prevents cell cycle progression until the DNA damage has been repaired (Höglund et al., 2011). Hence although CHK1 can function as a tumour suppressor early in cell transformation (preventing accumulation of DNA damage), CHK1 activity can also enhance "tumour fitness" by repairing DNA damage caused by replication stress and preventing apoptosis (López-Contreras et al., 2012). Consequently, CHK1 inhibitors are currently being trialled as potential chemotherapeutic agents, both in combination therapy to potentiate the efficacy of other DNA-damaging chemotherapeutics or as single agent therapies (McNeely et al., 2014). This also raises the possibility of whether targeting Tra2 proteins or their RNA targets might be exploited therapeutically in cancer. For instance, Tra2 protein inhibition may sensitise cancer cells to chemotherapeutic agents which cause stalled replication forks and high levels of replication stress, such as hydroxycarbamide (hydroxyurea).

Cancer-specific splicing changes have huge potential to provide new prognostic markers in cancer, as well as providing novel therapeutic targets (Shkreta et al., 2013). The application of RNA-seq to analyse cancer cell transcriptomes will likely uncover many more cancer-specific splice variants, many of which may be cancer and sub-type specific (Eswaran et al., 2013). Further research into the transcriptomic landscape in cancer is likely to prove an exciting field of future research.

Bibliography

Alekseev, O.M., Bencic, D.C., Richardson, R.T., Widgren, E.E. and O'Rand, M.G. (2003) 'Overexpression of the Linker Histone-binding Protein tNASP Affects Progression through the Cell Cycle', *Journal of Biological Chemistry*, 278(10), pp. 8846-8852.

Alekseev, O.M., Richardson, R.T. and O'Rand, M.G. (2009) 'Linker Histones Stimulate HSPA2 ATPase Activity Through NASP Binding and Inhibit CDC2/Cyclin B1 Complex Formation During Meiosis in the Mouse', *Biology of Reproduction*, 81(4), pp. 739-748.

Alekseev OM, R.T.R., James K Tsuruta and Michael G O'Rand (2011) 'Depletion of the histone chaperone tNASP inhibits proliferation and induces apoptosis in prostate cancer PC-3 cells', *Reproductive Biology and Endocrinology* 9 (50).

Alhopuro, P., Björklund, M., Sammalkorpi, H., Turunen, M., Tuupanen, S., Biström, M., Niittymäki, I., Lehtonen, H.J., Kivioja, T., Launonen, V., Saharinen, J., Nousiainen, K., Hautaniemi, S., Nuorva, K., Mecklin, J.-P., Järvinen, H., Orntoft, T., Arango, D., Lehtonen, R., Karhu, A., Taipale, J. and Aaltonen, L.A. (2010) 'Mutations in the Circadian Gene CLOCK in Colorectal Cancer', *Molecular Cancer Research*, 8(7), pp. 952-960.

Ameyar, M., Wisniewska, M. and Weitzman, J.B. (2003) 'A role for AP-1 in apoptosis: the case for and against', *Biochimie*, 85(8), pp. 747-752.

Anczuków, O., Rosenberg, A.Z., Akerman, M., Das, S., Zhan, L., Karni, R., Muthuswamy, S.K. and Krainer, A.R. (2012) 'The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation', *Nat Struct Mol Biol*, 19(2), pp. 220-228.

Anders, S. and Huber, W. (2010) 'Differential expression analysis for sequence count data', *Genome Biology*, 11(10), p. R106.

Anders, S., Reyes, A. and Huber, W. (2012) 'Detecting differential usage of exons from RNA-seq data', *Genome Research*, 22(10), pp. 2008-2017.

Anderson, E.S., Lin, C.-H., Xiao, X., Stoilov, P., Burge, C.B. and Black, D.L. (2012) 'The cardiotonic steroid digitoxin regulates alternative splicing through depletion of the splicing factors SRSF3 and TRA2B', *RNA*, 18(5), pp. 1041-1049.

Anko, M.-L., Muller-McNicoll, M., Brandl, H., Curk, T., Gorup, C., Henry, I., Ule, J. and Neugebauer, K. (2012) 'The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes', *Genome Biology*, 13(3), p. R17.

Badens, C., Lacoste, C., Philip, N., Martini, N., Courrier, S., Giuliano, F., Verloes, A., Munnich, A., Leheup, B., Burglen, L., Odent, S., Van Esch, H. and Levy, N. (2006) 'Mutations in PHD-like domain of the ATRX gene correlate with severe psychomotor impairment and severe urogenital abnormalities in patients with ATRX syndrome', *Clinical Genetics*, 70(1), pp. 57-62.

Baker, K.E. and Parker, R. (2004) 'Nonsense-mediated mRNA decay: terminating erroneous gene expression', *Current Opinion in Cell Biology*, 16(3), pp. 293-299.

Baltz, Alexander G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C. and Landthaler, M. (2012) 'The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts', *Molecular Cell*, 46(5), pp. 674-690.

Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B.J. and Frey, B.J. (2010) 'Deciphering the splicing code', *Nature*, 465(7294), pp. 53-9.

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.

Beil, B., Screaton, G. and Stamm, S. (1997) 'Molecular cloning of htra2-β-1 and htra2-β-2, two human homologs of tra-2 generated by alternative splicing', *DNA CellBiol.*, (16), pp. 679–690.

Ben-Dov, C., Hartmann, B., Lundgren, J. and Valcárcel, J. (2008) 'Genome-wide Analysis of Alternative Pre-mRNA Splicing', *Journal of Biological Chemistry*, 283(3), pp. 1229-1233.

Benderska, N., Becker, K., Girault, J.-A., Becker, C.-M., Andreadis, A. and Stamm, S. (2010) 'DARPP-32 binds to tra2-beta1 and influences alternative splicing', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1799(5–6), pp. 448-453.

Berget, S.M. (1995) 'Exon Recognition in Vertebrate Splicing', *Journal of Biological Chemistry*, 270(6), pp. 2411-2414.

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.

Best, A., Dalgliesh, C., Kheirollahi-Kouhestani, M., Danilenko, M., Ehrmann, I., Tyson-Capper, A. and Elliott, D. (2014a) 'Tra2 protein biology and mechanisms of splicing control', *Biochemical Society Transactions* 42(4), pp. 1152-1158.

Best, A., James, K., Dalgliesh, C., Hong, E., Kheirolahi-Kouhestani, M., Curk, T., Xu, Y., Danilenko, M., Hussain, R., Keavney, B., Wipat, A., Klinck, R., Cowell, I., Lee, K.C., Austin, C., Venables, J., Chabot, B., Koref, M.S., Tyson-Capper, A. and Elliott, D. (2014b) 'Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons', *Nature Communications*. Sep 11; 5:4760. doi: 10.1038/ncomms5760

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

Blencowe, B.J. (2000) 'Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases', *Trends Biochem Sci*, 25, pp. 106 - 110.

Blencowe, B.J. (2006) 'Alternative splicing: new insights from global analyses', *Cell*, 126, pp. 37 - 47.

Boutz, P.L., Stoilov, P., Li, Q., Lin, C.-H., Chawla, G., Ostrow, K., Shiue, L., Ares, M. and Black, D.L. (2007) 'A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons', *Genes & Development*, 21(13), pp. 1636-1652.

Burdall, S., Hanby, A., Lansdown, M. and Speirs, V. (2003) 'Breast cancer cell lines: friend or foe?', *Breast Cancer Res*, 5(2), pp. 89 - 95.

Burge, C.B., Padgett, R.A. and Sharp, P.A. (1998) 'Evolutionary Fates and Origins of U12-Type Introns', *Molecular Cell*, 2(6), pp. 773-785.

Burge., G.Y.a.C.B. (2004) 'Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals.', *Journal of Computational Biology.*, 11(2-3), pp. 377-394.

Cáceres, J.F. and Kornblihtt, A.R. (2002) 'Alternative splicing: multiple control mechanisms and involvement in human disease', *Trends in Genetics*, 18(4), pp. 186-193.

Cailleau R, O.M., Cruciger QV. (1978a) 'Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization.', *In Vitro.*, 14, pp. 911-915.

Cailleau R, O.M., Cruciger QV. (1978b) 'Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization.', *In Vitro.* , 14(11), pp. 911-5.

Campaner, S. and Amati, B. (2012) 'Two sides of the Myc-induced DNA damage response: from tumor suppression to tumor maintenance', *Cell Division*, 7(1), p. 6.

Carlson, M. (2012) 'org.Hs.eg.db: Genome wide annotation for Human. R package version 2.8.0'.

Cartegni, L., Chew, S.L. and Krainer, A.R. (2002) 'Listening to silence and understanding nonsense: exonic mutations that affect splicing', *Nat Rev Genet*, 3(4), pp. 285-298.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. and Krainer, A.R. (2003) 'ESEfinder: a web resource to identify exonic splicing enhancers', *Nucleic Acids Research*, 31(13), pp. 3568-3571.

Chang, Y.-F., Imam, J.S. and Wilkinson, M.F. (2007) 'The Nonsense-Mediated Decay RNA Surveillance Pathway', *Annual Review of Biochemistry*, 76(1), pp. 51-74.

Chen, Y. and Sanchez, Y. (2004) 'Chk1 in the DNA damage response: conserved roles from yeasts to mammals', *DNA Repair*, 3(8–9), pp. 1025-1032.

Cheng, D., Côté, J., Shaaban, S. and Bedford, M.T. (2007) 'The Arginine Methyltransferase CARM1 Regulates the Coupling of Transcription and mRNA Processing', *Molecular Cell*, 25(1), pp. 71-83.

Chi, S.W., Zang, J.B., Mele, A. and Darnell, R.B. (2009) 'Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps', *Nature*, 460(7254), pp. 479-486.

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.

Citri, A. and Yarden, Y. (2006) 'EGF–ERBB signalling: towards the systems level', *Nat Rev Mol Cell Biol*, 7(7), pp. 505-516.

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M.F., Kellis, M., Lindblad-Toh, K. and Lander, E.S. (2007) 'Distinguishing protein-coding and noncoding genes in the human genome', *Proceedings of the National Academy of Sciences*, 104(49), pp. 19428-19433.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A. and Lawrence, J.B. (2009) 'An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles', *Molecular Cell*, 33(6), pp. 717-726.

Cléry, A., Jayne, S., Benderska, N., Dominguez, C., Stamm, S. and Allain, F.H.T. (2011) 'Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1', *Nat Struct Mol Biol*, advance online publication.

Colwill, K., Feng, L.L., Yeakley, J.M., Gish, G.D., Cáceres, J.F., Pawson, T. and Fu, X.-D. (1996) 'SRPK1 and Clk/Sty Protein Kinases Show Distinct Substrate Specificities for Serine/Arginine-rich Splicing Factors', *Journal of Biological Chemistry*, 271(40), pp. 24569-24575.

Cramer, P., Srebrow, A., Kadener, S., Werbajh, S., de la Mata, M., Melen, G., Nogués, G. and Kornblihtt, A.R. (2001) 'Coordination between transcription and pre-mRNA processing', *FEBS Letters*, 498(2–3), pp. 179-182.

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Graf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Langerod, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Borresen-Dale, A.-L., Brenton, J.D., Tavare, S., Caldas, C. and Aparicio, S. (2012) 'The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups', *Nature*, 486(7403), pp. 346-352.

Damianov, A. and Black, D.L. (2010) 'Autoregulation of Fox protein expression to produce dominant negative splicing factors', *RNA*, 16(2), pp. 405-416.

Daoud, R., Da Penha Berzaghi, M., Siedler, F., Hübener, M. and Stamm, S. (1999) 'Activity-dependent regulation of alternative splicing patterns in the rat brain', *European Journal of Neuroscience*, 11(3), pp. 788-802.

Darnell, R.B. (2010) 'HITS-CLIP: panoramic views of protein–RNA regulation in living cells', *Wiley Interdisciplinary Reviews - RNA*, 1(2), pp. 266-286.

Dauwalder, B., Amaya-Manzanares, F. and Mattox, W. (1996) 'A human homologue of the Drosophila sex determination factor transformer-2 has conserved splicing regulatory functions', *Proc Natl Acad Sci U S A*, 93(17), pp. 9004-9.

De La Fuente, R., Baumann, C. and Viveiros, M.M. (2011) 'Role of ATRX in chromatin structure and function: implications for chromosome instability and human disease', *Reproduction*, 142(2), pp. 221-234.

Dehm, S.M. and Tindall, D.J. (2011) 'Alternatively spliced androgen receptor variants', *Endocrine-Related Cancer*, 18(5), pp. R183-R196.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Roder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Bar, N.S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O.J., Park, E., Persaud, K., Preall, J.B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L.H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S.E., Hannon, G., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigo, R. and Gingeras, T.R. (2012) 'Landscape of transcription in human cells', *Nature*, 489(7414), pp. 101-8.

Doherty, J.K., Bond, C., Jardim, A., Adelman, J.P. and Clinton, G.M. (1999) 'The HER-2/neu receptor tyrosine kinase gene encodes a secreted autoinhibitor', *Proceedings of the National Academy of Sciences*, 96(19), pp. 10869-10874.

Dreszer, T.R., Karolchik, D., Zweig, A.S., Hinrichs, A.S., Raney, B.J., Kuhn, R.M., Meyer, L.R., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B., Pohl, A., Malladi, V.S., Li, C.H., Learned, K., Kirkup, V., Hsu, F., Harte, R.A., Guruvadoo, L., Goldman, M., Giardine, B.M., Fujita, P.A., Diekhans, M., Cline, M.S., Clawson, H., Barber, G.P., Haussler, D. and James Kent, W. (2012) 'The UCSC Genome Browser database: extensions and updates 2011', *Nucleic Acids Res*, 40(Database issue), pp. D918-23.

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.

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. and Grellscheid, S.N. (2006) 'Alternative RNA splicing regulation in the testis', *Reproduction*, 132(6), pp. 811-819.

Elliott, Y.L.a.D.J. (2010) 'Coupling genetics and post-genomic approaches to decipher the cellular splicing code at a systems-wide level', *Biochemical Society Transactions* 38, pp. 237-241.

Erkelenz, S., Mueller, W.F., Evans, M.S., Busch, A., Schöneweis, K., Hertel, K.J. and Schaal, H. (2013) 'Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms', *RNA*, 19(1), pp. 96-102.

Eswaran, J., Horvath, A., Godbole, S., Reddy, S.D., Mudvari, P., Ohshiro, K., Cyanam, D., Nair, S., Fuqua, S.A.W., Polyak, K., Florea, L.D. and Kumar, R. (2013) 'RNA sequencing of cancer reveals novel splicing alterations', *Sci. Rep.*, 3.

Falcon, S. and Gentleman, R. (2007) 'Using GOstats to test gene lists for GO term association', *Bioinformatics*, 23(2), pp. 257-258.

Fang, R., Chen, F., Dong, Z., Hu, D., Barbera, A.J., Clark, E.A., Fang, J., Yang, Y., Mei, P., Rutenberg, M., Li, Z., Zhang, Y., Xu, Y., Yang, H., Wang, P., Simon, M.D., Zhou, Q., Li, J., Marynick, M.P., Li, X., Lu, H., Kaiser, U.B., Kingston, R.E. and Shi, Y.G. (2013) 'LSD2/KDM1B and its cofactor NPAC/GLYR1 endow a structural and molecular model for regulation of H3K4 demethylation', *Mol Cell*, 49(3), pp. 558-70.

Fasken, M.B., Leung, S.W., Banerjee, A., Kodani, M.O., Chavez, R., Bowman, E.A., Purohit, M.K., Rubinson, M.E., Rubinson, E.H. and Corbett, A.H. (2011) 'Air1 Zinc Knuckles 4 and 5 and a Conserved IWRXY Motif Are Critical for the Function and Integrity of the Trf4/5-Air1/2-Mtr4 Polyadenylation (TRAMP) RNA Quality Control Complex', *Journal of Biological Chemistry*, 286(43), pp. 37429-37445.

Faustino, N.A. and Cooper, T.A. (2003) 'Pre-mRNA splicing and human disease', *Genes & Development*, 17(4), pp. 419-437.

Feldman, B.J. and Feldman, D. (2001) 'The development of androgen-independent prostate cancer', *Nat Rev Cancer*, 1(1), pp. 34-45.

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', *Oncology Reports*, 11(5), pp. 1085-1090.

Friedenson, B. (2007) 'The BRCA1/2 pathway prevents hematologic cancers in addition to breast and ovarian cancers', *BMC Cancer*, 7(1), p. 152.

Fu, K., Mende, Y., Bhetwal, B.P., Baker, S., Perrino, B.A., Wirth, B. and Fisher, S.A. (2012) 'Tra2β Protein Is Required for Tissue-specific Splicing of a Smooth Muscle Myosin Phosphatase Targeting Subunit Alternative Exon', *Journal of Biological Chemistry*, 287(20), pp. 16575-16585.

Gabriel, B., Hausen, A.Z., Bouda, J., Boudova, L., Koprivova, M., Hirschfeld, M., JÄGer, M. and Stickeler, E. (2009) 'Significance of nuclear hTra2-beta1 expression in cervical cancer', *Acta Obstetricia et Gynecologica Scandinavica*, 88(2), pp. 216-221.

Gagou, M.E., Zuazua-Villar, P. and Meuth, M. (2010) 'Enhanced H2AX Phosphorylation, DNA Replication Fork Arrest, and Cell Death in the Absence of Chk1', *Molecular Biology of the Cell*, 21(5), pp. 739-752.

Ganesh, L., Yoshimoto, T., Moorthy, N.C., Akahata, W., Boehm, M., Nabel, E.G. and Nabel, G.J. (2006) 'Protein Methyltransferase 2 Inhibits NF-κB Function and Promotes Apoptosis', *Molecular and Cellular Biology*, 26(10), pp. 3864-3874.

Ge, H., Zuo, P. and L. Manley, J. (1991) 'Primary structure of the human splicing factor asf reveals similarities with drosophila regulators', *Cell*, 66(2), pp. 373-382.

Gelinas, R.E. and Roberts, R.J. (1977) 'One predominant 5'-undecanucleotide in adenovirus 2 late messenger RNAs', *Cell*, 11(3), pp. 533-544.

Gentleman, R. (2004) 'Bioconductor: open software development for computational biology and bioinformatics.', *Genome Biology*, 5(10).

Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. and Zhang, J. (2004) 'Bioconductor: open software development for computational biology and bioinformatics', *Genome Biology*, 5(10), p. R80.

Gibbons, R.J., Wada, T., Fisher, C.A., Malik, N., Mitson, M.J., Steensma, D.P., Fryer, A., Goudie, D.R., Krantz, I.D. and Traeger-Synodinos, J. (2008) 'Mutations in the chromatin-associated protein ATRX', *Human Mutation*, 29(6), pp. 796-802.

Glatz, D.C., Rujescu, D., Tang, Y., Berendt, F.J., Hartmann, A.M., Faltraco, F., Rosenberg, C., Hulette, C., Jellinger, K., Hampel, H., Riederer, P., Möller, H.-J., Andreadis, A., Henkel, K. and Stamm, S. (2006) 'The alternative splicing of tau exon 10 and its regulatory proteins CLK2 and TRA2-BETA1 changes in sporadic Alzheimer's disease', *Journal of Neurochemistry*, 96(3), pp. 635-644.

Gopinath, S. (2007) 'Methods developed for SELEX', *Analytical and Bioanalytical Chemistry*, 387(1), pp. 171-182.

Graham, F.L., Smiley, J., Russell, W.C. and Nairn, R. (1977) 'Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5', *Journal of General Virology*, 36(1), pp. 59-72.

Granneman, S., Kudla, G., Petfalski, E. and Tollervey, D. (2009) 'Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs',

Proceedings of the National Academy of Sciences of the United States of America - PNAS, 106(Article), pp. 9613-9618.

Graveley, B.R., Hertel, K.J. and Maniatis, T. (2001) 'The role of U2AF35 and U2AF65 in enhancer-dependent splicing', *RNA*, 7(6), pp. 806-818.

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. (2011a) 'Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2beta in development', *PLoS Genet*, 7(12), p. e1002390.

Grellscheid, S.N., Dalgliesh, C., Rozanska, A., Grellscheid, D., Bourgeois, C.F., Stévenin, J. and Elliott, D.J. (2011b) 'Molecular design of a splicing switch responsive to the RNA binding protein Tra2β', *Nucleic Acids Research*, 39(18), pp. 8092-8104.

Gutschner, T., Hämmerle, M., Eißmann, M., Hsu, J., Kim, Y., Hung, G., Revenko, A., Arun, G., Stentrup, M., Groß, M., Zörnig, M., MacLeod, A.R., Spector, D.L. and Diederichs, S. (2013) 'The Noncoding RNA MALAT1 Is a Critical Regulator of the Metastasis Phenotype of Lung Cancer Cells', *Cancer Research*, 73(3), pp. 1180-1189.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., Ulrich, A., Wardle, G.S., Dewell, S., Zavolan, M. and Tuschl, T. (2010) 'PAR-CliP - A Method to Identify Transcriptome-wide the Binding Sites of RNA Binding Proteins', (41), p. e2034.

Hagan, J.P., Piskounova, E. and Gregory, R.I. (2009) 'Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells', *Nat Struct Mol Biol*, 16(10), pp. 1021-5.
Hahn, W.C. and Weinberg, R.A. (2002) 'Rules for Making Human Tumor Cells', *New England Journal of Medicine*, 347(20), pp. 1593-1603.

Halazonetis, T.D., Gorgoulis, V.G. and Bartek, J. (2008) 'An Oncogene-Induced DNA Damage Model for Cancer Development', *Science*, 319(5868), pp. 1352-1355.

Hanahan, D. and Weinberg, Robert A. (2011) 'Hallmarks of Cancer: The Next Generation', *Cell*, 144(5), pp. 646-674.

Hansen, K.D., Brenner, S.E. and Dudoit, S. (2010) 'Biases in Illumina transcriptome sequencing caused by random hexamer priming', *Nucleic Acids Research*, 38(12), p. e131.

Harrison, P.M., Kumar, A., Lang, N., Snyder, M. and Gerstein, M. (2002) 'A question of size: the eukaryotic proteome and the problems in defining it', *Nucleic Acids Research*, 30(5), pp. 1083-1090.

Heckman, K.L. and Pease, L.R. (2007) 'Gene splicing and mutagenesis by PCR-driven overlap extension', *Nat. Protocols*, 2(4), pp. 924-932.

Hendzel MJ1, W.Y., Mancini MA, Van Hooser A, Ranalli T, Brinkley BR, Bazett-Jones DP, Allis CD. (1997) 'Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation.', *Chromosoma*, 106(6), pp. 348-60.

Hirschfeld, M., Jaeger, M., Buratti, E., Stuani, C., Grueneisen, J., Gitsch, G. and Stickeler, E. (2011) 'Expression of tumor-promoting Cyr61 is regulated by hTRA2-β1 and acidosis', *Human Molecular Genetics*, 20(12), pp. 2356-2365.

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-β1', *Human Molecular Genetics*, 11(17), pp. 2037-2049.

Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D. and Brown, P.O. (2008) 'Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system', *PLoS Biol*, 6(10), p. e255.

Höglund, A., Nilsson, L.M., Muralidharan, S.V., Hasvold, L.A., Merta, P., Rudelius, M., Nikolova, V., Keller, U. and Nilsson, J.A. (2011) 'Therapeutic Implications for the Induced Levels of Chk1 in Myc-Expressing Cancer Cells', *Clinical Cancer Research*, 17(22), pp. 7067-7079.

Holliday, D. and Speirs, V. (2011) 'Choosing the right cell line for breast cancer research', *Breast Cancer Research*, 13(4), p. 215.

Holly, A.C., Melzer, D., Pilling, L.C., Fellows, A.C., Tanaka, T., Ferrucci, L. and Harries, L.W. (2013) 'Changes in splicing factor expression are associated with advancing age in man', *Mechanisms of Ageing and Development*, 134(9), pp. 356-366.

Hsu, S. (2008) *Venn diagram for economics*. Available at: http://infoproc.blogspot.co.uk/2008/11/venn-diagram-for-economics.html.

Huang, K., Jia, J., Wu, C., Yao, M., Li, M., Jin, J., Jiang, C., Cai, Y., Pei, D., Pan, G. and Yao, H. (2013) 'Ribosomal RNA gene transcription mediated by the master genome regulator protein CTCF is negatively regulated by condensin complex', *J Biol Chem*.

Hurwitz, J. (2005) 'The Discovery of RNA Polymerase', *Journal of Biological Chemistry*, 280(52), pp. 42477-42485.

Jackson, C., Browell, D., Gautrey, H. and Tyson-Capper, A. (2013) 'Clinical Significance of HER-2 Splice Variants in Breast Cancer Progression and Drug Resistance', *International Journal of Cell Biology*, 2013, p. 8.

Jensen, O. (2004) 'Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry', *Current Opinion in Chemical Biology*, 8(1), pp. 33-41.

Jiang, Z., Tang, H., Havlioglu, N., Zhang, X., Stamm, S., Yan, R. and Wu, J.Y. (2003) 'Mutations in Tau Gene Exon 10 Associated with FTDP-17 Alter the Activity of an Exonic Splicing Enhancer to Interact with Tra2β', *Journal of Biological Chemistry*, 278(21), pp. 18997-19007.

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, pp. 2141 - 2144.

Jones, R.M., Mortusewicz, O., Afzal, I., Lorvellec, M., Garcia, P., Helleday, T. and Petermann, E. (2013) 'Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress', *Oncogene*, 32(32), pp. 3744-3753.

Kaehler, C., Isensee, J., Nonhoff, U., Terrey, M., Hucho, T., Lehrach, H. and Krobitsch, S. (2012) 'Ataxin-2-like is a regulator of stress granules and processing bodies', *PLoS One*, 7(11), p. e50134.

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 2β gene in human colon cancer cells', *Journal of Gastroenterology*, 48(11), pp. 1222-1233.

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.

Karunakaran, D.K.P., Banday, A.R., Wu, Q. and Kanadia, R. (2013) 'Expression Analysis of an Evolutionarily Conserved Alternative Splicing Factor, Sfrs10, in Age-Related Macular Degeneration', *PLoS One*, 8(9), p. e75964.

Keene, J.D. (2007) 'RNA regulons: coordination of post-transcriptional events', *Nat Rev Genet*, 8(7), pp. 533-43.

Kengo Tsuda, T.S., 1 Kanako Kuwasako,1 Mari Takahashi,1 Fahu He,1 Satoru Unzai,2 Makoto Inoue,1 Takushi Harada,1 Satoru Watanabe,1 Takaho Terada,1 Naohiro Kobayashi,1 Mikako Shirouzu,1 Takanori Kigawa,1 Akiko Tanaka,1 Sumio Sugano,3 Peter Güntert,1,4,5 Shigeyuki Yokoyama,1,6* and Yutaka Muto1 (2011) 'Structural basis for the dual RNA-recognition modes of human Tra2-b RRM', *Nucleic Acids Research*, 39(4), pp. 1538-1553.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S. (2013) 'TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions', *Genome Biology*, 14(4), p. R36.

Kiryu-Seo, S., Matsuo, N., Wanaka, A., Ogawa, S., Tohyama, M. and Kiyama, H. (1998) 'A sequence-specific splicing activator, Tra2[beta], is up-regulated in response to nerve injury', *Molecular Brain Research*, 62(2), pp. 220-223.

Kohli, M. and Tindall, D.J. (2010) 'New Developments in the Medical Management of Prostate Cancer', *Mayo Clinic Proceedings*, 85(1), pp. 77-86.

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. 119-124.

Kolb, S.J. and Kissel, J.T. (2011) 'Spinal Muscular Atrophy: A Timely Review', *Arch Neurol*, p. archneurol.2011.74.

König, J., Zarnack, K., Luscombe, N.M. and Ule, J. (2012) 'Protein–RNA interactions: new genomic technologies and perspectives', *Nat Rev Genet*, 13(2), pp. 77-83.

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. 909-915.

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. and Ule, J. (2011) 'iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution', *J Vis Exp*, (50).

Krcmery, J., Camarata, T., Kulisz, A. and Simon, H.-G. (2010) 'Nucleocytoplasmic functions of the PDZ-LIM protein family: new insights into organ development', *BioEssays*, 32(2), pp. 100-108.

Krueger, F. *Trim Galore!* Available at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

Lander, E.S. (2011) 'Initial impact of the sequencing of the human genome', *Nature*, 470(7333), pp. 187-197.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) 'Ultrafast and memory-efficient alignment of short DNA sequences to the human genome', *Genome Biol*, 10(3), p. R25.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C. and Brenner, S.E. (2007) 'Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements', *Nature*, 446(7138), pp. 926-9.

Lasfargues (1958) 'Cultivation of human breast carcinomas.', *J Natl Cancer Inst*, 21, pp. 1131-1147.

Lewis, B.P., Green, R.E. and Brenner, S.E. (2003) 'Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans', *Proceedings of the National Academy of Sciences of the United States of America*, 100(1), pp. 189-192.

Li, L., Feng, T., Lian, Y., Zhang, G., Garen, A. and Song, X. (2009) 'Role of human noncoding RNAs in the control of tumorigenesis', *Proceedings of the National Academy of Sciences*, 106(31), pp. 12956-12961.

Li, S.-J., Qi, Y., Zhao, J.-J., Li, Y., Liu, X.-Y., Chen, X.-H. and Xu, P. (2013) 'Characterization of Nuclear Localization Signals (NLSs) and Function of NLSs and Phosphorylation of Serine Residues in Subcellular and Subnuclear Localization of Transformer-2β (Tra2β)', *Journal of Biological Chemistry*, 288(13), pp. 8898-8909.

Liao, H., Winkfein, R.J., Mack, G., Rattner, J.B. and Yen, T.J. (1995) 'CENP-F is a protein of the nuclear matrix that assembles onto kinetochores at late G2 and is rapidly degraded after mitosis', *The Journal of Cell Biology*, 130(3), pp. 507-518.

Licatalosi, D.D. and Darnell, R.B. (2010) 'RNA processing and its regulation: global insights into biological networks', *Nat Rev Genet*, 11(1), pp. 75-87.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., Darnell, J.C. and Darnell, R.B. (2008) 'HITS-CLIP yields genome-wide insights into brain alternative RNA processing', *Nature*, 456(7221), pp. 464-469.

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

Liu, H.-X., Cartegni, L., Zhang, M.Q. and Krainer, A.R. (2001) 'A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes', *Nat Genet*, 27(1), pp. 55-58.

Liu, Y. (2009) *Identification and functional dissection of physiological targets of the RNA binding proteins RBMY, hnRNP G-T and T-STAR*. Newcastle University.

Llorian, M., Schwartz, S., Clark, T.A., Hollander, D., Tan, L.Y., Spellman, R., Gordon, A., Schweitzer, A.C., de la Grange, P., Ast, G. and Smith, C.W. (2010) 'Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB', *Nat Struct Mol Biol*, 17(9), pp. 1114-23.

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.

López-Contreras, A.J., Gutierrez-Martinez, P., Specks, J., Rodrigo-Perez, S. and Fernandez-Capetillo, O. (2012) 'An extra allele of Chk1 limits oncogene-induced replicative stress and promotes transformation', *The Journal of Experimental Medicine*, 209(3), pp. 455-461.

Lopez, A.J. (1998) 'ALTERNATIVE SPLICING OF PRE-mRNA: Developmental Consequences and Mechanisms of Regulation', *Annual Review of Genetics*, 32(1), pp. 279-305.

Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R. and Misteli, T. (2011) 'Epigenetics in Alternative Pre-mRNA Splicing', *Cell*, 144(1), pp. 16-26.

Lukas, J., Gao, D.-Q., Keshmeshian, M., Wen, W.-H., Tsao-Wei, D., Rosenberg, S. and Press, M.F. (2001) 'Alternative and Aberrant Messenger RNA Splicing of the mdm2 Oncogene in Invasive Breast Cancer', *Cancer Research*, 61(7), pp. 3212-3219.

Ly, T. (2014) *A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells*.

Mardon, H.J., Sebastio, G. and Baralle, F.E. (1987) 'A role for exon sequences in alternative splicing of the human fibronection gene', *Nucleic Acids Research*, 15(19), pp. 7725-7733.

Martin Dutertre, S.V.a.D.A., * (2010) 'Alternative splicing and breast cancer', *RNA Biology*, 7(4), pp. 403-411.

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. 123–147.

Matlin, A.J., Clark, F. and Smith, C.W.J. (2005) 'Understanding alternative splicing: towards a cellular code', *Nat Rev Mol Cell Biol*, 6(5), pp. 386-398.

Matsui, M., Horiguchi, H., Kamma, H., Fujiwara, M., Ohtsubo, R. and Ogata, T. (2000) 'Testis- and developmental stage-specific expression of hnRNP A2/B1 splicing isoforms, B0a/b', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1493(1–2), pp. 33-40.

Matsuo, N., Ogawa, S., Imai, Y., Takagi, T., Tohyama, M., Stern, D. and Wanaka, A. (1995) 'Cloning of a Novel RNA Binding Polypeptide (RA301) Induced by Hypoxia/Reoxygenation', *Journal of Biological Chemistry*, 270(47), pp. 28216-28222.

McCarthy, P.L., Mercer, F.C., Savicky, M.W.J., Carter, B.A., Paterno, G.D. and Gillespie, L.L. (2008) 'Changes in subcellular localisation of MI-ER1[alpha], a novel oestrogen receptor-[alpha] interacting protein, is associated with breast cancer progression', *Br J Cancer*, 99(4), pp. 639-646.

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

McNeely, S., Beckmann, R. and Bence Lin, A.K. (2014) 'CHEK again: Revisiting the development of CHK1 inhibitors for cancer therapy', *Pharmacology & Therapeutics*, 142(1), pp. 1-10.

Mee Young Kim, J.H.S.J. (2009) 'Emerging roles of RNA and RNA-binding protein network in cancer cells', *BMB reports online*, 42(3), p. 125~130.

Melamud, E. and Moult, J. (2009) 'Stochastic noise in splicing machinery', *Nucleic Acids Res*, 37(14), pp. 4873-86.

Mende, Y., Jakubik, M., Riessland, M., Schoenen, F., Rossbach, K., Kleinridders, A., Kohler, C., Buch, T. and Wirth, B. (2010a) '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.

Mende, Y., Jakubik, M., Riessland, M., Schoenen, F., Roßbach, K., Kleinridders, A., Köhler, C., Buch, T. and Wirth, B. (2010b) 'Deficiency of the splicing factor Sfrs10 results in early embryonic lethality in mice and has no impact on full-length SMN/Smn splicing', *Human Molecular Genetics*, 19(11), pp. 2154-2167.

Meyer, L.R., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Kuhn, R.M., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B., Raney, B.J., Pohl, A., Malladi, V.S., Li, C.H., Lee, B.T., Learned, K., Kirkup, V., Hsu, F., Heitner, S., Harte, R.A., Haeussler, M., Guruvadoo, L., Goldman, M., Giardine, B.M., Fujita, P.A., Dreszer, T.R., Diekhans, M., Cline, M.S., Clawson, H., Barber, G.P., Haussler, D. and Kent, W.J. (2013) 'The UCSC Genome Browser database: extensions and updates 2013', *Nucleic Acids Res*, 41(Database issue), pp. D64-9.

Milde-Langosch, K. (2005) 'The Fos family of transcription factors and their role in tumourigenesis', *European Journal of Cancer*, 41(16), pp. 2449-2461.

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. 1692-1694.

Mitra, D., Brumlik, M.J., Okamgba, S.U., Zhu, Y., Duplessis, T.T., Parvani, J.G., Lesko, S.M., Brogi, E. and Jones, F.E. (2009) 'An oncogenic isoform of HER2 associated with locally disseminated breast cancer and trastuzumab resistance', *Molecular Cancer Therapeutics*, 8(8), pp. 2152-2162.

Moore, M.J. and Sharp, P.A. (1993) 'Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing', *Nature*, 365(6444), pp. 364-368.

Nam, E.A. and Cortez, D. (2011) 'ATR signalling: more than meeting at the fork', *Biochemical Journal*, 436(3), pp. 527-536.

Nayler, O., Cap, C. and Stamm, S. (1998a) 'Human Transformer-2-beta Gene (SFRS10): Complete Nucleotide Sequence, Chromosomal Localization, and Generation of a Tissue-Specific Isoform', *Genomics*, 53(2), pp. 191-202.

Nayler, O., Strätling, W., Bourquin, J.-P., Stagljar, I., Lindemann, L., Jasper, H., Hartmann, A.M., Fackelmayer, F.O., Ullrich, A. and Stamm, S. (1998b) 'SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements', *Nucleic Acids Research*, 26(15), pp. 3542-3549.

Neugebauer, K.M. (2002) 'On the importance of being co-transcriptional', *J Cell Sci*, 115(20), pp. 3865-3871.

Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E. and Ares, M., Jr. (2007) 'Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay', *Genes Dev*, 21(6), pp. 708-18.

Nilsen, T.W. and Graveley, B.R. (2010) 'Expansion of the eukaryotic proteome by alternative splicing', *Nature*, 463(7280), pp. 457-463.

Novoyatleva, T., Heinrich, B., Tang, Y., Benderska, N., Butchbach, M.E.R., Lorson, C.L., Lorson, M.A., Ben-Dov, C., Fehlbaum, P., Bracco, L., Burghes, A.H.M., Bollen, M. and Stamm, S. (2008) 'Protein phosphatase 1 binds to the RNA recognition motif of several splicing factors and regulates alternative pre-mRNA processing', *Human Molecular Genetics*, 17(1), pp. 52-70.

Ohkura, N., Takahashi, M., Yaguchi, H., Nagamura, Y. and Tsukada, T. (2005) 'Coactivator-associated Arginine Methyltransferase 1, CARM1, Affects Pre-mRNA Splicing in an Isoform-specific Manner', *Journal of Biological Chemistry*, 280(32), pp. 28927-28935.

Oltean, S., Sorg, B.S., Albrecht, T., Bonano, V.I., Brazas, R.M., Dewhirst, M.W. and Garcia-Blanco, M.A. (2006) 'Alternative inclusion of fibroblast growth factor receptor 2 exon IIIc in Dunning prostate tumors reveals unexpected epithelial mesenchymal plasticity', *Proceedings of the National Academy of Sciences*, 103(38), pp. 14116-14121.

Owen, N., Zhou, H., Malygin, A.A., Sangha, J., Smith, L.D., Muntoni, F. and Eperon, I.C. (2011) 'Design principles for bifunctional targeted oligonucleotide enhancers of splicing', *Nucleic Acids Res*, 39(16), pp. 7194-208.

Pabla, N., Bhatt, K. and Dong, Z. (2012) 'Checkpoint kinase 1 (Chk1)-short is a splice variant and endogenous inhibitor of Chk1 that regulates cell cycle and DNA damage checkpoints', *Proc Natl Acad Sci U S A*, 109(1), pp. 197-202.

Pabla, N. and Dong, Z. (2012) 'Sibling rivalry in checkpoint control of cell cycle and DNA damage response', *Cell Cycle*, 11(10), pp. 1866-7.

Pajares, M.J., Ezponda, T., Catena, R., Calvo, A., Pio, R. and Montuenga, L.M. (2007) 'Alternative splicing: an emerging topic in molecular and clinical oncology', *The Lancet Oncology*, 8(4), pp. 349-357.

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

Patel, A.A. and Steitz, J.A. (2003) 'Splicing double: insights from the second spliceosome', *Nat Rev Mol Cell Biol*, 4(12), pp. 960-970.

Paull, T.T., Rogakou, E.P., Yamazaki, V., Kirchgessner, C.U., Gellert, M. and Bonner, W.M. (2000) 'A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage', *Current Biology*, 10(15), pp. 886-895.

Pedro A. F. Galante, D.S., Raquel de Sousa Abreu, Michael Gradassi, Natanja Slager, Christine Vogel, Sandro Jose de Souza and Luiz O.F. Penalva (2009) 'A comprehensive in silico expression analysis of RNA binding proteins in normal and tumor tissue; identification of potential players in tumor formation', *RNA Biology*, 6(4), pp. 426 - 433.

Peebles, C.L. and Finkbeiner, S. (2007) 'RNA decay back in play', *Nat Neurosci*, 10(9), pp. 1083-1084.

Perez, D.S., Hoage, T.R., Pritchett, J.R., Ducharme-Smith, A.L., Halling, M.L., Ganapathiraju, S.C., Streng, P.S. and Smith, D.I. (2008) 'Long, abundantly expressed non-coding transcripts are altered in cancer', *Human Molecular Genetics*, 17(5), pp. 642-655.

Perou, C.M., Jeffrey, S.S., van de Rijn, M., Rees, C.A., Eisen, M.B., Ross, D.T., Pergamenschikov, A., Williams, C.F., Zhu, S.X., Lee, J.C.F., Lashkari, D., Shalon, D., Brown, P.O. and Botstein, D. (1999) 'Distinctive gene expression patterns in human mammary epithelial cells and breast cancers', *Proceedings of the National Academy of Sciences*, 96(16), pp. 9212-9217.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lonning, P.E., Borresen-Dale, A.-L., Brown, P.O. and Botstein, D. (2000) 'Molecular portraits of human breast tumours', *Nature*, 406(6797), pp. 747-752.

Pertea, M. and Salzberg, S. (2010) 'Between a chicken and a grape: estimating the number of human genes', *Genome Biology*, 11(5), p. 206.

Peter Stoilov, E.M., Marieta Gencheva, David Glick, Hermona Soreq, Stefan Stamm. (2002) 'Defects in Pre-mRNA Processing as Causes of and Predisposition to Diseases', *DNA and Cell Biology*, 21(11), pp. 803-818.

Pilch, D.R., Sedelnikova, O.A., Redon, C., Celeste, A., Nussenzweig, A. and Bonner, W.M. (2003) 'Characteristics of γ-H2AX foci at DNA double-strand breaks sites', *Biochemistry and Cell Biology*, 81(3), pp. 123-129.

Prasad, J. and Manley, J.L. (2003) 'Regulation and Substrate Specificity of the SR Protein Kinase Clk/Sty', *Molecular and Cellular Biology*, 23(12), pp. 4139-4149.

Proudfoot, N.J., Furger, A. and Dye, M.J. (2002) 'Integrating mRNA Processing with Transcription', *Cell*, 108(4), pp. 501-512.

Qin, Z.-Y., Zhang, M., Guo, X.-R., Wang, Y.-M., Zhu, G.-Z., Ni, Y.-H., Zhao, Y.-P., Qiu, J., Kou, C.-Z., Qin, R. and Cao, X.-G. (2012) 'α-Lipoic acid ameliorates impaired glucose uptake in LYRM1 overexpressing 3T3-L1 adipocytes through the IRS-1/Akt signaling pathway', *Journal of Bioenergetics and Biomembranes*, 44(5), pp. 579-586.

Quinlan, A.R. and Hall, I.M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841-842.

R Foundation for Statistical Computing, V., Austria. 'R: A language and environment for statistical computing. '.

Raponi, M., Smith, L.D., Silipo, M., Stuani, C., Buratti, E. and Baralle, D. (2014) 'BRCA1 exon 11 a model of long exon splicing regulation', *RNA Biology*, 11(4), pp. 351-359.

Rappsilber, J., Ryder, U., Lamond, A.I. and Mann, M. (2002) 'Large-Scale Proteomic Analysis of the Human Spliceosome', *Genome Research*, 12(8), pp. 1231-1245.

Ren, S., Liu, Y., Xu, W., Sun, Y., Lu, J., Wang, F., Wei, M., Shen, J., Hou, J., Gao, X., Xu, C., Huang, J., Zhao, Y. and Sun, Y. (2013) 'Long Noncoding RNA MALAT-1 is a New Potential Therapeutic Target for Castration Resistant Prostate Cancer', *The Journal of Urology*, 190(6), pp. 2278-2287.

Richardson, R.T., Alekseev, O.M., Grossman, G., Widgren, E.E., Thresher, R., Wagner, E.J., Sullivan, K.D., Marzluff, W.F. and O'Rand, M.G. (2006) 'Nuclear Autoantigenic Sperm Protein (NASP), a Linker Histone Chaperone That Is Required for Cell Proliferation', *Journal of Biological Chemistry*, 281(30), pp. 21526-21534.

Robberson, B.L., Cote, G.J. and Berget, S.M. (1990) 'Exon definition may facilitate splice site selection in RNAs with multiple exons', *Mol. Cell. Biol.*, 10(1), pp. 84-94.

Roberts, J.M., Ennajdaoui, H., Edmondson, C., Wirth, B., Sanford, J. and Chen, B. (2013) 'Splicing factor TRA2B is required for neural progenitor survival', *J Comp Neurol*.

Roberts, J.M., Ennajdaoui, H., Edmondson, C., Wirth, B., Sanford, J.R. and Chen, B. (2014) 'Splicing factor TRA2B is required for neural progenitor survival', *Journal of Comparative Neurology*, 522(2), pp. 372-392.

Roth, M.B., Murphy, C. and Gall, J.G. (1990) 'A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle', *The Journal of Cell Biology*, 111(6), pp. 2217-2223.

Rubin, I. and Yarden, Y. (2001) 'The Basic Biology of HER2', *Annals of Oncology*, 12(suppl 1), pp. S3-S8.

Sanford, J.R., Coutinho, P., Hackett, J.A., Wang, X., Ranahan, W. and Caceres, J.F. (2008) 'Identification of Nuclear and Cytoplasmic mRNA Targets for the Shuttling Protein SF2/ASF', *PLoS ONE*, 3(10), p. e3369.

Schellenberg, M.J., Ritchie, D.B. and MacMillan, A.M. (2008) 'Pre-mRNA splicing: a complex picture in higher definition', *Trends in Biochemical Sciences*, 33(6), pp. 243-246.

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 Research*, 41(22), pp. 10170-10184.

Schmidt, M.J., West, S. and Norbury, C.J. (2011) 'The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation', *RNA*, 17(1), pp. 39-44.

Schoppy, D.W., Ragland, R.L., Gilad, O., Shastri, N., Peters, A.A., Murga, M., Fernandez-Capetillo, O., Diehl, J.A. and Brown, E.J. (2012) 'Oncogenic stress sensitizes murine cancers to hypomorphic suppression of ATR', *The Journal of Clinical Investigation*, 122(1), pp. 241-252.

Schwartz, S., Hall, E. and Ast, G. (2009) 'SROOGLE: webserver for integrative, user-friendly visualization of splicing signals', *Nucleic Acids Research*, 37(suppl 2), pp. W189-W192.

Scott, G.K., Robles, R., Park, J.W., Montgomery, P.A., Daniel, J., Holmes, W.E., Lee, J., Keller, G.A., Li, W.L. and Fendly, B.M. (1993) 'A truncated intracellular HER2/neu receptor produced by alternative RNA processing affects growth of human carcinoma cells', *Molecular and Cellular Biology*, 13(4), pp. 2247-2257.

Segade, F., Claudio, E., Wrobel, K., Ramos, S. and Lazo, P.S. (1995) 'Isolation of nine gene sequences induced by silica in murine macrophages', *The Journal of Immunology*, 154(5), pp. 2384-2392.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A. and Sattler, M. (2003) 'Structural Basis for the Molecular Recognition between Human Splicing Factors U2AF65 and SF1/mBBP', *Molecular Cell*, 11(4), pp. 965-976.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) 'Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks', *Genome Research*, 13(11), pp. 2498-2504.

Sharp, P.A. (1994) 'Split genes and RNA splicing', *Cell*, 77(6), pp. 805-815.

Shepard, P. and Hertel, K. (2009a) 'The SR protein family', *Genome Biol*, 10, p. 242.

Shiloh, Y., Shema, E., Moyal, L. and Oren, M. (2011) 'RNF20–RNF40: A ubiquitin-driven link between gene expression and the DNA damage response', *FEBS Letters*, 585(18), pp. 2795-2802.

Shkreta, L., Bell, B., Revil, T., Venables, J., Prinos, P., Elela, S. and Chabot, B. (2013) 'Cancer-Associated Perturbations in Alternative Pre-messenger RNA Splicing', in Wu, J.Y. (ed.) *RNA and Cancer*. Springer Berlin Heidelberg, pp. 41-94.

Siomi, H. and Dreyfuss, G. (1997) 'RNA-binding proteins as regulators of gene expression', *Current Opinion in Genetics & Development*, 7(3), pp. 345-353.

Skotheim, R.I. and Nees, M. (2007) 'Alternative splicing in cancer: Noise, functional, or systematic?', *The International Journal of Biochemistry & Cell Biology*, 39(7–8), pp. 1432-1449.

Smith, C.W.J. and Valcárcel, J. (2000) 'Alternative pre-mRNA splicing: the logic of combinatorial control', *Trends in Biochemical Sciences*, 25(8), pp. 381-388.

Smith, E.R., Cayrou, C., Huang, R., Lane, W.S., Cote, J. and Lucchesi, J.C. (2005) 'A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16', *Mol Cell Biol*, 25(21), pp. 9175-88.

Sorek, R. and Ast, G. (2003) 'Intronic Sequences Flanking Alternatively Spliced Exons Are Conserved Between Human and Mouse', *Genome Research*, 13(7), pp. 1631-1637.

Soulard, M., Valle, V.D., Siomi, M.C., Pin`ol-Roma, S., codogno, P., Bauvy, C., Bellini, M., Lacroix, J.-C., Monod, G., Dreyfuss, G. and Larsen, C.-J. (1993) 'hnRNP G: sequence and characterization of a glycosylated RNA-binding protein', *Nucleic Acids Research*, 21(18), pp. 4210-4217.

Soule HD, V.J., Long A, Albert S, Brennan M. (1973) 'A human cell line from a pleural effusion derived from a breast carcinoma.', *J Natl Cancer Inst.* , 51(5), pp. 1409-16.

Spellman, R., Llorian, M. and Smith, C.W. (2007) 'Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1', *Mol Cell*, 27(3), pp. 420-34.

Srebrow, A. and Kornblihtt, A.R. (2006) 'The connection between splicing and cancer', *J Cell Sci*, 119(13), pp. 2635-2641.

Srinivasan, Seetha V., Dominguez-Sola, D., Wang, Lily C., Hyrien, O. and Gautier, J. (2013) 'Cdc45 Is a Critical Effector of Myc-Dependent DNA Replication Stress', *Cell Reports*, 3(5), pp. 1629-1639.

Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A. and Tyers, M. (2006) 'BioGRID: a general repository for interaction datasets', *Nucleic Acids Research*, 34(suppl 1), pp. D535-D539.

Stella, A., Wagner, A., Shito, K., Lipkin, S.M., Watson, P., Guanti, G., Lynch, H.T., Fodde, R. and Liu, B. (2001) 'A Nonsense Mutation in MLH1 Causes Exon Skipping in Three Unrelated HNPCC Families', *Cancer Research*, 61(19), pp. 7020-7024.

Stewart, M. (2010) 'Nuclear export of mRNA', *Trends in Biochemical Sciences*, 35(11), pp. 609-617.

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.

Storbeck, M., Hupperich, K., Gaspar, J.A., Meganathan, K., Martínez Carrera, L., Wirth, R., Sachinidis, A. and Wirth, B. (2014) 'Neuronal-Specific Deficiency of the Splicing Factor <italic>Tra2b</italic> Causes Apoptosis in Neurogenic Areas of the Developing Mouse Brain', *PLoS One*, 9(2), p. e89020.

Stratton, M.R., Campbell, P.J. and Futreal, P.A. (2009) 'The cancer genome', *Nature*, 458(7239), pp. 719-724.

Syljuåsen, R.G., Sørensen, C.S., Hansen, L.T., Fugger, K., Lundin, C., Johansson, F., Helleday, T., Sehested, M., Lukas, J. and Bartek, J. (2005) 'Inhibition of Human Chk1 Causes Increased Initiation of DNA Replication, Phosphorylation of ATR Targets, and DNA Breakage', *Molecular and Cellular Biology*, 25(9), pp. 3553-3562.

Tacke, R., Tohyama, M., Ogawa, S. and Manley, J.L. (1998a) 'Human Tra2 Proteins Are Sequence-Specific Activators of Pre-mRNA Splicing', *Cell*, 93(1), pp. 139-148.

Takeo, K., Kawai, T., Nishida, K., Masuda, K., Teshima-Kondo, S., Tanahashi, T. and Rokutan, K. (2009) 'Oxidative stress-induced alternative splicing of transformer 2β (SFRS10) and CD44 pre-mRNAs in gastric epithelial cells', *American Journal of Physiology - Cell Physiology*, 297(2), pp. C330-C338.

Tian, M. and Maniatis, T. (1993) 'A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA', *Cell*, 74(1), pp. 105-114.

Tran, Q., Coleman, T.P. and Roesser, J.R. (2003) 'Human transformer 2[beta] and SRp55 interact with a calcitonin-specific splice enhancer', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1625(2), pp. 141-152.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., Blencowe, B.J., Prasanth, S.G. and Prasanth, K.V. (2010a) 'The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation', *Mol Cell*, 29, pp. 925 - 938.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., Blencowe, B.J., Prasanth, S.G. and Prasanth, K.V. (2010b) 'The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation', *Molecular Cell*, 39(6), pp. 925-938.

Turunen, J.J., Niemelä, E.H., Verma, B. and Frilander, M.J. (2013) 'The significant other: splicing by the minor spliceosome', *Wiley Interdisciplinary Reviews: RNA*, 4(1), pp. 61-76.

Ule, J., Jensen, K., Mele, A. and Darnell, R.B. (2005a) 'CLIP: A method for identifying protein-RNA interaction sites in living cells', *Methods*, 37(4), pp. 376-386.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A. and Darnell, R.B. (2003) 'CLIP Identifies Nova-Regulated RNA Networks in the Brain', *Science*, 302(5648), pp. 1212-1215.

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J. and Darnell, R.B. (2006) 'An RNA map predicting Nova-dependent splicing regulation', *Nature*, 444(7119), pp. 580-6.

Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.-S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., Zeeberg, B.R., Kane, D., Weinstein, J.N., Blume, J. and Darnell, R.B. (2005c) 'Nova regulates brain-specific splicing to shape the synapse', *Nat Genet*, 37(8), pp. 844-852.

Uren, P.J., Bahrami-Samani, E., Burns, S.C., Qiao, M., Karginov, F.V., Hodges, E., Hannon, G.J., Sanford, J.R., Penalva, L.O.F. and Smith, A.D. (2012) 'Site identification in high-throughput RNA–protein interaction data', *Bioinformatics*, 28(23), pp. 3013-3020.

Urlaub, H., Hartmuth, K. and Lührmann, R. (2002) 'A two-tracked approach to analyze RNA–protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles', *Methods*, 26(2), pp. 170-181.

Valadkhan, S. (2007) 'The spliceosome: caught in a web of shifting interactions', *Current Opinion in Structural Biology*, 17(3), pp. 310-315.

Valadkhan, S. and Jaladat, Y. (2010) 'The spliceosomal proteome: At the heart of the largest cellular ribonucleoprotein machine', *PROTEOMICS*, 10(22), pp. 4128-4141.

Valcarcel, J., Singh, R., Zamore, P.D. and Green, M.R. (1993) 'The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA', *Nature*, 362(6416), pp. 171-175.

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 Tra2β causes inclusion of a germ cell-specific exon', *Human Molecular Genetics*, 14(16), pp. 2289-2303.

Venables, J.P., Brosseau, J.-P., Gadea, G., Klinck, R., Prinos, P., Beaulieu, J.-F., Lapointe, E., Durand, M., Thibault, P., Tremblay, K., Rousset, F., Tazi, J., Abou Elela, S. and Chabot, B. (2013a) 'RBFOX2 Is an Important Regulator of Mesenchymal Tissue-Specific Splicing in both Normal and Cancer Tissues', *Molecular and Cellular Biology*, 33(2), pp. 396-405.

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 Tra2β and affect splicing', *Human Molecular Genetics*, 9(5), pp. 685-694.

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. (2009a) 'Cancer-associated regulation of alternative splicing', *Nat Struct Mol Biol*, 16(6), pp. 670-676.

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. (2009b) 'Cancer-associated regulation of alternative splicing', *Nat Struct Mol Biol*, 16, pp. 670 - 676.

Venables, J.P., Lapasset, L., Gadea, G., Fort, P., Klinck, R., Irimia, M., Vignal, E., Thibault, P., Prinos, P., Chabot, B., Abou Elela, S., Roux, P., Lemaitre, J.M. and Tazi, J. (2013c) 'MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation', *Nat Commun*, 4, p. 2480.

Wahl, M.C., Will, C.L. and Lührmann, R. (2009) 'The Spliceosome: Design Principles of a Dynamic RNP Machine', *Cell*, 136(4), pp. 701-718.

Wan, J., Sazani, P. and Kole, R. (2009) 'Modification of HER2 pre-mRNA alternative splicing and its effects on breast cancer cells', *International Journal of Cancer*, 124(4), pp. 772-777.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nat Rev Genet*, 10(1), pp. 57-63.

Wang, Z., Kayikci, M., Briese, M., Zarnack, K., Luscombe, N.M., Rot, G., Zupan, B., Curk, T. and Ule, J. (2010) 'iCLIP predicts the dual splicing effects of TIA-RNA interactions', *PLoS Biol*, 8(10), p. e1000530.

Warf, M.B. and Berglund, J.A. (2010) 'Role of RNA structure in regulating pre-mRNA splicing', *Trends in Biochemical Sciences*, 35(3), pp. 169-178.

Watermann, D.O., Tang, Y., zur Hausen, A., Jäger, M., Stamm, S. and Stickeler, E. (2006) 'Splicing Factor Tra2-β1 Is Specifically Induced in Breast Cancer and Regulates Alternative Splicing of the CD44 Gene', *Cancer Research*, 66(9), pp. 4774-4780.

Welch, J.E. and O'Rand, M.G. (1990) 'Characterization of a sperm-specific nuclear autoantigenic protein. II. Expression and localization in the testis', *Biology of Reproduction*, 43(4), pp. 569-578.

Westendorf, J.M., Konstantinov, K.N., Wormsley, S., Shu, M.D., Matsumoto-Taniura, N., Pirollet, F., Klier, F.G., Gerace, L. and Baserga, S.J. (1998) 'M phase phosphoprotein 10 is a human U3 small nucleolar ribonucleoprotein component', *Mol Biol Cell*, 9(2), pp. 437-49.

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., Gerstmair, A., Faerber, F. and Kuster, B. (2014) 'Mass-spectrometry-based draft of the human proteome', *Nature*, 509(7502), pp. 582-587.

Wimberly, H., Han, G., Pinnaduwage, D., Murphy, L., Yang, X., Andrulis, I., Sherman, M., Figueroa, J. and Rimm, D. (2014) 'ERβ splice variant expression in four large cohorts of human breast cancer patient tumors', *Breast Cancer Research and Treatment*, 146(3), pp. 657-667.

Witten, J.T. and Ule, J. (2011) 'Understanding splicing regulation through RNA splicing maps', *Trends in Genetics*, 27(3), pp. 89-97.

Wollerton, M.C., Gooding, C., Wagner, E.J., Garcia-Blanco, M.A. and Smith, C.W.J. (2004) 'Autoregulation of Polypyrimidine Tract Binding Protein by Alternative Splicing Leading to Nonsense-Mediated Decay', *Molecular Cell*, 13(1), pp. 91-100.

Wu, J.Y. and Maniatis, T. (1993) 'Specific interactions between proteins implicated in splice site selection and regulated alternative splicing', *Cell*, 75(6), pp. 1061-1070.

Wu, Q. and Krainer, A.R. (1997) 'Splicing of a divergent subclass of AT-AC introns requires the major spliceosomal snRNAs', *RNA*, 3(6), pp. 586-601.

Xiao, Z., Xue, J., Sowin, T.J., Rosenberg, S.H. and Zhang, H. (2004) 'A novel mechanism of checkpoint abrogation conferred by Chk1 downregulation', *Oncogene*, 24(8), pp. 1403-1411.

Yang, Y., Yang, D., Schluesener, H.J. and Zhang, Z. (2007) 'Advances in SELEX and application of aptamers in the central nervous system', *Biomolecular Engineering*, 24(6), pp. 583-592.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.-D. and Gage, F.H. (2009) 'An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells', *Nat Struct Mol Biol*, 16(2), pp. 130-137.

Zarnack, K., König, J., Tajnik, M., Martincorena, I., Eustermann, S., Stévant, I., Reyes, A., Anders, S., Luscombe, Nicholas 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. 453-466.

Zeman, M.K. and Cimprich, K.A. (2014) 'Causes and consequences of replication stress', *Nat Cell Biol*, 16(1), pp. 2-9.

Zhong, X.-Y., Ding, J.-H., Adams, J.A., Ghosh, G. and Fu, X.-D. (2009) 'Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones', *Genes & Development*, 23(4), pp. 482-495.

Zhou, Z. and Fu, X.D. (2013) 'Regulation of splicing by SR proteins and SR protein-specific kinases', *Chromosoma*, 122(3), pp. 191-207.

Zhu, C., Liu, Y.-Q., Chen, F.-K., Hu, D.-L., Yu, Z.-B. and Qian, L.-M. (2010) 'LYRM1, a Gene that Promotes Proliferation and Inhibits Apoptosis during Heart Development', *Molecules*, 15(10), pp. 6974-6982.

Zhu, J., Mayeda, A. and Krainer, A.R. (2001) 'Exon Identity Established through Differential Antagonism between Exonic Splicing Silencer-Bound hnRNP A1 and Enhancer-Bound SR Proteins', *Molecular Cell*, 8(6), pp. 1351-1361.

Appendix A

Complete list of primers used to monitor splicing inclusion of human endogenous exons in this thesis.

Primer	Sequence
ADAM9 F	TCCTGGGATGGTTAACGAAG
ADAM9 R	GGGAGGTGTCACTGGAGAAA
ANKRD1 F	GGAAAGAAGAATGGCAATGG
ANKRD1 R	GCAGCCTTCAGAAACGTAGG
ANKRD17 F	GCTGTATCCCCTGTTGCTGT
ANKRD17 R	TGCCTGCTGATTCATTTGAG
ANLN F	GAAAAGGTGACCGAAAACCA
ANLN R	GTTCTTCGCTGCTTTCTGCT
ATRX F	GAACTTGCAATGAAGGGTGTC
ATRX R	TGAAAAACCTTTGGATGATGAA
ATXN2 F	GGGATCCCAATGATATGTTTC
ATXN2 R	CCAGAATTCGGGTTGAAATC
BDP1 F	TTCCAGAAGTCCAACAAGAGAA
BDP1 R	TTTTTGGCAGGAGGAATGTG
BTBD7 F	AAACATGACCAGGGCAGAAG
BTBD7 internal	GCATCCTGCTTATGTCTCACTG
BTBD7 R	GTCGGTGCACCCATTTAGAG
CALD1 F	GGAAGAGGAGAAGCCAAAGC
CALD1 R	TGTGGGTCATGAATTCTCCA
CANX F	TTGTTGAGGCTCATGATGGA
CANX R	CACCTGGAAGCTTTGACTCC
CARS F	TGGAGTATTTGTCGGTTTCTGA
CARS R	CCCACAGTGGTGAAACTGGT
CCDC112 F	GGGAAGCCAACATTTATGGA
CCDC112 R	TTTGGTTCGTTCTTCCCAAC
CCNL1 F	TTTAGCACCCTCCTCTCTGC
CCNL1 R	TCCGCCAGTTAAGAGGAAAA
CDCA7L F	TCTTGGTGGGGAACTGAAAG
CDCA7L R	GTTGGCTTCCGAGATGATGT
CEP95 F	TCCTGGAAATATTTGATGGTTTG
CEP95 R	AGGTGTGTGCTGTGTCTCCA
CHEK1 F	GACTGGGACTTGGTGCAAAC
CHEK1 R	TGCCATGAGTTGATGGAAGA
CLOCK F	TTCTGAGACTTATGGTTGGTCA
CLOCK R	CAATCGAGCTCATTTTACTACAGC
CSDE1 F	AAGTCTTGCAGGTTGCCATT
CSDE1 F2	CTCCAGTTTCCAGCTGAACG

CSDE1 internal	ACTGTGAGATTGCCCGGTA
CSDE1 R	CGCGAGAGAAGCGAGATTTA
CSDE1 R2	TATCATCGGACCGACGGACT
CSPP1 F1	CTCCCTCCACCATCACAGTT
CSPP1 F2	CGGGAAAGAGAAGAAAGAAGG
CSPP1 R1	TTTTCCAGGAATCTGCTGCT
CSPP1 R2	AGTTCAGTTGGAGGCTCACC
DDX24 F	GGTGGTGCTTCGGTGTTACT
DDX24 internal	AGCTTGTGCCTTTCTCTTGG
DDX24 R	GGCGGTTTCTGAGGTTCTTC
DEK F	ACAGCCAAAAGAGAAAAACCT
DEK F	AGTGCCTGGCCTGTTGTAAA
DEK R	GTTGTTTTTATGAAATCTTTTCTTTCA
DEK R	AGAGGAGAGCGAGGAGGAAG
DNAJA1 F	CTACGATGTTTTGGGGGTCA
DNAJA1 R	TTGCAGAGCCAGTTTTCTTG
DUSP10 F	AGGAACAGGAAGGGCAAGAT
DUSP10 internal	GTTAGCAGGGCAGGTGGTAG
DUSP10 R	TGAATGTGCGAGTCCATAGC
EIF5B F	ATTTTGGCCTTCGATGTGAG
EIF5B R	TCCAAACATTTTGGGTGACTC
F3 F	CCCACTCCTGCCTTTCTACA
F3 R	TAAGCCTCCGGGATGTTTTT
FAM120B F	CTGGTGTTGGCGCATCTCTA
FAM120B internal	TTTTAATACGGACAGACCCTGAA
FAM120B R	AGCCTGCAGCTCTTCTCTTG
FNBP4 F	TCAAGGTCTGTTTGGCAAGA
FNBP4 R	TCAGGATGCAGCAGAACAAC
GALNT11 F	CTGTGGCGGGAGAGAAGAT
GALNT11 R	CAAGCGGTCACTGATAAGCA
GLYR1 F	CTAGACTCCGGGATGGTGAG
GLYR1 R	GCCTGGATCAAAGTGGAACA
HDLBP F	CTGAGTGATTTTGGCCAGGT
HDLBP R	GAAGGAACAGTTGGCTCAGG
HIF1A F1	TCAGAGAAAGCGAAAAATGGA
HIF1A F2	CACCTCTGGACTTGCCTTTC
HIF1A R1	CCCTGCAGTAGGTTTCTGCT
HIF1A R2	AAAACCATCCAAGGCTTTCA
HMGN1 F	GCTTCATCAGAGGCTGGACT
HMGN1 R	CCTGTGCTTTTCAGATTCTTCA
HNRPDL F	GGCTGGTAATTGTTTTGGTGA
HNRPDL R	GTTGCACAACCCAAAGAGGT
IBTK F	TTCAGTTTATTACCAGTGTTCTTGG
IBTK R	CTTTGCAATCCTGCAGTCAG

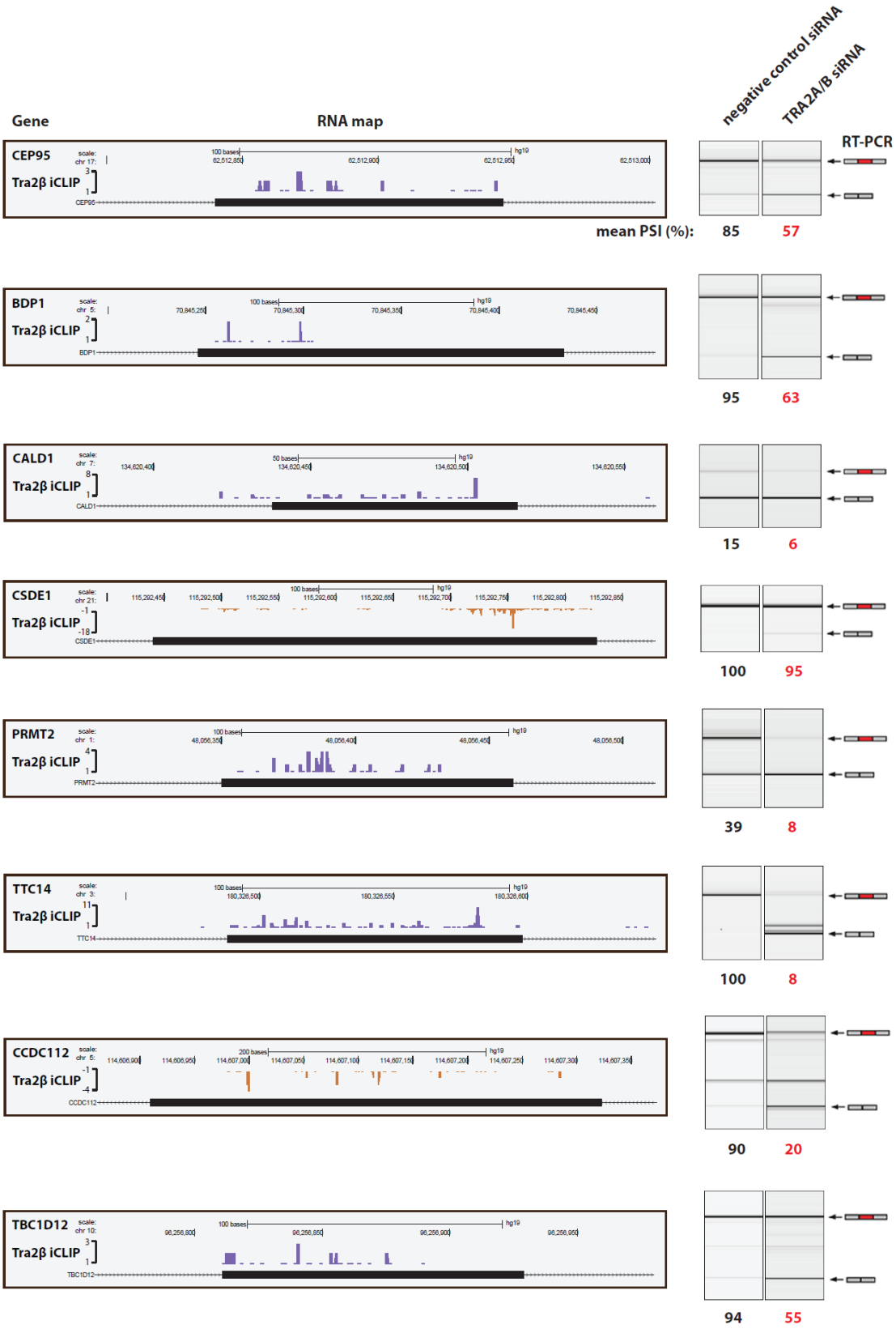
IGF2BP2 F	CGGGAAATCAATCTGTCTGG
IGF2BP2 R	GGCTCAATATGGGACAGTGG
ITFG3 F	CAGATGCACCAGCAGGAGT
ITFG3 R	AGGATCGGCTGAAATTGTTG
ITGA3 F	GACATTGACTCGGAGCTGGT
ITGA3 R	CTGGTCACCCAGTGCTTCTT
IWS1 F	TCAGCAGCTTCATTCATCTTG
IWS1 R	ATTGCAGACATATTTGGAGAATC
JARID2 F	CATCCCAAGTGTCCTCCACT
JARID2 R	CATTGTTGGTGGCTGTTTTG
KDM3A F	GGGGAGGAGGTTTCTCAGTC
KDM3A R	CAAACCAGCTTTGTCCAACA
KDM5A F	TGGTCAGGAGCAGAGGAGTC
KDM5A R	CCCTGCTTCTTTGCACAGTT
KIAA0586 F	CAGAAAGACAGTGGATGAATGG
KIAA0586 R	CTTCGATGGCCCTGATAAAC
KIF14 F	TTGCTTTCTCTGCCCATTTT
KIF14 R	GAAGCTAGCAGCACCACTCC
KIF4A F	CCCTTACTGAAGTGCGTGGT
KIF4A R	TGCTGTTGCTCCATTCTGAC
KTN1a F	CGAGAACATTTGGAAATGGAA
KTN1a R	TGCAAATCACCAGCTACCTTC
KTN1b F	AAGAAATAAGTGGTCTCTGGAATGA
KTN1b R	AAGGGACAGACACCTTTGGA
LYPLAL1 F	GAATGTGGATCAAGCAGGTTT
LYPLAL1 R	CCATGCATCCTCCCATAGAG
LYST F	GAAAGATAAGAGTGGCATTGTGG
LYST internal	TGTGTACAAATCAGATAATCCTGTCA
LYST R	TGCCAGAGAGTATTGGTGGA
MATR3 F	CTGGGGGTTCTCCGTGTC
MATR3 internal	TGACAGAATGGATTATGAAGATGA
MATR3 R	GACGACTGTGACTTGCTCCA
METAP2 F	TTGGAAAGATCAGCATTGGA
METAP2 R	CATCTTGTGTGGGTGGGTATT
MIER1 F	CCATCTGTTGAATCTTCAAGTCC
MIER1 R	TTTTCCCCACTACAGCCACT
MORF4L2 F	AGGAGGCCCCTCTCATTTTA
MORF4L2 R	GATTGCAGAACCTTCCCTGA
MPHOSPH10 F	AGCGGTGTCTGACGGAAGT
MPHOSPH10 R	TTGTTTGAATCCAGCTCATCA
MSL3 F	GCCTGGTGTGGACTCTGTCT
MSL3 R	TCATCCTCCAGCTGCTTCTT
N4BP2L2 F	TGGGAGGTAAAGGTCAATGC
N4BP2L2 internal	AATGGCTGCAGTATGCAAGA

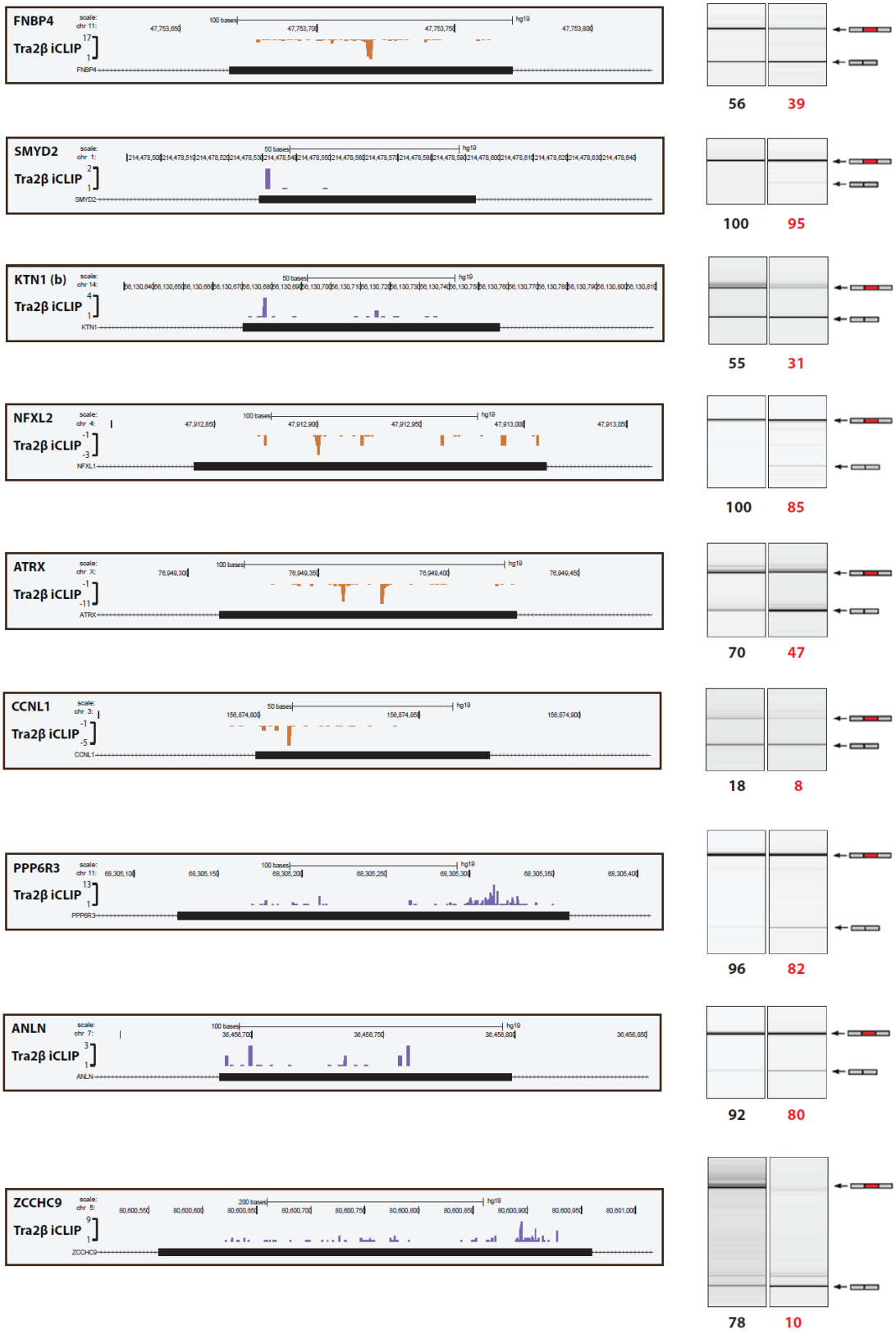
N4BP2L2 R	TGTTCTGGTTCCAGTCATGG
NAP1L1 F1	CTTGCAGGTTTTTGAGAGCA
NAP1L1 F2	GGAATTCCTTTGGGGTCTTC
NAP1L1 R1	CCCTACGTGCTGACGCTAAT
NAP1L1 R2	AATGCTCTCAAAAACCTGCAA
NASP F	AATGGAGAATGGTGTGTTGG
NASP internal	TCCCTAAGGATGGTGCAGTC
NASP R	GAATCATTTGGCATTTCTTCG
NCOA1 F	CCCTGCTAACCCAGACTCAC
NCOA1 R	CCAAAAGAAGAGGTCCCAAG
NEXN F	CTTACACCGGGAAAACTGGA
NEXN R	TTGCAGGCCTAACATCAACA
NFXL1 F	AGTTCCTTCTGGCACCAGTC
NFXL1 R	CATAGCCCCTGCTTGAAAAG
NGLY1 F	AGCGAGCTTCAAACCCTACA
NGLY1 R	GCTTGTATTCCGGTCCAAGA
NIPBL F	ACACAGGAAGAGCCGTTGTT
NIPBL R	ATGTCACTCCGCAGGAAGTC
NUB1 F	TTCAAGCAAAATTGACCCAGT
NUB1 R	GGACCTCAGTTCTCTGCCAGT
NUFIP2 F	TTGCTTCTGCAAATTCCAAAT
NUFIP2 internal	ATTCCTGCTTTCCCAAAGGT
NUFIP2 R	ACCAGCAGCCTCACCAATAC
PAK1 F	TGTGTACACGGTGCCTGAAG
PAK1 R	AGTGCCACCAGTTTCAGAAGA
PAM F	AGCAGCCAAAACGAGAAGAA
PAM R	GAAACCTGGCCTGGTAACAA
PDCD6IP F	CCTGGCTCAAGATGGTGTG
PDCD6IP R	TCTTTCTGTCTTCCGTGCAA
PHF14 F	AAGCCTTTGGCAGCTTCTCT
PHF14 internal	AAAGGGAGATCTGCGTCTCA
PHF14 R	TGCCACAATTGTCACACTGA
PPP1R7 F	AGTCGCAGGAGATGATGGAG
PPP1R7 R	TCCCTATGCGATAGTGATTCAA
PPP6R3 F	GCAAGGAGCCACAAAGAAGA
PPP6R3 R	AGGGGAATCGTTTAGGAGGA
PRMT2 F	CTCAGGCTCCTGGAAAGGAC
PRMT2 R	GGACAGTCACCTGATGTTGCC
PRPF3 F	GTTTGAGGCTGTGGAGGAAG
PRPF3 R	GGGGGCTAATGAAGCTCAG
SETD2 F	TTGTGTTTTCCTTTTCAGTTTGA
SETD2 R	CGCTGAAATAGAGCCCAAAG
SMC4 F	TGGATGTAGCCCAGTCAGAA
SMC4 R	CCAGTGCATGACAACAGGAT

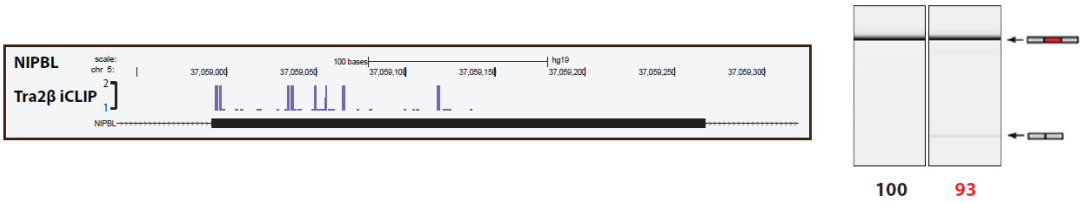
SMN2 F	ACCACCTCCCATATGTCCAG
SMN2 R	TTTGAAGAAATGAGGCCAGTT
SMYD2 F	GTGGGGGACTTGCTGTTCT
SMYD2 R	AGTCTCCGAGGGATTCCAGT
SON F	ACACCCATTGAAGGAAACCA
SON internal	TGACTCAGATTCTTCTGGCTGT
SON R	ATGGCAGCTGCATTAGCTTT
SPECC1 F	GTGGTTCGCCTGAAGAAGAC
SPECC1 internal	TGGCCAAAACTTTGGAAGAG
SPECC1 R	TGCAGCTTTAGGCTGGTTCT
SSR1 F	AGACTCCTCCCCCGCTTG
SSR1 R	CAAACAGTATAGTTGTATCTGCACTCG
TBC1D12 F	AATCTTCCTGCCAAATCTGTG
TBC1D12 R	TTTTCCACCGTTCTTTTGCT
TRA2A F	TTTGGAAACCCTTGATGGAC
TRA2A poison	TTCCCCAATTCAAACAATTCA
TRA2A R	AAAACAACTTCGAGGGCAGA
TRA2B F	ATCCGTGAGCACTTCCACTT
TRA2B poison	CATCTTCCCCACTTCACACA
TRA2B R	GCGTCACATCCGGTAGAGTT
TTC14 F	AGCTTTGGCTTTGGATGAGA
TTC14 R	AGGATGCCCTAGATGAATGC
UACA F	TGGAGATTGAAAATGAAGATTTGA
UACA R	TCCTTAAGCTTTCTTCATGTTGC
UBE3C F	CTGCCAGGATGTTCAGCTTC
UBE3C R	GCCATTAGCAATGGGAAAAG
VAPB F	CCAACCTAAAGCTTGGCAAC
VAPB R	CATCCGCAGTCCATCTTCTT
ZBTB40 F	TCATGGTTACCCAGGATGTTT
ZBTB40 R	ATTCCTCCTTGGCTGTTTCC
ZCCHC11 F	GGAGCCACTGAGATGAGGAG
ZCCHC11 internal	GGAAGGTTGGCAGCATTTAC
ZCCHC11 R	GCCTCTGCTCAGGTGTCAAT
ZCCHC7 F	GCCTCCTTACCTCATTTGTCC
ZCCHC7 R	TTTGTCACTGATGGACCAGTTT
ZCCHC9 F	GCACTATGTGCTTCCGGTCT
ZCCHC9 R	TCTGTGGACCCACACCTGTA
ZZZ3 F	TCTGGTGCCTGCCTAAACTT
ZZZ3 internal	CATGATTCTCGGGTACTCTGC
ZZZ3 R	TCATGGCTGAAGCTGAGAGA

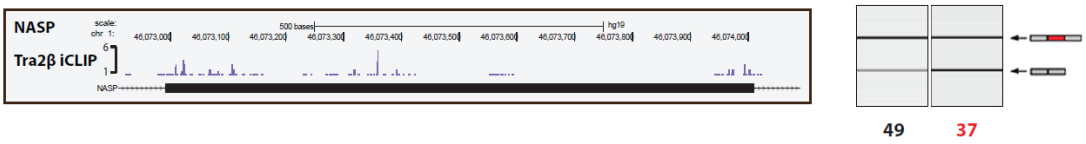
Appendix B

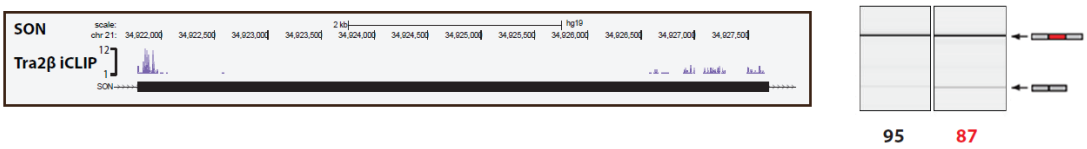
High-resolution iCLIP maps of the Tra2 protein dependent exons identified and validated in this thesis.

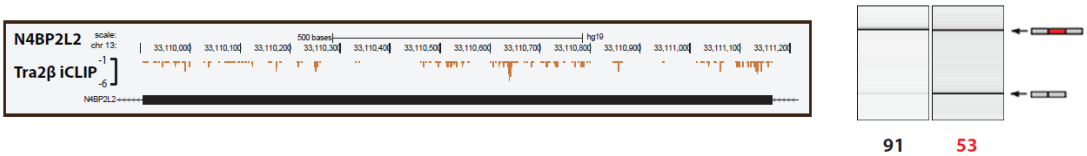



FNBP4
Tra2β iCLIP
56
39
SMYD2
Tra2β iCLIP
100
95
KTN1 (b)
Tra2β iCLIP
55
31
NFXL2
Tra2β iCLIP
100
85
ATRX
Tra2β iCLIP
70
47
CCNL1
Tra2β iCLIP
18
8
PPP6R3
Tra2β iCLIP
96
82
ANLN
Tra2β iCLIP
92
80
ZCCHC9
Tra2β iCLIP
78
10

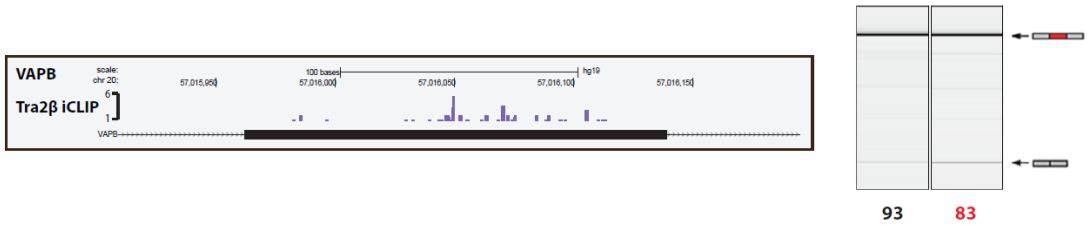

NIPBL
Tra2β iCLIP
100
93

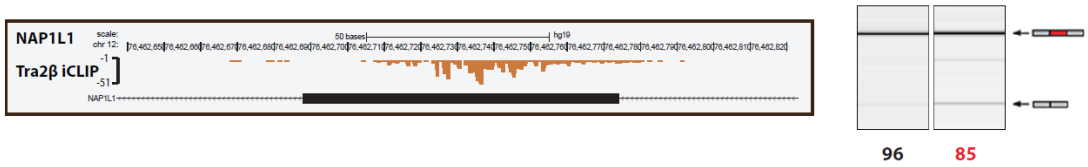

NASP
Tra2β iCLIP
49
37

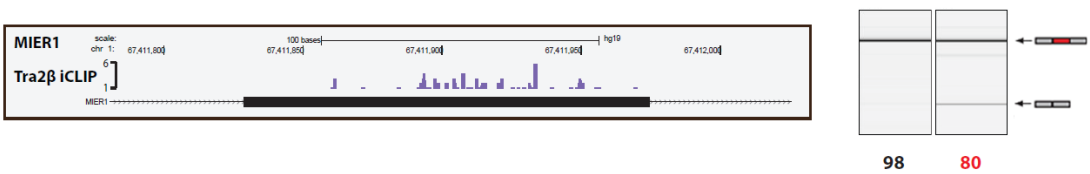

SON
Tra2β iCLIP
95
87

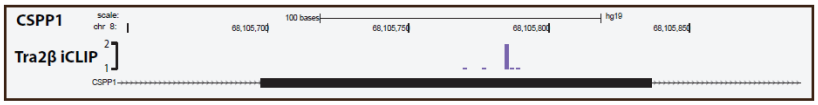

N4BP2L2
Tra2β iCLIP
91
53

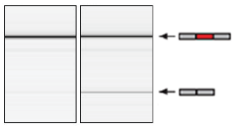

SMC4
Tra2β iCLIP
100
34


VAPB
Tra2β iCLIP
93
83

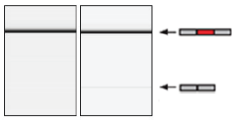

NAP1L1
Tra2β iCLIP
96
85


MIER1
Tra2β iCLIP
98
80

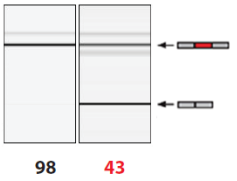

CSPP1
Tra2β iCLIP

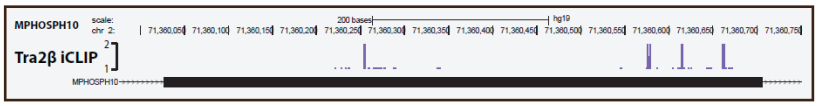

96
82

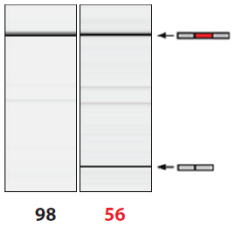

PDCD6IP
Tra2β iCLIP

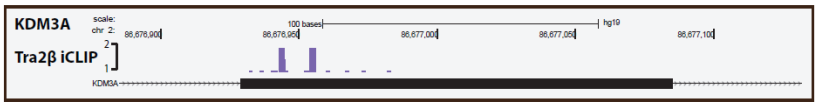

96
82

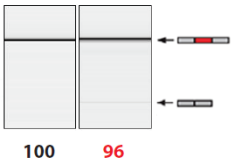

ATXN2
Tra2β iCLIP

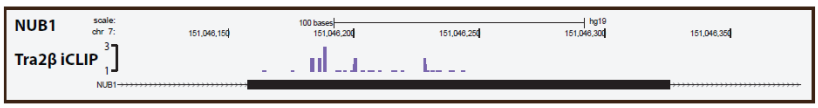

98
43

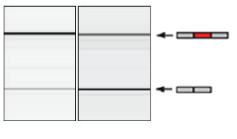

MPHOSPH10
Tra2β iCLIP

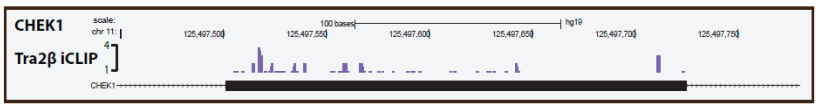

98
56

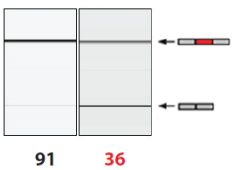

KDM3A
Tra2β iCLIP

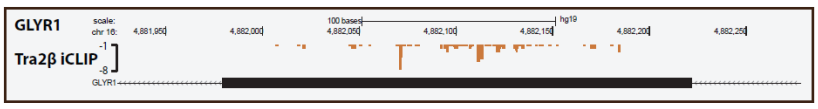

100
96

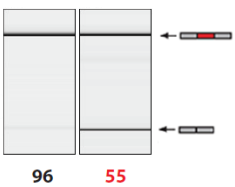

NUB1
Tra2β iCLIP

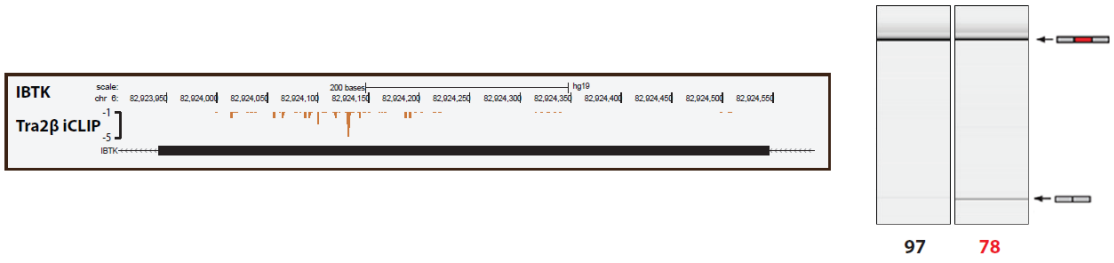

81
39

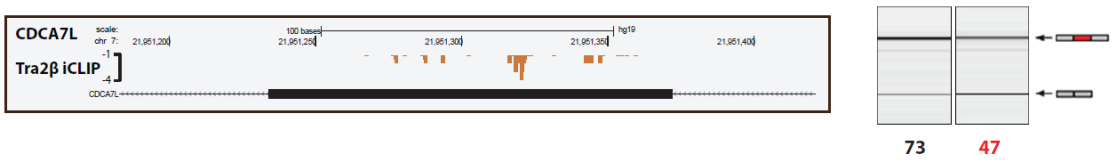

CHEK1
Tra2β iCLIP

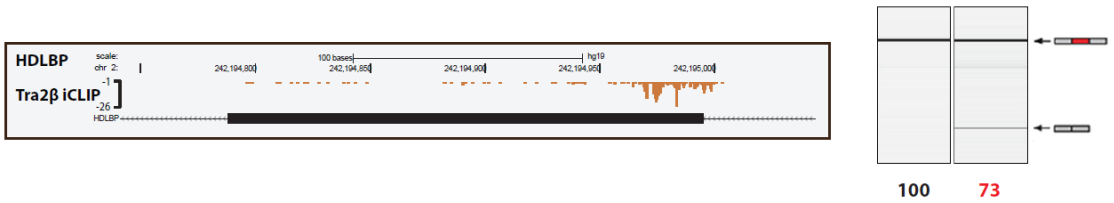

91
36


GLYR1
Tra2β iCLIP


96
55

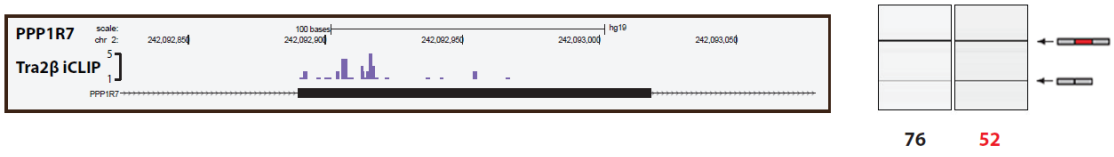

IBTK
Tra2β iCLIP
97
78

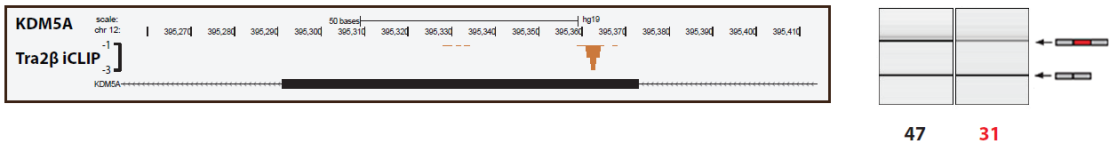

CDCA7L
Tra2β iCLIP
73
47


HDLBP
Tra2β iCLIP
100
73


KIF14
Tra2β iCLIP
99
84

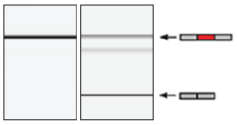

NEXN
Tra2β iCLIP
94
46

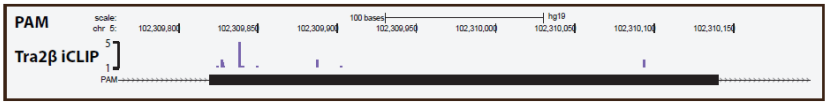

PPP1R7
Tra2β iCLIP
76
52

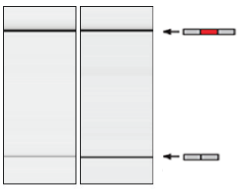

KDM5A
Tra2β iCLIP
47
31


ANKRD17
Tra2β iCLIP
100
96

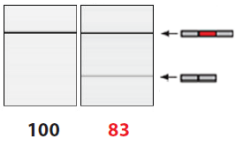

ANKRD1
Tra2β iCLIP

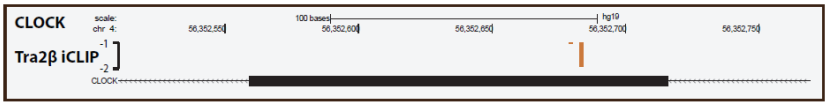

100
27

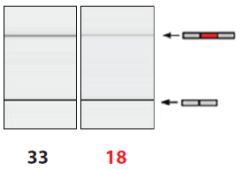

PAM
Tra2β iCLIP


78
63

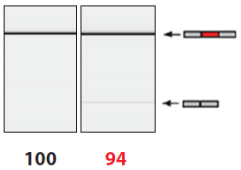

IWS1
Tra2β iCLIP

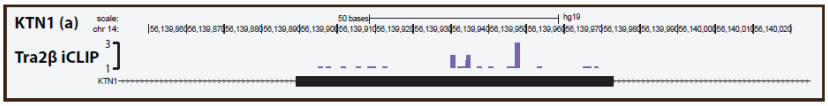

100
83

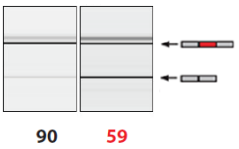

CLOCK
Tra2β iCLIP

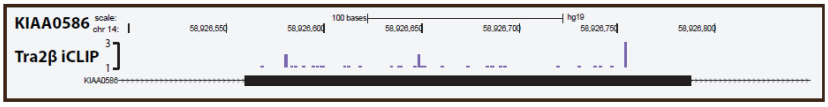

33
18

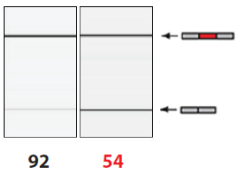

NCOA1
Tra2β iCLIP


100
94

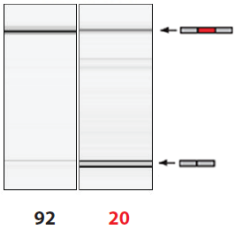

KTN1 (a)
Tra2β iCLIP

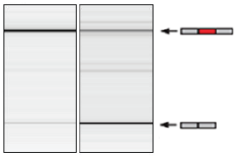

90
59


KIAA0586
Tra2β iCLIP

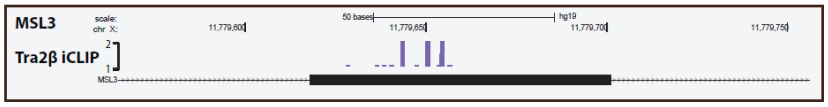

92
54

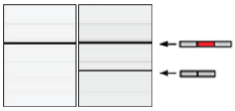

ZCCHC7
Tra2β iCLIP


92
20

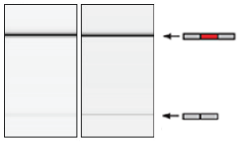


69 22





100 59



96 91

Appendix C

List of publications associated with this thesis.

- Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons
  Best A, James K, Dalgliesh C, Hong E, Kheirolahi-Kouhestani M, Curk T, Xu Y, Danilenko M, Hussain R, Keavney B, Wipat A, Klinck R, Cowell I, Lee K, Austin CA, Venables JP, Chabot B, Santibanez-Koref M, Tyson-Capper A, Elliott DJ.
  Nature Communications. 2014 Sep 11;5:4760. doi: 10.1038/ncomms5760.

- Tra2 protein biology and mechanisms of splicing control.
  Best A, Dalgliesh C, Kheirollahi-Kouhestani M, Danilenko M, Ehrmann I, Tyson-Capper A, Elliott DJ.
  Biochem Soc Trans. 2014 Aug 1;42(4):1152-8. doi: 10.1042/BST20140075.

- Expression of Tra2β in Cancer Cells as a Potential Contributory Factor to Neoplasia and Metastasis.
  Best A, Dagliesh C, Ehrmann I, Kheirollahi-Kouhestani M, Tyson-Capper A, Elliott DJ.
  Int J Cell Biol. 2013;2013:843781. doi: 10.1155/2013/843781.

- How does Tra2β protein regulate tissue-specific RNA splicing?
  Elliott DJ, **Best A**, Dalgliesh C, Ehrmann I, Grellscheid S.
  Biochem Soc Trans. 2012 Aug;40(4):784-8. doi: 10.1042/BST20120036. Review.

- Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2β in development.
  Grellscheid S, Dalgliesh C, Storbeck M, **Best A**, Liu Y, Jakubik M, Mende Y, Ehrmann I, Curk T, Rossbach K, Bourgeois CF, Stévenin J, Grellscheid D, Jackson MS, Wirth B, Elliott DJ.
  PLoS Genetics. 2011 Dec;7(12):e1002390. doi: 10.1371/journal.pgen.1002390.

nature COMMUNICATIONS

ARTICLE

Received 17 Apr 2014 | Accepted 22 Jul 2014 | Published 11 Sep 2014

DOI: 10.1038/ncomms5760 OPEN

Human Tra2 proteins jointly control a *CHEK1* splicing switch among alternative and constitutive target exons

Andrew Best[1], Katherine James[2], Caroline Dalgliesh[1], Elaine Hong[3], Mahsa Kheirolahi-Kouhestani[1], Tomaz Curk[4], Yaobo Xu[1], Marina Danilenko[1], Rafiq Hussain[1], Bernard Keavney[1,5], Anil Wipat[2], Roscoe Klinck[6], Ian G. Cowell[7], Ka Cheong Lee[7], Caroline A. Austin[7], Julian P. Venables[1], Benoit Chabot[6], Mauro Santibanez Koref[1], Alison Tyson-Capper[3] & David J. Elliott[1]

Alternative splicing—the production of multiple messenger RNA isoforms from a single gene—is regulated in part by RNA binding proteins. While the RBPs transformer2 alpha (Tra2α) and Tra2β have both been implicated in the regulation of alternative splicing, their relative contributions to this process are not well understood. Here we find simultaneous—but not individual—depletion of Tra2α and Tra2β induces substantial shifts in splicing of endogenous Tra2β target exons, and that both constitutive and alternative target exons are under dual Tra2α–Tra2β control. Target exons are enriched in genes associated with chromosome biology including *CHEK1*, which encodes a key DNA damage response protein. Dual Tra2 protein depletion reduces expression of full-length CHK1 protein, results in the accumulation of the DNA damage marker γH2AX and decreased cell viability. We conclude Tra2 proteins jointly control constitutive and alternative splicing patterns via paralog compensation to control pathways essential to the maintenance of cell viability.

[1] Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle NE1 3BZ, UK. [2] School of Computing Science, Claremont Tower, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. [3] Institute for Cellular Medicine, Newcastle University, Framlington Place, Newcastle NE2 4HH, UK. [4] Faculty of Computer and Information Science, University of Ljubljana, Trzaska cesta 25, SI-1000, Ljubljana, Slovenia. [5] Institute of Cardiovascular Sciences, The University of Manchester, Manchester M13 9NT, UK. [6] Department of Microbiology and Infectious Diseases, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, Québec, Canada J1E 4K8. [7] Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle NE2 4HH, UK. Correspondence and requests for materials should be addressed to D.J.E. (email: David.Elliott@ncl.ac.uk)

Human genes encode long precursor messenger RNAs (mRNAs) that are extensively processed before nuclear export. This maturation includes the splicing of exons, which normally occurs with high fidelity to create functional mRNAs[1]. Constitutive exons splice into all mRNAs transcribed from a gene, while alternative exons are sometimes included and sometimes skipped[2]. Human protein-coding genes each produce an average of three mRNA isoforms through alternative splicing, many of which are differentially regulated[3]. RNA binding proteins play a key role in transforming precursor RNAs into mRNAs. Although RNA-binding proteins can regulate many transcripts in parallel, some splicing regulatory proteins preferentially engage with transcripts belonging to specific functional classes, including Nova proteins (synapse functions), Fox proteins (neuromuscular, cytoskeleton and EMT functions), PTB proteins (cytoskeleton functions) and T-STAR (synapse functions)[4–8].

Transformer2 (Tra2) proteins are involved in splicing control[9,10]. First discovered in insects, Tra2 proteins form an essential component of the alternative splicing complex that controls fly sexual differentiation[11,12]. Tra2 proteins are conserved across the animal kingdom, but separate gene paralogs encoding Tra2α and Tra2β proteins evolved early in vertebrate evolution[13,14]. Knockout experiments in mice show that Tra2β is essential for embryonic and brain development[15–18]. In humans, Tra2β expression levels change in several cancers (reviewed by Best *et al.*[19]), and Tra2β is implicated in the pathology of other diseases including spinal muscular atrophy[20], Alzheimer's disease[21] and frontotemporal dementia and Parkinsonism linked to chromosome 17 (ref. 22).

Tra2 proteins have amino- and carboxy-terminal domains enriched in arginine and serine residues (RS domains) flanking a single central RNA recognition motif (RRM) and so resemble the relatively well characterized core group of 12 SR proteins that control both constitutive and alternative splicing as well as other aspects of RNA metabolism[23–25]. Each core SR protein contains N-terminal RRMs and single C-terminal RS domains. However, unlike the core SR proteins all current data implicate Tra2 proteins solely in alternative splicing rather than constitutive splicing[10,26], and only SR proteins and not Tra2 proteins can provide splicing activity to S100 extracts[26].

To regulate splicing inclusion Tra2β binds to AGAA-rich and CAA-rich target RNA sequences. These RNA protein interactions have been resolved at the atomic level[9,27]. Endogenous Tra2β target RNAs have been identified using HITS-CLIP[18], RIP-seq[28], shRNA depletion[29] and microarrays[17], but important fundamental questions still remain as to the identity of the biological targets and the functions of vertebrate Tra2 proteins. These include whether endogenous Tra2α and Tra2β proteins jointly control the same splicing targets, and if so what these shared targets are? Although Tra2α and Tra2β both activate splicing of the same model exons when overexpressed in transfected HEK-293 cells (suggesting redundant functions)[18], the *Tra2a* gene alone is not sufficient to maintain viability in *Tra2b* knockout mice (suggesting specific functions)[15]. Another question relates to how Tra2α and Tra2β interact with each other? We previously found that Tra2β protein binds to a poison exon in the *TRA2A* gene to activate poison exon inclusion[18]. Poison exons introduce premature translation termination codons into mRNAs so as to inhibit translation of full-length proteins and are often regulatory[18,30–32], but whether Tra2α might reciprocally control Tra2β expression is not known.

Here we address these questions in human MDA-MB-231 cells that model invasive breast cancer. We find asymmetric splicing feedback control pathways between Tra2α and Tra2β that buffer splicing defects caused by depletion of either Tra2α or Tra2β protein alone. Overriding these feedback control pathways by joint depletion of both Tra2α and Tra2β globally identifies Tra2-dependent target exons, and reveals critical roles for these proteins in DNA damage control and cell viability.

Results

Tra2β efficiently suppresses Tra2α protein expression. To test for *in vivo* interactions between Tra2α and Tra2β proteins, we monitored their expression levels using western blots. Consistent with predictions from our previous study[18], Tra2α protein levels were normally very low but significantly increased after small interfering RNA (siRNA)-mediated depletion of Tra2β (Fig. 1a top panel, compare lanes 1 and 3, and Fig. 1b). Although weak, the Tra2α western blot signal was of the predicted size and was almost completely eliminated following transfection with a *TRA2A*-specific siRNA (Fig. 1a top panel, compare lanes 1 and 2). Tra2α protein depletion had less effect on Tra2β protein levels (Fig. 1a, middle panel and Fig. 1b). Western blot analysis confirmed this effect for two independent sets of siRNAs targeted against different parts of the respective mRNAs (Supplementary Fig. 1).

Consistent with Tra2β protein repressing Tra2α expression via poison exon activation, siRNA-mediated depletion of Tra2β led to strongly reduced splicing inclusion of the *TRA2A* poison exon (Fig. 1d, upper panel). siRNA-mediated depletion of Tra2α protein led to a smaller but detectable effect on splicing inclusion of the *TRA2B* poison exon (Fig. 1d, lower panel). Analysis of *TRA2A* and *TRA2B* steady state mRNA expression levels by quantitative PCR confirmed that each protein also negatively regulates the expression of the other at the RNA level (Fig. 1c).

The *TRA2A* and *TRA2B* genes are differentially expressed. RNA-seq of MDA-MB-231 cells indicated that the *TRA2B* gene is expressed at much higher levels than the *TRA2A* gene (Fig. 1e shows one of three biological replicate RNA-seq analyses, with the height of the *y* axis showing read depth and so indicating relative gene expression levels). This provides a potential mechanism for why Tra2β represses Tra2α protein expression more than vice versa, since lower cellular concentrations of Tra2α would be less able to activate splicing of the *TRA2B* poison exon.

We used iCLIP[33] to systematically map the transcriptome-wide binding sites of human Tra2β in MDA-MB-231 cells. Endogenous Tra2β protein was efficiently immunoprecipitated along with radiolabelled crosslinked RNA. A single radiolabelled RNA protein adduct of ~40 kDa was identified at high RNase concentrations, just above the known molecular weight of uncrosslinked endogenous Tra2β protein (37 kDa) (arrowed in Supplementary Fig. 2a). Lower RNase concentrations enabled endogenous Tra2β binding sites to be mapped across the MDA-MB-231 cell transcriptome in biological triplicate iCLIP experiments. Following deep sequencing, 7,443,903 reads were successfully mapped back to the human genome, of which 3,338,710 were unique cDNA reads used for downstream analysis (Supplementary Data 1). These individual sequencing reads are subsequently referred to as iCLIP tags. The only clusters of Tra2β iCLIP tags, which mapped to the human *TRA2B* and *TRA2A* genes from all three biological replicates were within their respective poison exons (Fig. 1e). Despite much lower levels of overall *TRA2A* gene expression, the *TRA2A* poison exon had a similar number of Tra2β iCLIP tags as the *TRA2B* poison exon. This suggests the *TRA2A* poison exon is a stronger physiological target for Tra2β binding than the *TRA2B* poison exon (the *TRA2A* poison exon also has a much higher density of AGAA Tra2β binding sites than the *TRA2B* poison exon[18]).

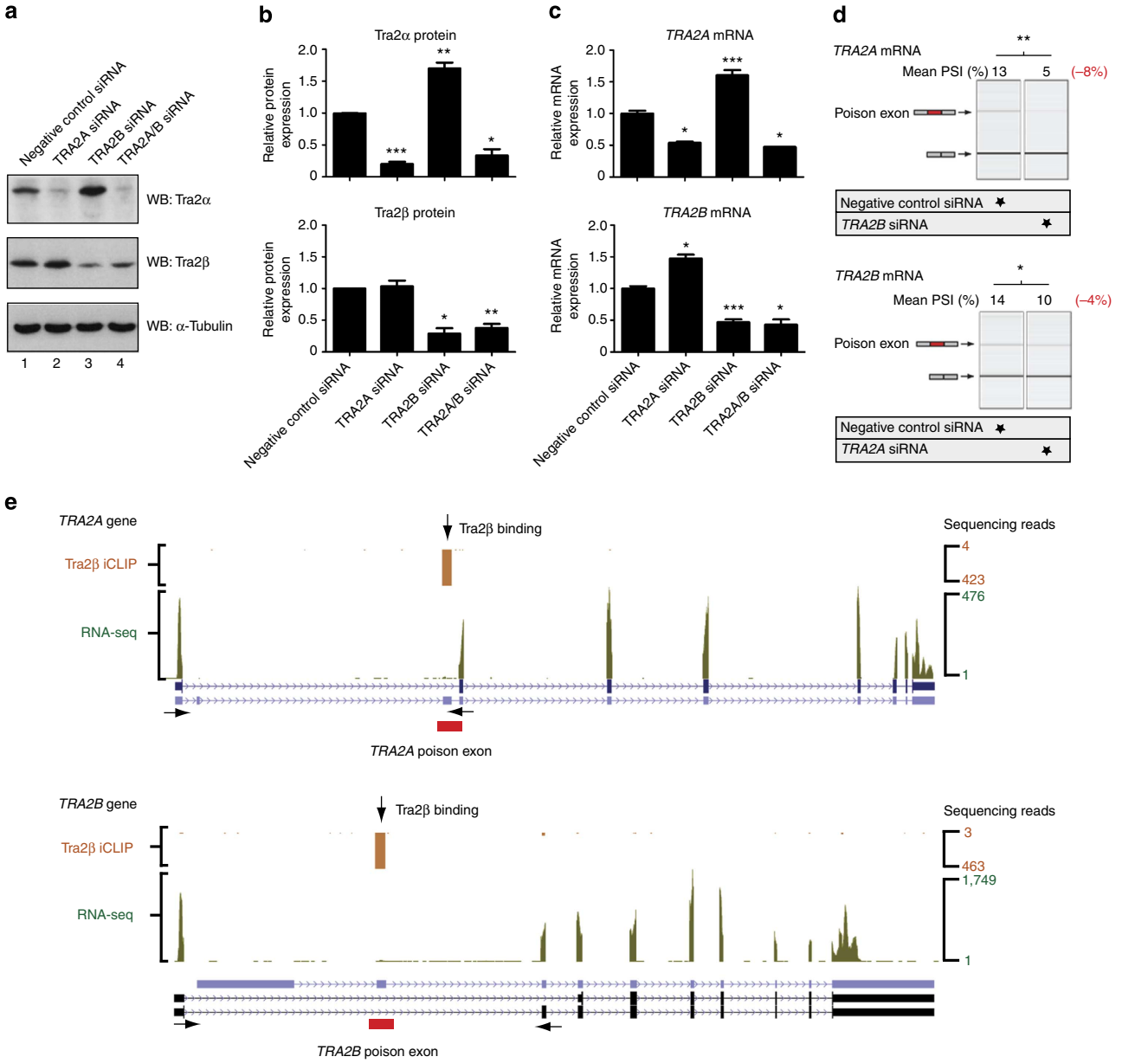


Figure 1 | Tra2β regulates Tra2α protein expression. (**a**) Western blot analysis showing depletion of *TRA2B* induces reciprocal upregulation of Tra2α protein expression, whereas depletion of *TRA2A* had minimal effect on Tra2β protein expression. (**b**) Quantitation of cross-regulation between Tra2α and Tra2β at the protein level (Tra2α and Tra2β protein expression were quantified relative to α-Tubulin from three western blots using independent biological replicates). (**c**) Quantitation of cross-regulation between Tra2α and Tra2β at the RNA level from quantitative PCR analysis of three independent biological replicates in MDA-MB-231 cells. (**d**) Splicing inclusion of the *TRA2A* poison exon is strongly reduced by depletion of endogenous Tra2β protein, whereas splicing inclusion of the *TRA2B* poison exon is less affected by depletion of Tra2α protein. Splicing patterns were monitored by RT–PCR between flanking exons (arrowed) followed by capillary electrophoresis. (**e**) Screenshot from the UCSC genome browser[35] showing the *TRA2B* and *TRA2A* genes, and the positions of aligned RNA-seq reads (green peaks) and Tra2β binding (orange clusters of significant cross-linking by Tra2β protein identified by biological triplicate iCLIP experiments) in MDA-MB-231 cells. Probability (*P*) values were calculated using an independent two-sample *t*-test between negative control siRNA-treated cells and the gene-specific siRNA-treated cells (statistical significance shown as: $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.0001$). All data represented by bar charts was generated from three biological replicates and error bars represent the s.e.m.

Endogenous Tra2α functionally compensates for loss of Tra2β. The most frequently enriched pentamers recovered in the iCLIP tags were highly enriched in AGAA nucleotide sequences (Supplementary Fig. 2b), which is the Tra2β binding site predicted by HITS-CLIP for endogenous mouse Tra2β, RIP-seq and from SELEX experiments using purified Tra2β proteins[18,26,28]. However, our combined human iCLIP data in MDA-MB-231 cells provided substantially more coverage than previously obtained in mouse testis[18] (in which just 177,457 reads were mapped back to the mouse genome). In total, 1,546,290 (44.8%) of unique cDNAs mapped to intronic regions, suggesting the Tra2β iCLIP experiment largely captured Tra2β interactions with pre-mRNAs. However, a further 1,169,374 (33.8%) of unique cDNAs mapped to exons (5′UTR, 3′UTR or ORF), despite exons comprising only ~1% of the genome. After correcting for the relative size of each genomic region (by dividing the

number of unique cDNAs mapping to each genomic region by the relative size of that region within the genome), we find that Tra2β binding is highly enriched within exons: 76.8% of Tra2β iCLIP tags mapped to exons (5′UTR, 3′UTR or ORF), while a further 20.4% mapped to non-coding RNAs (Supplementary Fig. 2c).

We used our iCLIP data to screen for endogenous exons controlled by human Tra2β. Alternative exon junctions (+/−300 bp of either splice site) were stratified to find those with the highest number of iCLIP tags relative to overall iCLIP coverage within the same gene. To test whether upregulation of Tra2α protein expression could functionally compensate in splicing regulation for depletion of Tra2β, we then monitored percentage splicing inclusion (PSI) of associated exons after single depletion of either Tra2α or Tra2β, or after combined depletion of both Tra2α and Tra2β proteins.

Clusters of Tra2β iCLIP tags mapped to alternative exons in the *ATRX*, *GLYR1* and *CEP95* genes (Fig. 2a). Single depletion of either Tra2α or Tra2β had only a small effect on the endogenous splicing pattern of these three exons, but joint depletion of both Tra2α and Tra2β substantially decreased their splicing inclusion (Fig. 2b). We obtained similar results with 14/14 Tra2β target exons identified by iCLIP analysis (Fig. 2c). Individual data for each of these tested exons are shown in Supplementary Fig. 3). In fact, of these 14 tested exons, some less-responsive exons including cassette exons within *PAM* and *BDP1* only responded to depletion of both Tra2 proteins, and not to single depletion of Tra2β at all (Supplementary Fig. 3).

These data are consistent with maintenance of splicing patterns via paralog compensation, that is, following depletion of Tra2β, upregulated Tra2α is able to functionally substitute for Tra2β and largely maintain Tra2 target exon inclusion. The Tra2β target exons inhibited more substantially by joint Tra2 protein depletion compared with single depletion of either Tra2α or Tra2β included *SMN2* exon 7 (Supplementary Fig. 3), which is a candidate target for gene therapy in spinal muscular atrophy[20].

Tra2α and Tra2β control constitutive exon splicing patterns. Splicing profiles of candidate Tra2β target exons (containing Tra2β iCLIP tag coverage) were next analysed using RNA-seq after joint depletion of Tra2α and Tra2β proteins, and changes

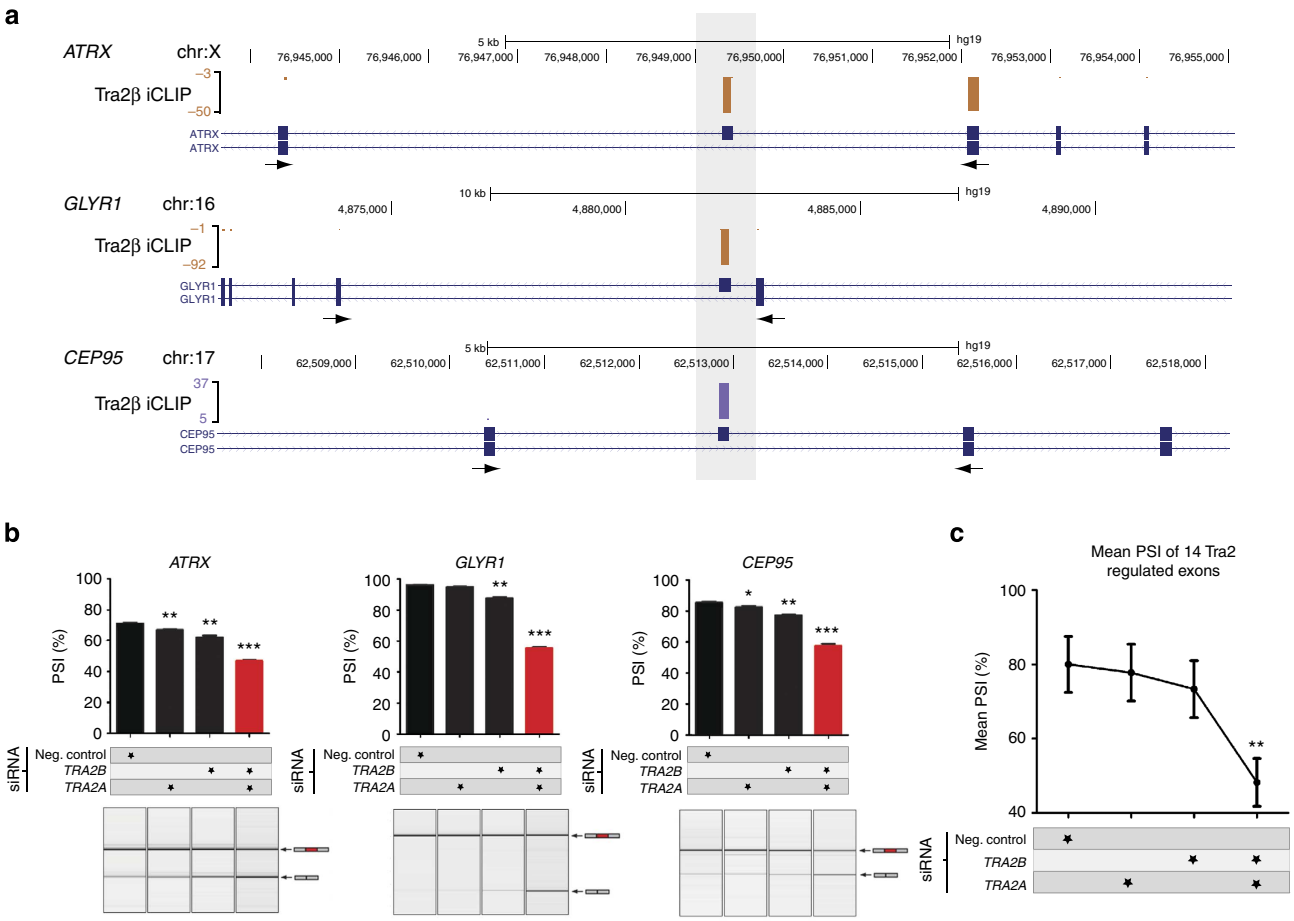


Figure 2 | Endogenous Tra2α functionally compensates for loss of Tra2β. (**a**) UCSC genome browser screenshot[35] showing significant clusters of iCLIP tags mapping directly to alternatively spliced exons within the *CEP95*, *GLYR1* and *ATRX* genes (position of target alternative exons highlighted in grey). (**b**) Splicing inclusion of novel Tra2β target exons within *ATRX*, *GLYR1* and *CEP95* were only slightly affected by depletion of either endogenous Tra2α or Tra2β proteins, but were strongly affected by joint depletion of both Tra2α and Tra2β (red). PSI levels were measured by RT–PCR and capillary gel electrophoresis (lower panels) in three biological replicates (upper panels). (**c**) Splicing inclusion of 14 novel Tra2β target exons showed minimal splicing response to single depletion of either Tra2α or Tra2β, but showed highly significant splicing changes after joint depletion of both Tra2 proteins (complete data for all 14 exons is provided in Supplementary Fig. 4). Probability (*P*) values were calculated using an independent two-sample *t*-test between PSI levels of negative control siRNA-treated cells and *TRA2A*/*TRA2B* siRNA-treated cells (statistical significance: $*P<0.05$, $**P<0.01$, $***P<0.0001$). All data represented by bar charts was generated from three biological replicates where error bars represent the s.e.m.

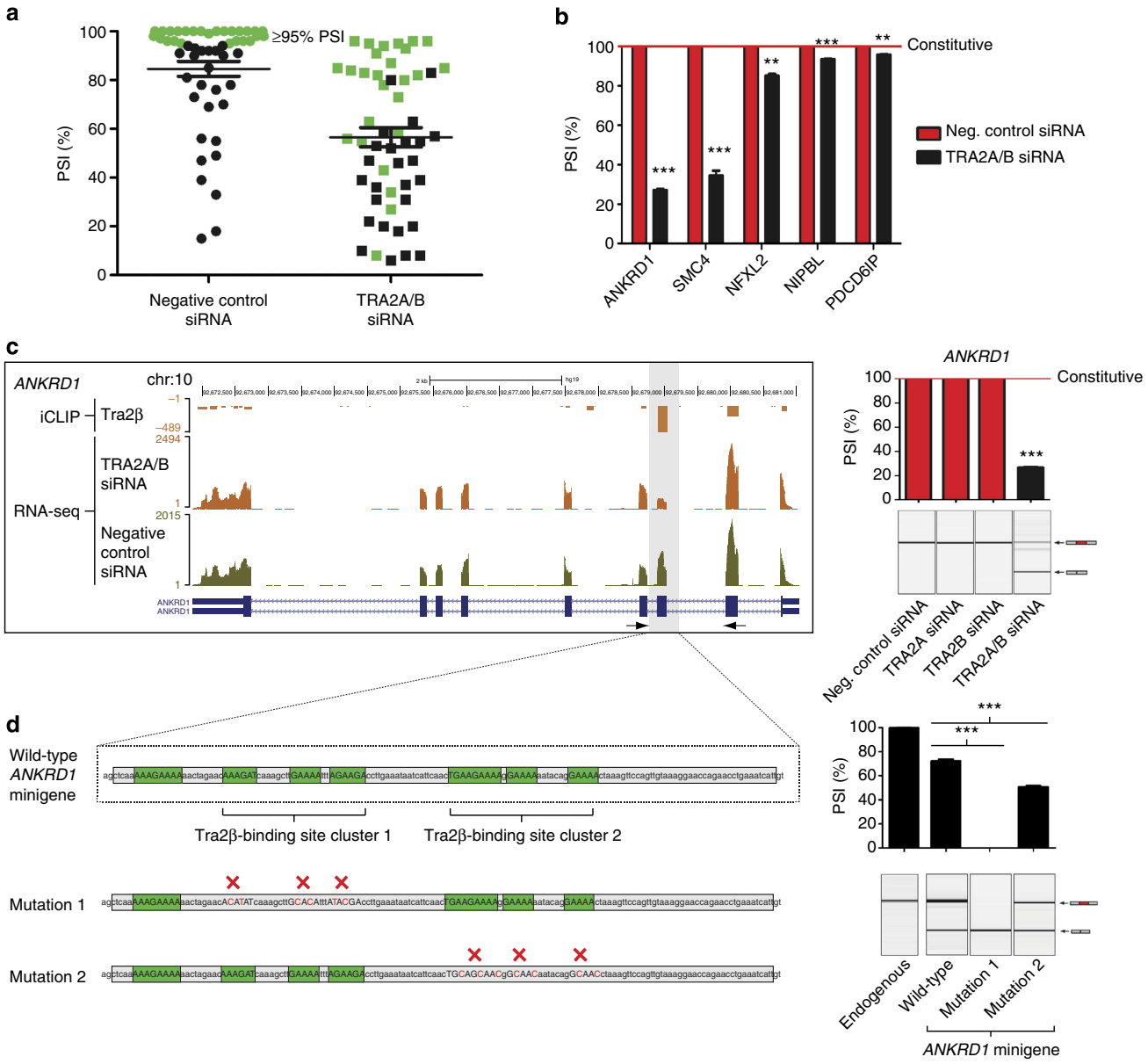


Figure 3 | Tra2 proteins control splicing of constitutively spliced target exons. (**a**) Many novel Tra2α/β-responsive exons normally have high levels of splicing inclusion in MDA-MB-231 cells. PSI levels of target exons are shown from negative control siRNA-treated cells and after joint Tra2 protein depletion. Exons included at equal to or greater than 95% PSI in control MDA-MB-231 cells are highlighted in green. (**b**) The inclusion of five constitutively spliced Tra2β target exons is reduced after joint Tra2α/Tra2β protein depletion. PSI levels were measured by RT–PCR and capillary gel electrophoresis. (**c**) The splicing profile of the *ANKRD1* gene changes in response to joint depletion of Tra2 proteins. Combined iCLIP and RNA-seq data were visualized on the UCSC genome browser[35] (left panel), and splicing inclusion levels directly measured using RT–PCR (right panel). Probability (*P*) values were calculated using an independent two-sample *t*-test between the PSI levels of negative control siRNA-treated cells and the gene-specific siRNA-treated cells (statistical significance: $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.0001$). All data represented by bar charts was generated from three biological replicates and error bars represent s.e.m. (**d**) Tra2β binding sites in the *ANKRD1* exon are essential for its splicing inclusion. The wild-type minigene contained two clusters of Tra2β binding sites, which were independently altered by mutagenesis (left panel). The PSI of the resulting exons was measured after transfection into HEK-293 cells at endogenous Tra2β protein concentrations (right panel). Probability (*P*) values were calculated using an independent two-sample *t*-test between PSI levels of the wild-type and mutated versions of the *ANKRD1* minigene (statistical significance: $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.0001$). All data represented by bar charts was generated from three biological replicates and error bars represent the s.e.m.

validated by reverse transcriptase (reverse transcriptase–PCR). From the initial panel of 30 Tra2 protein-responsive alternative exons that we identified, 7/30 were included at 100% and 17/30 had a PSI greater or equal to 95% in MDA-MB-231 cells (Fig. 3a). These results suggested that Tra2 proteins might be important for the inclusion of constitutive exons (in addition to their expected function in alternative exon splicing regulation), or alternatively that Tra2 protein expression levels in MDA-MB-231 cells are sufficient to promote 100% inclusion of some alternative exons.

To distinguish between these possibilities, we used our iCLIP and RNA-seq data to search for splicing changes in exons that have never previously been annotated as alternatively spliced in human cells[34,35]. Such exons were found in the *ANKRD1*, *SMC4*, *NFXL2*, *NIPBL* and *PDCD6IP* genes. Each of these five exons were spliced at 100% PSI in MDA-MB-231 cells but skipped at

different levels after Tra2 protein depletion (Fig. 3b). The constitutively spliced exon within *ANKRD1* showed the largest change (Fig. 3c), a -73% point switch in PSI after joint depletion of Tra2α and Tra2β. To further dissect splicing control, we cloned the *ANKRD1* exon and its flanking intronic sequences between β–globin exons in a minigene construct. Transfection experiments showed the *ANKRD1* exon is included at 72% PSI when expressed from this minigene, indicating it lacks some important sequences for splicing (the endogenous *ANKRD1* exon was 100% included in these transfected HEK-293 cells, Fig. 3d). Two clusters of GAA-rich Tra2β binding sites were present in the *ANKRD1* exon (Tra2β binding site clusters 1 and 2). Mutation of either Tra2β binding site cluster negatively impacted splicing inclusion, showing an essential role for these binding sites in modulating the inclusion of the *ANKRD1* exon. In particular, mutation of Tra2β binding site cluster 1 (to create Mutant M1) completely abolished splicing inclusion of the *ANKRD1* constitutive exon (Fig. 3d, right panel).

In total, we identified and validated 53 human splicing targets, which were both directly bound and jointly controlled by Tra2 proteins, including both alternative and constitutive exons (Fig. 4a; Supplementary Fig. 10; Supplementary Data 2). As well as *SMN2* exon 7 (ref. 20), these included the *NASP-T* exon and the *TRA2A* poison exon, orthologs of which have both been previously identified as functional Tra2β splicing targets in the mouse testis by HITS-CLIP[18]. The *NASP-T* exon was also in the data set of Tra2β targets identified by RIP-seq[28]. However, the vast majority of the dual Tra2α/Tra2β target exons identified here are novel. The PSI changes for individual genes in response to joint Tra2α and Tra2β protein depletion ranged between −4 and −92% points (measured by RT–PCR, Fig. 4a; Supplementary Data 2), indicating that individual Tra2-dependent exons have different intrinsic requirements for Tra2 proteins. The length of Tra2-dependent exons ranged from 64 nucleotides at the smallest (a cassette exon in the *SMYD2* gene, which showed a −5% point switch in PSI in response to Tra2 protein depletion), to 5916 nucleotides at the largest (an unusually large internal cassette exon in the *SON* gene, which showed a −8 point PSI change in response to endogenous Tra2 protein depletion) (Supplementary Data 2). A further 38 exons had Tra2β iCLIP tag coverage, but did not detectably respond to Tra2 protein depletion in MDA-MB-231 cells: possibly this latter class of exons either might need a relatively small amount of Tra2 protein to be included, or alternatively they might not be controlled by Tra2 proteins in MDA-MB-231 cells (Supplementary Data 2). In a comparison between Tra2 responsive and non-responsive exons, the only statistically significant difference was the density of Tra2β binding sites in the more highly responsive exons (exons showing >15 PSI change following joint Tra2α/Tra2β depletion, Fig. 4b). High resolution iCLIP maps of the individual exons are shown in Supplementary Fig. 10.

Tra2 proteins are highly specific splicing regulators. To establish the relative role of Tra2 proteins in controlling the identified panel of target exons, we probed a custom plate containing cDNAs, where we had systematically knocked down 53 known splicing regulators in MDA-MB-231 cells[5] (Fig. 4c). Strikingly, of all the knockdowns tested, only double knockdown of Tra2 proteins shifted *ANKRD1* splicing (constitutive exon) and joint Tra2 depletion also had the largest effect on splicing for *GLYR1* (alternative exon). An intermediate situation was observed for *SMC4* (constitutive exon) in which knockdown of *SNRP70* (encoding U170K) also reduced splicing inclusion, as did knockdown of *SRPK1*. The splicing inclusion pattern of *CHEK1* was strongly shifted (−78 point PSI switch) by joint depletion of Tra2 proteins, but consistent with broader mechanisms of combinatorial control, significant shifts were also seen after depletion of three core U2 snRNP components, which are thought to be important for splice site commitment for all exons. Depletion of other constitutive splicing factors such as SFRS2, hnRNPK, hnRNPC2, KHSRP and CDC5L also affected *CHEK1* exon 3 splicing.

Since our panel of splicing factor knockdowns was not exhaustive, we cannot exclude all combinations of combinatorial control. However, our data are at least consistent with Tra2 proteins being among the most quantitatively important splicing regulators for their individual target exons.

Tra2 splicing targets associate with chromosome biology. Gene ontology (GO) enrichment analysis of the 53 human genes containing Tra2-dependent exons revealed an enrichment of five functionally similar biological processes (Fig. 5a). Eleven of the 53 genes were annotated to one or more of these processes, with ten of the eleven genes being annotated to the term 'chromosome organization'. There was significant overlap in annotation between this process and annotation to the conceptually related terms 'histone modification' and 'chromatin modification' (Fig. 5b).

The BioGRID database[36] was used to retrieve a network of functional interaction data involving genes containing Tra2-dependent exons. In addition to the eleven genes directly annotated to the five enriched GO processes in Fig. 5a, a further 23 of the 53 genes (43.4%) that contain Tra2-dependent exons also have functional interactions with genes annotated to these terms (Supplementary Fig. 4, and summarized in Fig. 5c). Interestingly, although indirectly connected within the network via these annotated genes, none of the 53 genes containing validated Tra2-dependent exons directly interact with one another in the BioGRID database (Supplementary Fig. 4).

Tra2 proteins control splicing of a key checkpoint protein. Among the Tra2 target exons involved in chromosome biology was exon 3 of the *CHEK1* gene, which encodes the serine/threonine protein kinase CHK1 that is involved in checkpoint control in response to DNA damage. iCLIP analysis identified significant Tra2β binding over *CHEK1* exon 3 (Fig. 6a), and we observed a −55 point PSI switch for this exon in MDA-MB-231 cells after joint Tra2 protein depletion (Fig. 6b). Joint depletions of Tra2α and Tra2β proteins also indicate *CHEK1* exon 3 splicing is under similar control in multiple cell types including MCF7, PC3 and HeLa (Fig. 6b).

Tra2β iCLIP tags mapped throughout *CHEK1* exon 3, but were particularly enriched towards the 3′ splice site (Fig. 7a). We confirmed *CHEK1* exon 3 is a direct target for Tra2β binding *in vitro* by electrophoretic mobility shift assays (EMSAs) using radiolabelled RNA probes corresponding to portions of the exon sequence (Fig. 7b). RNA probe A corresponds to the part of *CHEK1* exon 3 with the most Tra2β iCLIP tags, and also contains the most predicted binding sites for Tra2β (shaded green in Fig. 7B, right hand side). RNA probe A was very efficiently shifted by even the lowest tested concentrations (25 ng) of Tra2β protein. RNA probe B did not bind Tra2β protein as tightly (around 200 ng Tra2β protein was needed to see a comparable shift) and also contained fewer Tra2β binding sites and mapped iCLIP tags. A control RNA probe corresponding to the flanking intron sequence did not shift even at the highest concentrations of Tra2β protein (this intron sequence contained no predicted Tra2β binding sites).

We confirmed that *CHEK1* exon 3 is a direct target for Tra2β splicing regulation using a minigene construct in which *CHEK1* exon 3 is flanked by β-globin exons (Fig. 7c). After transfection of

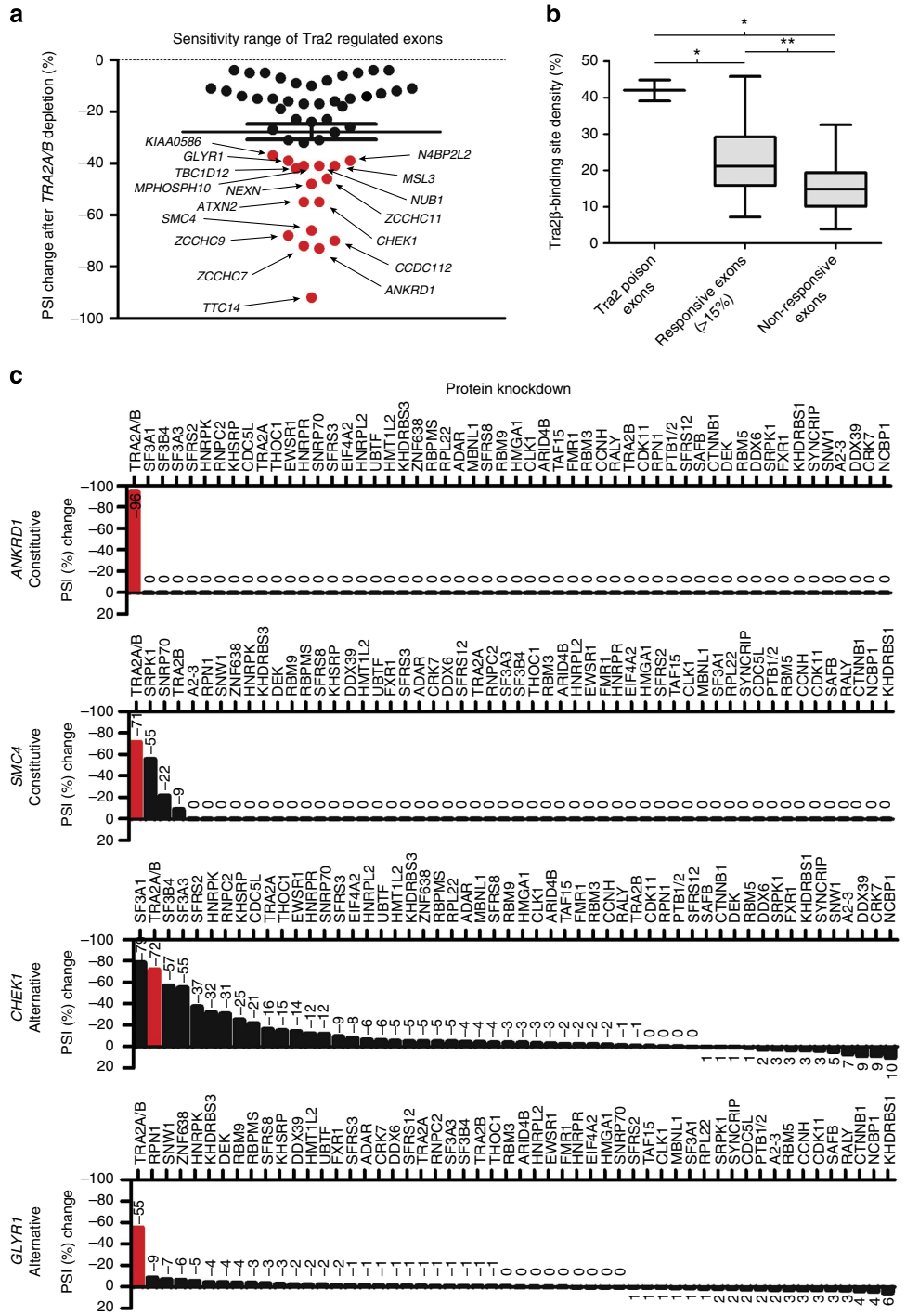


Figure 4 | Identification of exons jointly controlled by Tra2α and Tra2β proteins. (**a**) Scatterplot showing amplitude of splicing response of 53 exons to joint depletion of endogenous Tra2α and Tra2β in MDA-MB-231 cells. The genes corresponding to the highest amplitude PSI changes after joint Tra2 protein depletion are labelled and highlighted in red. (**b**) Analysis of Tra2β binding site density (measured as a percentage of exon content) within groups of Tra2β target exons identified by iCLIP. Tra2β binding site density comparisons are shown between the Tra2α and Tra2β poison exons; all exons that showed a greater than 15% point PSI change following joint Tra2α and Tra2β depletion; and in the exons that bound Tra2β based on iCLIP tag coverage but did not respond to Tra2α and Tra2β depletion. Probability (*P*) values were calculated using an independent two-sample *t*-test (statistical significance: $*P<0.05$, $**P<0.01$, $***P<0.0001$). (**c**) Regulation of *ANKRD1*, *SMC4*, *GLYR1* and *CHEK1* splice variants following knockdown of a panel of RNA binding proteins (RBPs) in MDA-MB-231 cells[5]. The *y* axis shows PSI change after joint Tra2α and Tra2β depletion, with a negative number indicating splicing repression in the absence of these proteins.

this minigene into HEK-293 cells, *CHEK1* exon 3 was skipped, but its splicing inclusion was strongly induced in response to co-transfection with either Tra2β-GFP or Tra2α-GFP. No *CHEK1* exon 3 splicing activation was observed after co-transfection of either Tra2βΔRRM-GFP (lacking the RRM) or GFP alone. Furthermore, point mutations of the Tra2β binding sites within the exon (wild-type binding sites shaded green, mutations shaded red in Fig. 7c right hand side) completely

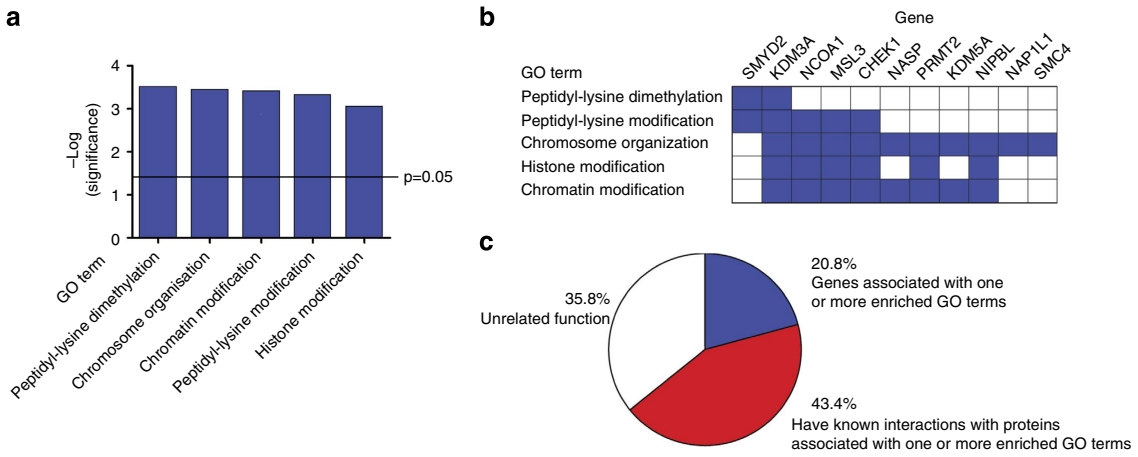


Figure 5 | Tra2 splicing targets are enriched in GO terms associated with chromosome biology. (**a**) GO enrichment analysis reveals splicing targets responsive to endogenous Tra2α/Tra2β protein concentrations are enriched in particular biological processes associated with chromosome biology. (**b**) GO enrichment analysis showed some joint Tra2α/Tra2β-responsive exons were annotated to multiple overlapping biological processes. (**c**) Summary of GO and network analyses of joint Tra2α and Tra2β-dependent splicing targets. Individual segments of the pie chart show the percentage of Tra2α/Tra2β target genes directly annotated to the GO biological processes shown in part **a**; the percentage of Tra2α/Tra2β target genes that interact within the BioGrid database with partners known to be involved in the biological processes shown in part **a**; and the percentage of Tra2α/Tra2β target genes, which have unknown or unrelated functions. Full details of the BioGrid analysis are given in Supplementary Fig. 4.

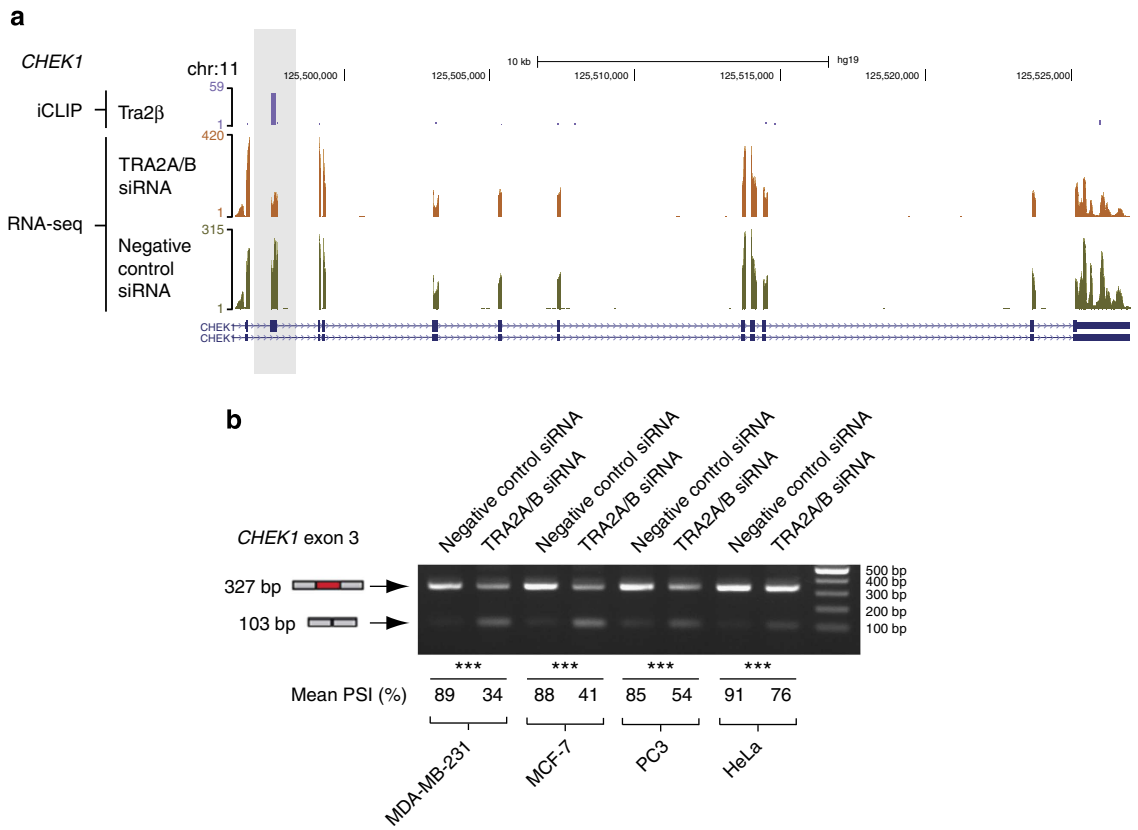


Figure 6 | Splicing inclusion of *CHEK1* exon 3 decreases following joint depletion of Tra2α and Tra2β. (**a**) Screenshot[35] from the UCSC genome browser showing the *CHEK1* gene. Tra2β iCLIP tags are shown at the top in purple, and map predominantly to a known alternative exon (*CHEK1* exon 3) that is skipped in one mRNA isoform (shaded grey). Example RNA-seq tracks from control MDA-MB-231 cells and after joint Tra2α and Tra2β protein depletion are shown in green and orange respectively. (**b**) Splicing inclusion of *CHEK1* exon 3 is inhibited by siRNA-mediated Tra2α and Tra2β protein depletion in multiple cell types. Splicing inclusion was monitored by RT–PCR and agarose gel electrophoresis, and shown as a mean PSI value after analysis of three independent samples. Probability (*P*) values were calculated using an independent two-sample *t*-test between PSI levels of negative control siRNA-treated cells and TRA2A/B siRNA-treated cells (statistical significance: $*P<0.05$, $**P<0.01$, $***P<0.0001$). An uncropped gel is shown in Supplementary Fig. 13.

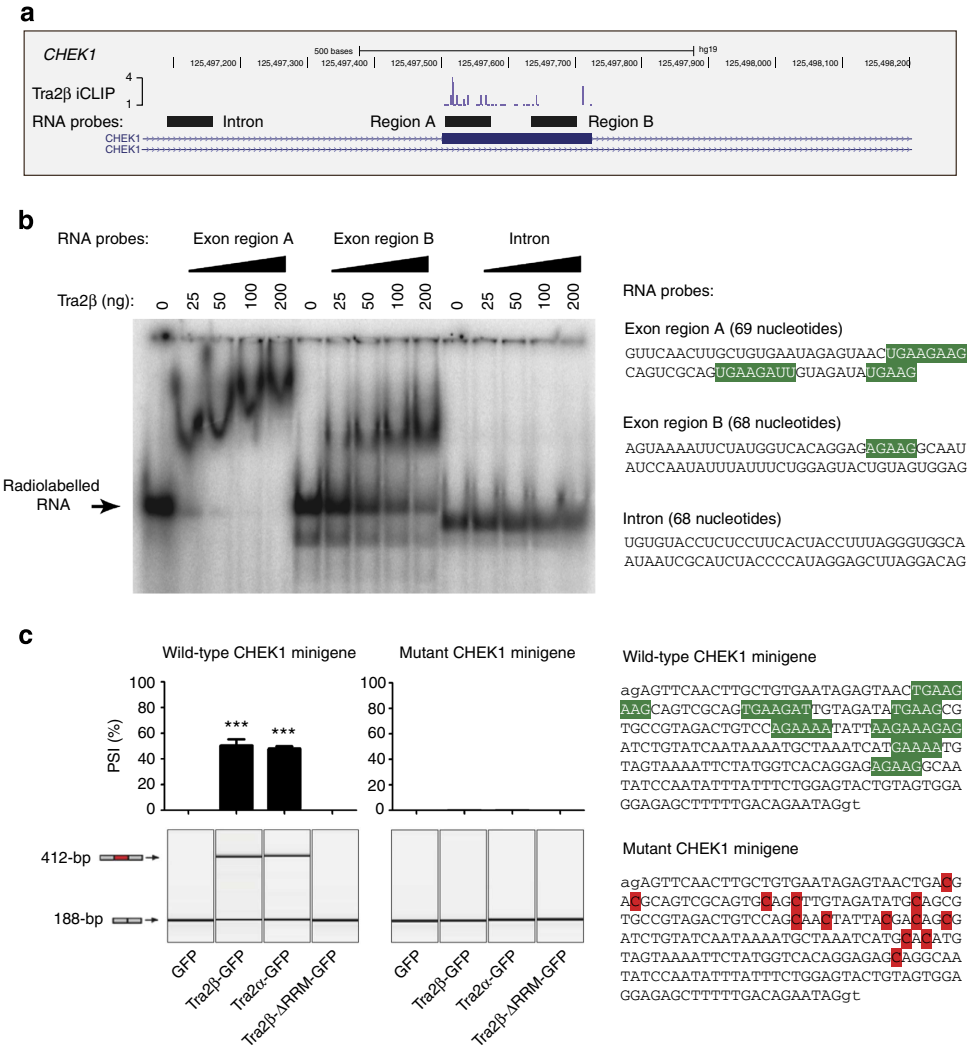


Figure 7 | Human Tra2α and Tra2β regulate *CHEK1* exon 3 through direct RNA protein interactions. (**a**) High resolution map *CHEK1* exon 3 showing mapped iCLIP tags (purple bars) and the three subregions of the pre-mRNA, which were used to generate RNA probes for EMSAs. This screenshot was downloaded from the UCSC genome browser[35]. (**b**) Molecular interactions between purified Tra2β protein and RNA probes in and around *CHEK1* exon 3 (the location of these probes is shown in part **a**). The sequences of the probes are shown to the right, with predicted Tra2β binding sites shaded green. (**c**) Splicing patterns of mRNAs made from a minigene containing *CHEK1* exon 3 in response to coexpressed fusion proteins, expressed either as a PSI (upper bar chart, $n=3$ independent experiments) or shown as one of the original capillary electrophoresis gel-like images from a single experiment (lower image). The sequence of *CHEK1* exon 3 is shown to the right, with the predicted Tra2β binding sites shaded green (above) and the altered sequence after these sites were mutated (below, the altered nucleotides are shaded red). Probability (*P*) values were calculated using an independent two-sample *t*-test between PSI levels of the minigene-derived CHEK1 exon 3 in cells cotransfected with GFP and each of the different Tra2 constructs (statistical significance: $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.0001$). All data represented by bar charts was generated from three biological replicates and error bars represent s.e.m.

abolished splicing activation in response to coexpressed Tra2 proteins.

Tra2 proteins are required for CHK1 protein expression. We carried out further experiments to test if Tra2 proteins are also required for expression of full-length CHK1 protein. On western blots, we could detect expression of a single major protein CHK1 isoform in MDA-MB-231 cells, corresponding to the expected size of full-length CHK1 protein (54 kDa). This band was substantially reduced following treatment with an siRNA directed against *CHEK1* mRNA (Fig. 8a). Consistent with joint control of *CHEK1* expression by Tra2α and Tra2β, levels of full-length CHK1 protein were also substantially reduced after joint depletion of Tra2α and Tra2β. Expression of full-length CHK1 protein was also reduced after joint Tra2α and Tra2β protein depletion in MCF7, PC3 and to a lesser extent HeLa cells (Supplementary Fig. 7a).

A shorter isoform of the CHK1 protein (termed CHK1-S) has previously been reported to be translated from an alternative downstream translational initiation site in exon 3-skipped *CHEK1* mRNA[37]. In our experiments, although depletion of Tra2 proteins switched splicing of *CHEK1* exon 3, they did not lead to an observable increase in any shorter isoform of the CHK1 protein. We detected much lower expression levels of possible shorter CHK1 protein isoforms on western blots compared with full-length CHK1 protein (the ~43 KDa protein that would correspond in size to CHK1-S could only be seen on long

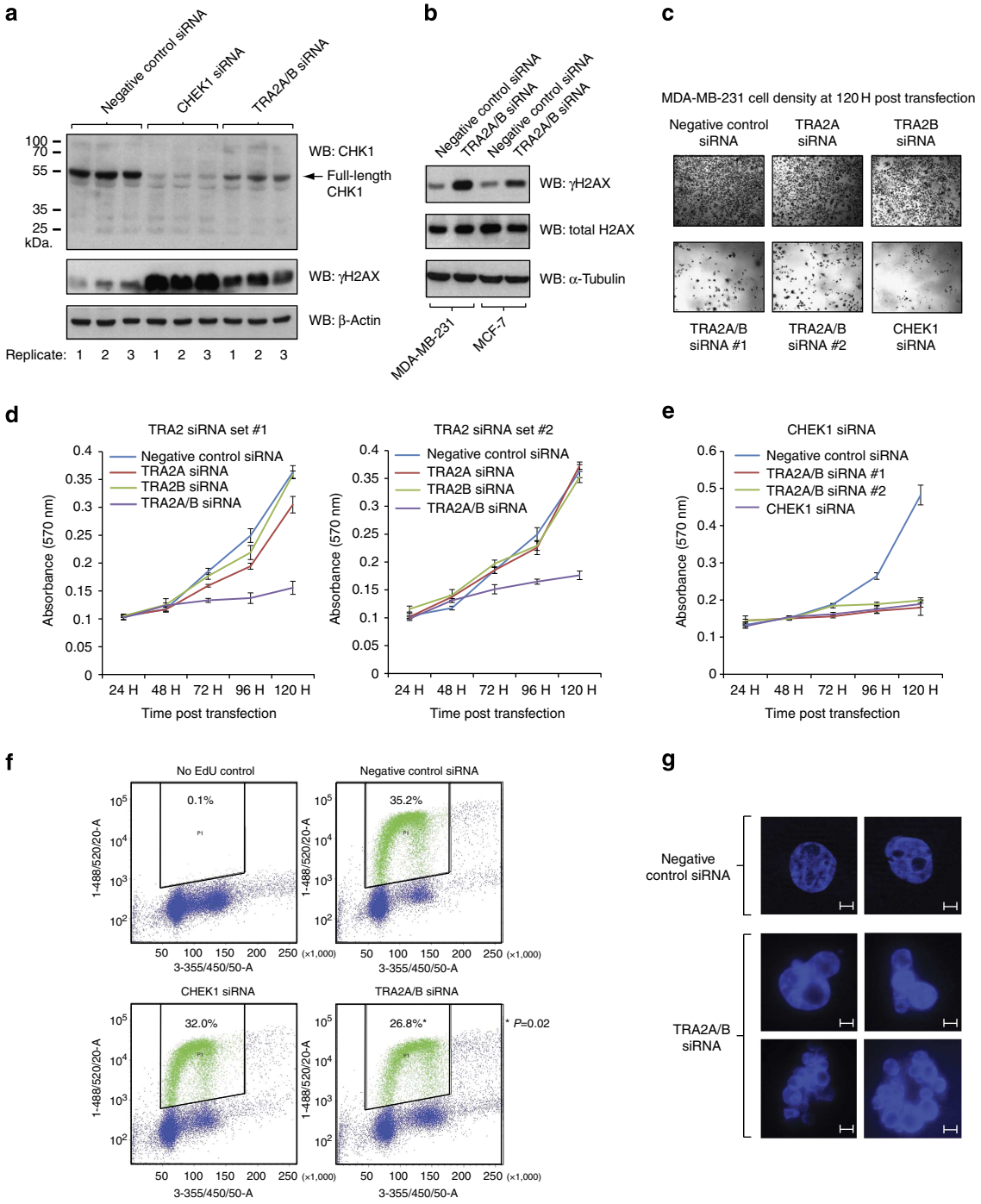


Figure 8 | Human Tra2 proteins are essential for expression of full-length CHK1 protein and to maintain cell viability. (**a**) Full-length CHK1 protein expression is depleted by siRNAs specific to *CHEK1* mRNA and also by joint siRNAs specific to the *TRA2A* and *TRA2B* mRNAs. In each case, samples from three independent replicate experiments were analysed in parallel. Also detected in these samples are levels of γH2AX and α-tubulin. (**b**) Expression of total H2AX and γH2AX after joint Tra2α and Tra2β depletion, or depletion with a control siRNA in MDA-MB-231 cells and MCF7 cells. (**c**) Measurement of cell density 120 h after transfection of siRNAs targeting different regions of the *TRA2A* and *TRA2B* mRNAs or *CHEK1* mRNA. (**d**) Joint (but not single) depletion of Tra2α and Tra2β proteins reduced MDA-MB-231 cell viability measured by MTT assays after siRNA transfection. (**e**) Depletion of CHK1 protein alone was sufficient to reduce viability of MDA-MB-231 cells measured by MTT assay at different time points after siRNA transfection. (**f**) Joint depletion of Tra2α and Tra2β reduced the proportion of EdU-positive MDA-MB-231 cells 96 h after siRNA transfection. Separate panels, shown clockwise from top left, show fluorescence-activated cell sorting analysis of control MDA-MB-231 cells incubated without EdU; cells transfected with a negative control siRNA and incubated with EdU; cells transfected with siRNAs specific for *TRA2A* and *TRA2B* and incubated with EdU; and cells transfected with a single siRNA specific to *CHEK1* and incubated with EdU. Probability (*P*) values were calculated using an independent two-sample *t*-test comparing the percentage of EdU-positive cells of negative control siRNA-treated cells and the gene-specific siRNA-treated cells (statistical significance: $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.0001$). Data were generated from three biological replicates. (**g**) Examples of abnormal nuclear morphology observed within cells transfected with siRNAs specific for *TRA2A* and *TRA2B* (lower panel) compared with the normal morphology seen in negative control siRNA-treated cells (upper panel). Cells were stained with 4′,6-diamidino-2-phenylindole, and these images were taken 96 h after siRNA transfection. The scale bar shows 5 μM. Uncropped western blots are shown in Supplementary Figs 11–15.

exposure, and decreased on siRNA treatment,Supplementary Fig. 7b). CHK1-S protein is reported to be regulated over the cell cycle and in tumours[37]. To test for cell cycle regulated splicing inclusion of *CHEK1* exon 3, we prepared RNAs from KG1 cell populations enriched in different cell cycle stages prepared using elutriation (Supplementary Fig. 5). When analysed by RT–PCR, very similar patterns of *CHEK1* exon 3 splicing inclusion were observed in each of the cell populations even though they contain different cell cycle profiles. We could detect high levels of both *CHEK1* splice isoforms in RNA purified from a small panel of breast cancer tissues, although we did not see an enrichment of either isoform in any particular tumour type at the RNA level (Supplementary Fig. 6). Overall, the above data are most consistent with Tra2α and Tra2β activity being essential for expression of full-length CHK1 protein rather than inducing expression of a shorter protein isoform of CHK1.

Tra2 protein depletion affects DNA damage and cell viability. Although it is not as a direct target of CHK1 phosphorylation, γH2AX has been used as a marker for the replication stress that can be induced by depleted CHK1 levels[38,39]. Similar to previous observations[38], we observed greatly increased levels of the DNA damage marker γH2AX following depletion of CHK1 protein by siRNA, compared with cells treated with a negative control siRNA (Fig. 8a). Increased γH2AX levels were also observed after joint depletion of Tra2α and Tra2β proteins in MDA-MB-231 cells and in MCF7 cells (Fig. 8a,b). The relative increased levels of γH2AX following CHEK1 or TRA2A/B siRNA treatment appear proportional to the reduction in full-length CHK1 protein expression observed by western blot.

Microscopy and MTT assays also indicated reduced cell viability 120 hours after joint Tra2α and Tra2β depletion (Fig. 8c,d). In contrast, single depletion of either Tra2α or Tra2β had negligible effect on cell viability compared with mock depleted cells. Similar results were obtained using two independent sets of siRNAs targeted at different regions of the mRNAs. This reduction in cell viability from joint removal of Tra2α and Tra2β, compared with the negligible effects of removing either protein alone, suggest that Tra2α and Tra2β are functionally interchangeable for maintaining cell viability in MDA-MB-231 cells, as well as in splicing control.

Depletion of CHK1 also reduced cell viability in MDA-MB-231 cells (Fig. 8e). This suggests that depletion of full-length CHK1 protein would likely be sufficient by itself to contribute to the loss of cell viability observed after joint Tra2α and Tra2β depletion. To test if re-introduction of full-length CHK1 protein would be sufficient to restore viability of joint Tra2α and Tra2β protein-depleted cells, we made a stable cell line in the FLP-in HEK-293 cell background in which a full-length FLAG-tagged CHK1 protein was expressed under control of a tetracycline promoter. Similar to the result obtained in MDA-MB-231 cells, joint depletion of Tra2α and Tra2β reduced cell viability in this stable HEK-293 cell line. However, although the full-length FLAG-tagged CHK1 protein was efficiently induced by tetracycline, it was not sufficient to rescue cell viability after joint Tra2 protein depletion (Supplementary Figs 8 and 9). While we cannot rule out that the tagged full-length CHK1 protein failed to rescue viability of this cell line for another reason, this result is consistent with multiple exons controlled by Tra2 proteins (including *CHEK1* exon 3) being important for cell viability.

Finally, we monitored incorporation of the thymidine analogue EdU using flow cytometry to determine whether joint Tra2 protein depletion affected cell proliferation of MDA-MB-231 cells (Fig. 8f). After joint Tra2 protein depletion, we observed a significant reduction in the proportion of EdU-positive cells 96 h after siRNA transfection (an 8.4% reduction, $P = 0.02$), indicating fewer cells had initiated DNA replication after joint Tra2 protein depletion. A slight reduction in the proportion of EdU-positive cells was observed after single CHK1 protein depletion, but this was not statistically significant when compared with negative control siRNA-treated cells. Joint Tra2 protein depletion also caused an increase in the proportion of cells containing abnormally shaped nuclei 96 h after siRNA transfection, consistent with major biological defects (Fig. 8g).

Discussion

Here we find that only joint depletion of both Tra2α and Tra2β proteins (and not single depletion of either protein alone) could induce substantial splicing switches in endogenous Tra2β target exons in MDA-MB-231 cells. This joint depletion strategy has enabled us to derive the most comprehensive map of dual Tra2-dependent target exons in any organism to date. Among the jointly regulated exons identified here was a key exon in the *CHEK1* gene, which encodes a protein essential for monitoring DNA damage and controlling cell cycle progression[37,39]. Exon 3 of the *CHEK1* gene is 224 nucleotides long; hence, skipping of this exon in the absence of Tra2 proteins would frameshift the reading frame of the *CHEK1* mRNA if a downstream translational initiation site is not selected[37]. Joint depletion of both Tra2 proteins quantitatively switched *CHEK1* pre-mRNA splicing, reduced expression of full-length CHK1 protein, and led to an increase in DNA damage as monitored by accumulation of γH2AX. We also confirmed *SMN2* exon 7 as a joint Tra2α/Tra2β target exon. Joint control by Tra2α provides an explanation why *SMN2* exon 7 is a target for Tra2β in transfected cells, but not appreciably affected in *Tra2b* single knockout mice[15].

This strategy also reveals that Tra2 proteins are required for splicing inclusion of some constitutively spliced exons. To the best of our knowledge, Tra2 proteins have only previously been described as alternative splicing factors[26]. Low levels of apparent alternative splicing of constitutive exons might be ascribed to error prone exon recognition by the spliceosome[40]. However, the *ANKRD1* and *SMC4* exons are not annotated as alternatively spliced in any tissue consistent with them being true constitutive exons, yet also show high-amplitude splicing changes upon Tra2 protein depletion. A role in constitutive splicing brings the Tra2 proteins closer to the core SR group in described molecular functions[23]. Consistent with this newly discovered role, Tra2β protein is fairly evenly expressed across mouse tissues, so would be available in most cells for splicing inclusion of constitutive exons[10,18]. Although we only detected exons activated by Tra2 proteins in this study, exons have previously been described that are repressed by Tra2 proteins[29,41]. Such repressed exons might have either eluded our search criteria or occur less frequently.

Genes containing Tra2-dependent exons are enriched in processes associated with chromosome biology. Although GO annotations are known to be incomplete and can differ in accuracy[42], this pattern of functional enrichment observed in our data is notably coherent. A potential association between the Tra2-dependent exons and chromosome biology is additionally supported by the connectivity of the BioGRID functional interaction network. In addition to the regulated *CHEK1* alternative splice, 6/17 of the strongest Tra2 protein-responsive exons (showing >40 point PSI change after Tra2 depletion) were also in genes involved in chromosome structure and epigenetic regulation, including *MSL3* (−41 point PSI exon switch after Tra2 depletion). *MSL3* is the human ortholog[43] of the *Drosophila melanogaster MSL3* gene, which regulates chromatin remodelling during sex determination, and is under control of Sex Lethal (a protein just upstream of Tra2 in the *Drosophila* sex determination

pathway) in flies. The *SMC4* gene (−66 point PSI exon switch after Tra2 depletion) encodes a protein important for DNA repair and chromosome condensation, and also interacts with the CTCF transcription factor that modifies chromatin structure[44]. The *ANKRD1* (−73 point PSI exon switch after Tra2 depletion) encodes a transcription factor, which is a negative regulator of cardiac genes[45]. Also among the genes with highly Tra2 protein-responsive exons was *GLYR1* (−41 point PSI exon switch after Tra2 depletion), which is a cofactor for histone demethylation[46], and the zinc-finger protein *ZCCHC9* (−68 point PSI exon switch after Tra2 depletion), which targets histone mRNAs for degradation[47]. *ZCCHC7* (−72 point PSI exon switch after Tra2 depletion) and *ZCCHC11* (−46 point PSI exon switch after Tra2 depletion) encode zinc-finger proteins homologous to ZCCHC9, but with roles in non-coding RNA metabolism[47–49]. The *MPHOSPH10* gene (−42 point PSI exon switch after Tra2 depletion) encodes a protein involved in ribosomal RNA processing in interphase, and is associated with chromosomes during mitosis[50]. Other genes involved in chromatin modification that are controlled by Tra2 proteins but did not fit into the most responsive group include the *NASP* (*NASP-T*, −12 point PSI exon switch after Tra2 depletion) and *ATRX* genes (−23 point PSI exon switch after Tra2 depletion)[18,51]. Interestingly, the two known *Drosophila* Tra2 splicing targets *Doublesex* and *Fruitless* are both transcription factors[12], and one of the major functions of the *Drosophila* sex determination pathway is dosage compensation via chromatin modification. Our data lend further support to the association of particular splicing regulators with the regulation of coherent cellular functions, also described for NOVA, RBFox2, PTB and T-STAR[4–7].

Our data indicate a high degree of functional redundancy between Tra2α and Tra2β, and a powerful homeostatic repressive feedback activity of Tra2β over Tra2α that buffers splicing changes when just one of these proteins is missing. Although splicing defects after single Tra2α and Tra2β depletion were small, they were often individually statistically significant (for example, in the *ATRX* gene). Such fine tuning of splicing profiles by joint Tra2 protein concentration by splicing feedback control might be important in whole organisms and at particular points of development (for example, in brain or testis development). Even individually, small splicing defects over many Tra2β-target exons might cumulatively cause physiological defects. This might explain why *Tra2b*-null mice are embryonic lethal despite containing a *Tra2a* gene[15,17,18]. Similar asymmetric expression patterns, in which a dominantly expressed splicing factor cross-regulates other family member proteins, have been found in the PTB family, where PTBP1 cross-regulates PTBP2 and PTBP3 (ref. 53). Comprehensive identification of PTBP1 targets similarly required joint depletion of PTBP1 and PTBP2 (ref. 53) Future studies of Tra2β-regulated splicing may also benefit by considering expression levels of both Tra2 proteins, rather than in the context of Tra2β expression alone.

Joint depletion of both Tra2 protein levels reduced cell viability in MDA-MB-231 cells, likely at least in part because of the requirement for productive splicing of the *CHEK1* mRNA. CHK1 protein expression is critical to reduce replication stress in cancer cells undergoing rapid proliferation driven by oncogenes including *RAS* and *MYC*[54,55]. Our data thus suggest the Tra2 proteins may represent novel targets to inhibit cancer cell growth.

Methods

Cell culture. MDA-MB-231 cells and MCF7 cells were maintained in DMEM (no phenol red) plus 10% fetal bovine serum and 1% Penicillin Streptomycin. HEK-293, HeLa and PC3 cells were maintained in DMEM plus 10% fetal bovine serum. Cells lines were originally purchased from the American Type Culture Collection and LGC Standards, Europe.

iCLIP. Triplicate iCLIP experiments were performed following the iCLIP protocol[33]. Briefly, MDA-MB-231 cells were irradiated with 400 mJ cm^{-2} ultraviolet-C light on ice, lysed and subject to partial RNase digestion. The crosslinked Tra2β-RNA complexes were then immunoprecipitated using Protein A Dynabeads (Invitrogen) and a rabbit polyclonal anti-Tra2β antibody (Abcam, ab31353). cDNA libraries were prepared according to the published iCLIP protocol. High throughput sequencing of cDNA libraries was performed using an Illumina GAIIx.

RNA-seq. RNA was extracted from cells using RNeasy Plus Mini Kit (Qiagen) following manufacturer's instructions and re-suspended in nuclease-free water. All RNA samples were DNase treated using DNA-free kit (Invitrogen) and stored at −80 °C prior to RNA quality control check using 2100 Agilent Bioanalyser and mRNA library prep using TruSeq mRNA library kit (Illumina). Pair-end sequencing was done in total for six samples (three biological replicates of negative control siRNA-treated cells and three biological replicates from *TRA2A* and *TRA2B* siRNA-treated cells) using an Illumina HiSeq 2000.

Bioinformatics (iCLIP and RNA-seq analysis). iCLIP data analysis, crosslink site identification and quantification, randomization of iCLIP positions and pentamer enrichment analysis were performed according to published procedures[33]. Briefly, we used the human genome annotation version hg19, and gene annotations from Ensembl 59. Experiment barcode and random barcodes were registered and removed from iCLIP reads. After trimming, we ignored reads shorter than 11 nucleotides. Remaining trimmed reads were then mapped using Bowtie[56], allowing two mismatches and accepting only reads with single hits. Crosslink sites were initially identified as the first nucleotide upstream of the iCLIP tag, and then filtered to determine statistically significant crosslink sites and those which occurred in clusters within 15 nucleotides windows and with a significant iCLIP tag count, compared with randomized positions, as described in Konig *et al.*[56]. For RNA-seq analysis, the base quality of raw sequencing reads were checked with FastQC (ref. 57) and refined with Seqtk (ref. 58) and Trim-galore (ref. 59). Reads were mapped to the hg19 reference with Tophat2 (ref. 60) and matches analyzed with Bedtools (ref. 61). Differentially expressed genes and exon usage were determined with DESeq (ref. 62) and DEXSeq (ref. 63) respecitvely.

siRNA transfection. Efficient knockdown of endogenous Tra2α, Tra2β and CHK1 proteins were achieved by transfecting cells with Silencer Select Pre-designed siRNAs (Ambion), targeting *TRA2A* mRNA (Ambion IDs: s26664 and s26665), *TRA2B* mRNA (Ambion IDs: s12749 and s12751) or *CHEK1* mRNA (Ambion ID: s503) respectively, with siPORT NeoFX Transfection Agent (Ambion). Control cells were transfected with a negative control siRNA (Ambion Cat#: 4390843). MDA-MB-231 cells grown in 100 mm tissue culture dishes were transfected with either 24 µl of 10 µM negative control siRNA (control), 12 µl of 10 µM siRNA targeting *TRA2A, TRA2B or CHEK1* (single Tra2α, Tra2β or CHK1 knockdown) or 12 µl of 10 µM siRNA targeting *TRA2A* and 12 µl of 10 µM siRNA targeting *TRA2B* (joint Tra2α and Tra2β knockdown). Cells were incubated for 72 h post siRNA transfection, before RNA extraction or western blotting.

Splicing assays: RNA extraction, RT–PCR and PCR. RNA was extracted using standard Trizol RNA extraction. cDNA was synthesized from 500 ng total RNA in a 10 µl reaction, using Superscript VILO cDNA synthesis kit (Invitrogen) following manufacturer's instructions. Splicing profiles were monitored by PCR using primers in flanking exons. For each PCR, 1 µl diluted cDNA (1/8) was used as template in a 10 µl PCR reaction using Phusion High-Fidelity PCR Kit (NEB, UK) following manufacturer's instructions. Splicing profiles were monitored and quantified using the Qiaxcel capillary electrophoresis system (Qiagen) and PSI was calculated as described previously[18]. All primers used for splicing assays are provided in Supplementary Data 3.

Quantitative PCR. Relative gene expression was determined by quantitative real-time PCR using the SYBR Green PCR Master Mix kit (Applied Biosystems) and an Applied Biosystems 7900HT Fast Real-Time PCR Machine. cDNA was generated from equal quantities of total RNA for each sample using Superscript VILO cDNA synthesis kit (Invitrogen) following manufacturer's instructions. Gene expression was calculated relative to three housekeeping genes *ACTB*, *GAPDH* and *TUBB*. C_t values for each sample were calculated using SDS 2.4 software (Applied Biosystems) and relative mRNA expression was calculated using the $2^{-\Delta\Delta Ct}$ method.

Calculation of Tra2β binding site density. Tra2β binding site density was calculated as the percentage of nucleotides within an exon that correspond to the top 10 kmers identified from the Tra2β iCLIP experiments (Supplementary Fig. 2b).

Detection of proteins using western blotting. Endogenous proteins were detected by western blot analysis using the following primary antibodies and dilutions: Tra2α (Novus Biologicals, H00029896-B01P;1:500 dilution), Tra2β (Abcam, ab31353;1:2,000 dilution), CHEK1 (Proteintech, 10362-1-AP;1:250 dilution), Histone H2AX (Santa Cruz Biotechnology, sc54-606;1:500 dilution),

γH2AX (Ser 139) (Santa Cruz Biotechnology, sc-101696;1:500 dilution), FLAG (Sigma-Aldrich, F3040;1:2,000 dilution), β-Actin (Sigma-Aldrich, A5441,1:2,000 dilution) and α-Tubulin (Sigma-Aldrich, T5168;1:2,000 dilution).

***ANKRD1 and CHEK1* minigene construction and mutagenesis.** The *ANKRD1* constitutive exon and ~200 nucleotides of flanking intronic region was amplified from human genomic DNA using the cloning primers ANKRD1 F (5′-AAAA AAAAAGAATTCAAAATCTAAGACTTGCTTATGGCATT-3′) and ANKRD1 R (5′AAAAAAAAAGAATTCAGCATGAGAGTTACCGTGAGC-3′). The PCR products were digested with *BamH1* restriction enzyme and cloned into the pXJ41 vector[64] using the *Mfe*1 site midway through the 757 nucleotide rabbit β-globin intron 2. Tra2β binding site mutations were made using site directed mutagenesis with the following primers; ANKRD1 M1F (5′- AGAACACATATCAAAGCTT GCACATTTATACGACCTTGAAA-3′), ANKRD1 M1R (5′-CAAGGTCGTATAA ATGTGCAAGCTTTGATATGTGTTCTAG-3′), ANKRD1 M2F (5′-ATCATT CAACTGCAGCAACGGCAACAATACAGGCACACTAAAG-3′) and ANKRD1 M2R (5′-GAACTTTAGTGTGCCTGTATTGTTGCCGTTGCTGCAGTTGAA TG-3′). The *CHEK1* alternative exon and approximately 250 nucleotides of flanking intronic region was synthesised *in vitro* and similarly cloned into the pXJ41 vector. A mutated version that disrupted Tra2β binding sites was also synthesised (sequence provided in Fig. 7c) and cloned into the pXJ41 vector. Analysis of splicing patterns of mRNAs transcribed from minigenes was carried out in HEK-293 cells as previously described[18,24], using primers within the β-globin exons of pXJ41; PXJRTF (5′-GCTCCGGATCGATCCTGAGAACT-3′) and PXJB (5′-GCTGCAATAAACAAGTTCTGCT-3′).

EMSAs. Gel shift experiments[18,65] were performed using full-length Tra2β protein and *in vitro*-translated RNA probes made from constructs containing amplified regions of the human *CHEK1* gene, cloned into the pBluescript vector. Three regions of the human *CHEK1* gene were amplified using the following primers: CHEK1 intronic F (5′-AAAAAAAAAGGTACCTGTGTACCTCTCCTTCACTA CC-3′), CHEK1 intronic R (5′-AAAAAAAAAGAATTCCTGTCCTAAGCTCCT ATGGGG-3′), CHEK1 exon region A F (5′-AAAAAAAAAGGTACCgttcaacttgc tgtgaatagagt-3′), CHEK1 exon region A R (5′-AAAAAAAAAGAATTCggcacg CTTCAtatctacaATCT-3′), CHEK1 exon region B F (5′-AAAAAAAAAGGTACC agtaaaattctatggtcacagga-3′) and CHEK1 exon region B R (5′-AAAAAAAAAGAA TTCctccactacagtactccagaaat-3′).

GO and functional network analysis. GO[66] enrichment analysis was carried out using the Bioconductor GOstats package version 2.24.0refs [67, 68]. Enrichments of GO biological process terms were calculated using the conditional hypergeometric test with a significance cut-off of 0.001 and using a background of genes that are normally expressed in MDA-MB-231 cells. Annotations were taken from the Bioconductor *Homo sapiens* annotation package org.Hs.eg.db version 2.8.0ref. [69]. The analysis was run in the open source statistical package R version 3.0.1ref. [70].

Interaction data for *Homo sapiens* was retrieved from the BioGRID database (version 110). These data were integrated into a network in which nodes represented genes or gene products, and edges represented any type of BioGRID interaction between the nodes. The network was visualised using the Cytoscape visualization platform[71], and was coloured based on annotations to top five enriched GO biological processes (as downloaded from QuickGO[71]). Where a protein was annotated to more than one term, the most specific annotation was chosen.

MTT assay. MTT assays were performed using MTT Cell Proliferation Assay Kit (Cayman Chemical), following manufacturer's instructions. An siRNA transfection mix was added to a suspension of $\sim 2 \times 10^5$ MDA-MB-231 cells in 10 ml media. The siRNA/cell suspension was gently mixed and a 100 µl aliquot was added per well to a 96-well plate. Absorbance from the MTT assay was measured at 24, 48, 72, 96 and 120 h after siRNA transfection/seeding of cells. Relative density of cells was also compared 120 h after seeding cells by microscopy.

Fluorescence-activated cell sorting analysis of EdU-positive cells. MDA-MB-231 cells were incubated with 10 µM EdU for 4 h, 96 h after siRNA transfection. Cell fixation, permeabilization and EdU detection was performed using the Click-iT EdU Flow Cytometry Assay Kit (Life Technologies) following the manufacturer's instructions. Data were collected and analysed using a BD LSR II flow cytometer using 488 nm excitation and a 520/20 band-pass for detection of EdU Alexa Fluor488 azide and 355 nm excitation and a 450/50 band-pass for detection of 4',6-diamidino-2-phenylindole. Experiments were performed with biological triplicate samples and 30,000 cells were analysed per sample. A no-EdU control sample was used to inform our gating strategy to calculate the proportion of EdU-positive cells.

Analysis of nuclear morphology. To investigate nuclear morphology, cells were fixed with 4% paraformaldehyde followed by nuclear staining with 4',6-diamidino-2-phenylindole, 96 h after siRNA transfection (siRNA transfection as described above).

Elutriation and cell cycle evaluation. Elutriation: Cells were size fractionated by centrifugal elutriation, using flow rates of 10, 13, 17, 20, 24 and 28 ml min^{-1} (ref. 72). Cell cycle evaluation: Cell cycle phase enrichment of cells was assessed using immunofluorescence staining for CENPF (late S, G_2, G_2/M)[73] and phospho-histone H3S10 (G_2/M, M)[74]. Asynchronous and elutriated KG1 cells were suspended in PBS, spotted onto poly-lysine-coated slides and processed for immunofluorescence. Images were captured and cells were scored for CENPF and phospho-H3S10 staining.

Generation of tetracycline-inducible HEK-293 cells. Full-length CHK1-FLAG cDNA was amplified from the pcDNA4-Chk1-Flag plasmid (Addgene plasmid #22894) using the primers CHEK1 FLAG F (5′-AAAAAAAAAGCGGCCGC atggcagtgccctttgtggaagac-3′) and CHEK1 FLAG R (5′-AAAAAAAAAGTCGAC tcatgtggcaggaagccaaatcttc-3′) and cloned into the Flp-In expression vector (pCDNA5). To generate an inducible cell line, the CHK1-FLAG-pCDNA5 vector was cotransfected with the Flp recombinase plasmid (pOG44) into Flp-In HEK-293 cells and selected for using treatment with Hygromycin B. Following Hygromycin B selection, CHK1-FLAG expression was induced by the addition of tetracycline to promote expression CHK1-FLAG expression via a tetracycline-inducible promoter.

References

1. Fox-Walsh, K. L. & Hertel, K. J. Splice-site pairing is an intrinsically high fidelity process. *Proc. Natl Acad. Sci. USA* **106,** 1766–1771 (2009).
2. Kelemen, O. *et al.* Function of alternative splicing. *Gene* **514,** 1–30 (2013).
3. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489,** 101–108 (2012).
4. Ehrmann, I. *et al.* The tissue-specific RNA binding protein T-STAR controls regional splicing patterns of neurexin pre-mRNAs in the brain. *PLoS Genet* **9,** *e1003474 (2013).*
5. Venables, J. P. *et al.* RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. *Mol. Cell Biol.* **33,** 396–405 (2013).
6. Ule, J. *et al.* Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet.* **37,** 844–852 (2005).
7. Zhang, C. *et al.* Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22,** 2550–2563 (2008).
8. Llorian, M. *et al.* Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat. Struct. Mol. Biol.* **17,** 1114–1123 (2010).
9. Tsuda, K. *et al.* Structural basis for the dual RNA-recognition modes of human Tra2-beta RRM. *Nucleic Acids Res.* **39,** 1538–1553 (2011).
10. Elliott, D.J., Best, A., Dalgliesh, C., Ehrmann, I. & Grellscheid, S. How does Tra2beta protein regulate tissue-specific RNA splicing? *Biochem Soc Trans* **40,** 784–8 (2012).
11. Baker, B. S. Sex in flies: the splice of life. *Nature* **340,** 521–524 (1989).
12. Forch, P. & Valcarcel, J. Splicing regulation in *Drosophila* sex determination. *Prog. Mol. Subcell. Biol.* **31,** 127–151 (2003).
13. Dauwalder, B., Amaya-Manzanares, F. & Mattox, W. A human homologue of the *Drosophila* sex determination factor transformer-2 has conserved splicing regulatory functions. *Proc. Natl Acad. Sci. USA* **93,** 9004–9009 (1996).
14. Beil, B., Screaton, G. & Stamm, S. Molecular cloning of htra2-beta-1 and htra2-beta-2, two human homologs of tra-2 generated by alternative splicing. *DNA Cell Biol.* **16,** 679–690 (1997).
15. Mende, Y. *et al.* 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,** 2154–2167 (2010).
16. Roberts, J. M. *et al.* Splicing factor TRA2B is required for neural progenitor survival. *J. Comp. Neurol.* **522,** 372–392 (2014).
17. Storbeck, M. *et al.* Neuronal-specific deficiency of the splicing factor tra2b causes apoptosis in neurogenic areas of the developing mouse brain. *PLoS ONE* **9,** e89020 (2014).
18. Grellscheid, S. *et al.* Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2beta in development. *PLoS Genet.* **7,** e1002390 (2011).
19. Best, A. *et al.* Expression of Tra2 beta in cancer cells as a potential contributory factor to neoplasia and metastasis. *Int. J. Cell Biol.* **2013,** 843781 (2013).
20. Hofmann, Y., Lorson, C. L., Stamm, S., Androphy, E. J. & Wirth, B. 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. USA* **97,** 9618–9623 (2000).
21. Glatz, D. C. *et al.* The alternative splicing of tau exon 10 and its regulatory proteins CLK2 and TRA2-BETA1 changes in sporadic Alzheimer's disease. *J. Neurochem.* **96,** 635–644 (2006).
22. Long, J. C. & Caceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* **417,** 15–27 (2009).
23. Zhou, Z. & Fu, X. D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* **122,** 191–207 (2013).

24. Venables, J. P. *et al.* Up-regulation of the ubiquitous alternative splicing factor Tra2beta causes inclusion of a germ cell-specific exon. *Hum. Mol. Genet.* **14,** 2289–2303 (2005).
25. Busch, A. & Hertel, K. J. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip. Rev. RNA* **3,** 1–12 (2012).
26. Tacke, R., Tohyama, M., Ogawa, S. & Manley, J. L. Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* **93,** 139–148 (1998).
27. Clery, A. *et al.* Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-beta1. *Nat. Struct. Mol. Biol.* **18,** 443–450 (2011).
28. Uren, P. J. *et al.* Site identification in high-throughput RNA-protein interaction data. *Bioinformatics* **28,** 3013–3020 (2012).
29. Anderson, E. S. *et al.* The cardiotonic steroid digitoxin regulates alternative splicing through depletion of the splicing factors SRSF3 and TRA2B. *RNA* **18,** 1041–1049 (2012).
30. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446,** 926–929 (2007).
31. Stoilov, P., Daoud, R., Nayler, O. & Stamm, S. Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. *Hum. Mol. Genet.* **13,** 509–524 (2004).
32. Saltzman, A.L., Pan, Q. & Blencowe, B. J. Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev* **25,** 373–84 (2011).
33. Konig, J. *et al.* iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J. Vis. Exp.* **50:** pii 2638 (2011).
34. Dreszer, T. R. *et al.* The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res.* **40,** D918–D923 (2012).
35. Meyer, L. R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* **41,** D64–D69 (2013).
36. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34,** D535–D539 (2006).
37. Pabla, N., Bhatt, K. & Dong, Z. Checkpoint kinase 1 (Chk1)-short is a splice variant and endogenous inhibitor of Chk1 that regulates cell cycle and DNA damage checkpoints. *Proc. Natl Acad. Sci. USA* **109,** 197–202 (2012).
38. Gagou, M. E., Zuazua-Villar, P. & Meuth, M. Enhanced H2AX phosphorylation, DNA replication fork arrest, and cell death in the absence of Chk1. *Mol. Biol. Cell* **21,** 739–752 (2010).
39. Syljuasen, R. G. *et al.* Inhibition of human Chk1 causes increased initiation of DNA replication, phosphorylation of ATR targets, and DNA breakage. *Mol. Cell Biol.* **25,** 3553–3562 (2005).
40. Melamud, E. & Moult, J. Stochastic noise in splicing machinery. *Nucleic Acids Res.* **37,** 4873–4886 (2009).
41. Chandler, D. S., Qi, J. L. & Mattox, W. Direct repression of splicing by transformer-2. *Mol. Cell Biol.* **23,** 5174–5185 (2003).
42. Skunca, N., Altenhoff, A. & Dessimoz, C. Quality of computationally inferred gene ontology annotations. *PLoS Comput. Biol.* **8,** e1002533 (2012).
43. Smith, E.R. *et al.* A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol Cell Biol* **25,** 9175–88 (2005).
44. Huang, K. *et al.* Ribosomal RNA gene transcription mediated by the master genome regulator protein CCCTC-binding factor (CTCF) is negatively regulated by the condensin complex. *J Biol Chem* **288,** 26067–77 (2013).
45. Zolk, O. *et al.* Cardiac ankyrin repeat protein, a negative regulator of cardiac gene expression, is augmented in human heart failure. *Biochem Biophys Res Commun* **293,** 1377–1382 (2002).
46. Fang, R. *et al.* LSD2/KDM1B and its cofactor NPAC/GLYR1 endow a structural and molecular model for regulation of H3K4 demethylation. *Mol. Cell* **49,** 558–570 (2013).
47. Schmidt, M. J., West, S. & Norbury, C. J. The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. *RNA* **17,** 39–44 (2011).
48. Fasken, M. B. *et al.* Air1 zinc knuckles 4 and 5 and a conserved IWRXY motif are critical for the function and integrity of the Trf4/5-Air1/2-Mtr4 polyadenylation (TRAMP) RNA quality control complex. *J. Biol. Chem.* **286,** 37429–37445 (2011).
49. Hagan, J. P., Piskounova, E. & Gregory, R. I. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **16,** 1021–1025 (2009).
50. Westendorf, J.M. *et al.* M phase phosphoprotein 10 is a human U3 small nucleolar ribonucleoprotein component. *Mol Biol Cell* **9,** 437–49 (1998).
51. Eustermann, S. *et al.* Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. *Nat. Struct. Mol. Biol.* **18,** 777–782 (2011).
52. Ule, J. *et al.* An RNA map predicting Nova-dependent splicing regulation. *Nature* **444,** 580–586 (2006).
53. Spellman, R., Llorian, M. & Smith, C. W. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol. Cell* **27,** 420–434 (2007).
54. Bartek, J., Mistrik, M. & Bartkova, J. Thresholds of replication stress signaling in cancer development and treatment. *Nat. Struct. Mol. Biol.* **19,** 5–7 (2012).
55. Murga, M. *et al.* Exploiting oncogene-induced replicative stress for the selective killing of Myc-driven tumors. *Nat. Struct. Mol. Biol.* **18,** 1331–1335 (2011).
56. Konig, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17,** 909–15 (2010).
57. Andrews, S. FastQC A quality control tool for high throughput sequence data. Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
58. Li, H. seqtk. Available at https://github.com/lh3/seqtk.
59. Krueger, F. Trim Galore! Available at http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
60. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14,** R36 (2013).
61. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
62. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biology* **11,** R106 (2010).
63. Anders, S. *et al.* Detecting differential usage of exons from RNA-seq data. *Genome Research* **22,** 2008–2017 (2012).
64. Bourgeois, C. F., Popielarz, M., Hildwein, G. & Stevenin, J. Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation. *Mol. Cell. Biol.* **19,** 7347–7356 (1999).
65. Grellscheid, S. N. *et al.* Molecular design of a splicing switch responsive to the RNA binding protein Tra2beta. *Nucleic Acids Res.* **39,** 8092–8104 (2011).
66. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29 (2000).
67. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23,** 257–258 (2007).
68. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).
69. Carlson, M. org.Hs.eg.db: Genome wide annotation for Human. R package version 2.8.0. (2012).
70. Rcoreteam. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
71. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–2504 (2003).
72. Ly, T. *et al.* A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *Elife* **3,** e01630 (2014).
73. Liao, H., Winkfein, R. J., Mack, G., Rattner, J. B. & Yen, T. J. CENP-F is a protein of the nuclear matrix that assembles onto kinetochores at late G2 and is rapidly degraded after mitosis. *J. Cell Biol.* **130,** 507–518 (1995).
74. Hendzel, M. J. *et al.* Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* **106,** 348–360 (1997).

Acknowledgements

We thank Professor Jernej Ule and Dr Julian Konig for help with the iCLIP analysis, and Professor Ian Eperon, Dr Ingrid Ehrmann and Dr Jennifer Munkley for comments on the manuscript. Andrew Best was a Breast Cancer Campaign funded PhD student. This project was supported by the Breast Cancer Campaign (Grant 2009NovPhd21), the Wellcome Trust (Grant numbers WT080368MA and WT089225/Z/09/Z), the BBSRC (grant numbers BB/D013917/1 and BB/I006923/1), a Canadian Institute of Health Research grant to B.C. (MOP93917) and the Leukaemia and Lymphoma Research Specialist Programme Grant 12031 (to C.A.A. and I.G.C). B.K. was a BHF Professor of Cardiology.

Author contributions

A.B. performed iCLIP, RNA-seq, splicing and cellular analyses, and performed experiments; K.J. performed GO and BioGrid analyses; C.D. performed EMSAs; E.H. helped with iCLIP experiments; M.K.-K. made stable HEK-293 cell lines; T.C. performed iCLIP sequence analysis; Y.X. and M.S.K. performed RNA-seq bioinformatic analysis; M.D. made minigenes; R.H. and B.K. performed iCLIP sequencing; A.W. helped with analysis; J.P.V., R.K. and B.C. analysed the role of multiple RNA binding proteins for CHEK1, SMC4, GLYR1 and ANKRD1; I.G.C., K.C.L. and C.A.A. purified and characterized KG1 cell fractions; A.B., A.T.-C. and D.J.E. were involved in study design and data analysis. A.B. and D.J.E. prepared the manuscript. All authors discussed the results and commented on the manuscript.

Additional information

Accession codes: iCLIP and RNA-seq data available via the Gene Expression Omnibus (GEO) using GEO accession number GSE59335.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Best, A. *et al.* Human Tra2 proteins jointly control a *CHEK1* splicing switch amongst alternative and constitutive target exons. *Nat. Commun.* 5:4760 doi: 10.1038/ncomms5760 (2014).

Tra2 protein biology and mechanisms of splicing control

Andrew Best*, Caroline Dalgliesh*, Mahsa Kheirollahi-Kouhestani*, Marina Danilenko*, Ingrid Ehrmann*, Alison Tyson-Capper† and David J. Elliott*[1]

*Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne NE1 3BZ, U.K.
†Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne NE1 3BZ, U.K.

Abstract

Tra2 proteins regulate pre-mRNA splicing in vertebrates and invertebrates, and are involved in important processes ranging from brain development in mice to sex determination in fruitflies. In structure Tra2 proteins contain two RS domains (domains enriched in arginine and serine residues) flanking a central RRM (RNA recognition motif). Understanding the mechanisms of how Tra2 proteins work to control splicing is one of the key requirements to understand their biology. In the present article, we review what is known about how Tra2 proteins regulate splicing decisions in mammals and fruitflies.

Introduction

Tra2 proteins are nuclear RNA-binding proteins involved in splicing regulation (Figure 1A). Tra2 proteins are conserved across the animal kingdom. Invertebrates have a single Tra2 protein copy (of which the fruitfly Tra2 homologue is the best characterized), whereas vertebrates have two distinct protein copies called Tra2α and Tra2β. Hence a gene duplication probably took place early in vertebrate evolution, with invertebrates retaining one copy and vertebrates obtaining two copies.

In terms of functions during development, fruitfly Tra2 protein is required for female sex determination, and also has roles in spermatogenesis [1,2]. Expression of human Tra2α protein in *Tra2*-null fruitflies is able to partially rescue female sex determination and replace the endogenous Tra2 protein in some but not all regulated splicing events. This shows substantial functional conservation between invertebrate and vertebrate Tra2 homologues [3]. In mammals, gene knockout experiments have shown Tra2β protein is essential for mouse embryonic development [4]. Ubiquitous knockout of the *Tra2b* gene results in disorganized embryos at E7.5 (embryonic day 7.5) that die during early development. Cell-specific conditional deletion of the *Tra2b* gene in the nervous system prevents proper brain development [5–7]. Mice with a cortex-specific *Tra2b* knockout (a conditional knockout driven by *Emx-cre* in the brain) survive to adulthood, but have severe brain abnormalities due to death of cortical neural progenitor cells by apoptosis [6]. Mice with broader neuronal knockout of the *Tra2b* gene (a conditional knockout driven by *Nestin-cre*) die shortly after birth with severe abnormalities in the cortex and thalamus [5].

In terms of their physiological functions, mouse Tra2β protein expression affects the properties of smooth muscle cells via splicing regulation of a myosin phosphatase targeting subunit [8], and is also targeted by the cardiotonic steroid digitonin [9]. Levels of expression of the *Drosophila* Tra2 protein are regulated by the DNA topisomerase inhibitor drug camptothecin [10]. Tra2β protein is both up-regulated and aberrantly localized in the nuclei of retinal cells in age-related macular degeneration [11], and differentially expressed on aging in humans [12]. Tra2β is also differentially expressed in some cancers (this activity of Tra2β was recently reviewed [13]), and Tra2β is implicated in the pathology of other diseases including SMA (spinal muscular atrophy) [14], Alzheimer's disease [15] and FTDP-17 (frontotemporal dementia and parkinsonism linked to chromosome 17) [16].

Central to understanding the biological actions of Tra2 proteins is how they regulate splicing. In the present article, we review the biology of these interesting proteins, and particularly what is known about the mechanisms through which they regulate splicing decisions.

Tra2 proteins contain N- and C-terminal domains enriched in serine and arginine residues

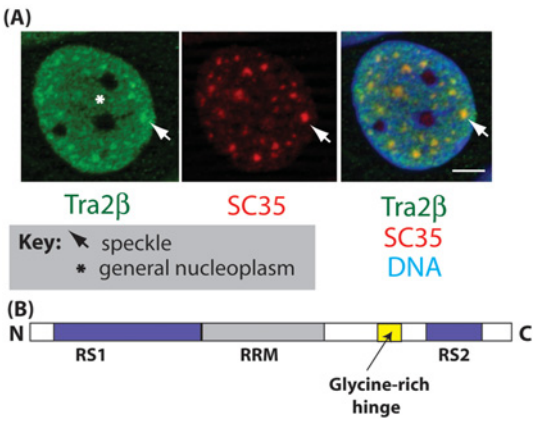
Tra2 proteins belong to a larger family of RS-domain (domains enriched in serine and arginine residues) containing proteins. Among these proteins Tra2 proteins have a distinctive modular structure, including N- and C-terminal RS-domains called the RS1 and RS2 domains (Figure 1B). Although there are two RS domains in Tra2 proteins these likely have somewhat distinct functions. In humans, the RS1

Key words: alternative splicing, *Drosophila*, gene expression, RNA recognition motif, RS domain, Tra2.

Abbreviations: DOA, Darkener of Apricot; ESE, exonic splicing enhancer; hnRNP, heterogeneous nuclear ribonucleoprotein; ISS, intron splicing silencer; RRM, RNA recognition motif; RS domain, domain enriched in arginine and serine residues; snRNP, small nuclear ribonucleoprotein.

[1]**To whom correspondence should be addressed (email David.Elliott@ncl.ac.uk).**

Figure 1 | **Tra2 proteins are nuclear RNA-binding proteins**
(**A**) Human Tra2β is found within the general nucleoplasm and also in nuclear speckles. Here localization of Tra2β (green) is shown in the nucleus of a human Saos2 cell alongside SC35 (red) and DNA (blue, stained with DAPI). The yellow signal represents overlap. The general nucleoplasm (shown by the asterisk) is where active gene expression takes place. Nuclear speckles (arrowed) are storage sites for splicing components and sites of post-transcriptional splicing [59]. Scale bar, 5 μm. (**B**) General structure of Tra2 proteins.



domain is longer than the RS2 domain. The RS1 domain is essential for splicing control in human cells [17], whereas RS2 is essential for controlling *doublesex* (creates male- and female-specific transcription factors) splicing patterns in fruitflies [18]. The fruitfly RS1 and RS2 domains also have different splicing activation activities in experiments where Tra2–MS2 fusion proteins were targeted to RNA through an MS2-binding site, again consistent with these RS domains not being functionally equivalent [19]. There are at least three nuclear localization signals within the RS1 domain that mediate nuclear import, and also two targeting sequences in RS1 that direct Tra2β protein to nuclear speckles once within the nucleus [20]. However, either one of the RS1 and RS2 domains are sufficient for nuclear import of Tra2β proteins in human cells [20]. In structure, the Tra2 proteins are similar to the core group of SR proteins that contain single RS domains and one or two RRMs (RNA recognition motifs) [21–23].

The serines in the RS domains of Tra2 proteins are modified by cycles of protein phosphorylation and dephosphorylation. Similar to other SR proteins, protein phosphorylation in the cytoplasm by the kinase SRPK1 is important for the nuclear import of Tra2 proteins from the cytoplasm. Some SR proteins including SRSF1 contain extended runs of serine–arginine repeats, and are very efficiently serine phosphorylated by the SRPK1 catalytic domain after attachment to a docking groove in SRPK1 [24]. Although it has two RS domains, the RS content of these domains are less in Tra2 proteins compared with some of the core SR proteins such as SRSF1. This makes Tra2β phosphorylation by SRPK1 less efficient than for SRSF1, and phosphorylation of the C-terminal RS domain of Tra2β has been shown to proceed with different kinetics and without use of the SRPK1 docking groove [24].

Once inside the nucleus phosphorylation of Tra2 proteins is carried out by a family of LAMMER kinases. A single LAMMER kinase is found in fruitflies called Darkener of Apricot (abbreviated DOA) [25]. Phosphorylation of fruitfly Tra2 protein by DOA relocalizes Tra2 from nuclear speckles (which are sites of splicing factor storage) into the nucleoplasm (where active steps of gene expression take place) [26]. Like Tra2 proteins, the DOA kinase is also essential for female sex determination in fruitflies. DOA phosphorylation of Tra2 is needed for splicing regulation of *Doublesex* exon 4 in females but not for female-specific splicing regulation of *Fruitless* mRNA (see below for more details about these regulated splicing events) indicating that different alternative events might depend on different phosphorylated serine residues in Tra2 [26].

Tra2 proteins contain a single central RRM

The RS domains of Tra2 proteins flank a central RRM which binds to target RNA sequences. Tra2β protein has a dual mode of RNA binding that enables it to bind to both CA-rich sequences and GAA-rich RNA sequences in target RNAs [27–29], and Tra2β regulates pre-mRNAs containing both kinds of binding sequence (see below). Tra2β protein–RNA interactions have been dissected at atomic resolution, and involve numerous hydrogen bonds and base stacking interactions between the RRM and GAA repeats [30,31]. The RRM of Tra2β uses a slightly different mode of binding to enable it to interact with single-stranded CAA sequences embedded in stem–loop structures [31].

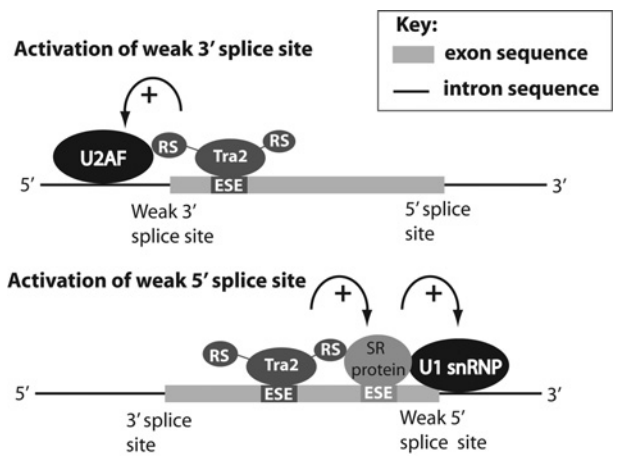
Transcriptome-wide RNA targets for mouse Tra2β have been mapped by HITS-CLIP (high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation) experiments [17]. The most frequent pentamers recovered in these experiments were AGAA-rich. The binding data are consistent with mouse Tra2β protein binding preferentially to exon sequences, but there were also binding sites identified in introns and intergenic sequences [17]. Recently Tra2β protein has also been found to interact with ribonucleoprotein from L1 retrotransposons, so might be involved in the nuclear metabolism of these genome parasites [32].

Tra2 proteins as splicing activators

The best characterized function of Tra2 proteins is in splicing regulation. Tra2 proteins frequently operate as splicing activators through binding to ESEs (exonic splicing enhancers). Binding of Tra2 proteins to ESE sequences closely adjacent to regulated splice sites acts to stabilize the assembly of the spliceosome, and this in turn enhances the use of these regulated splice sites (Figure 2) [33]. However, Tra2β binding to an exon is not sufficient to predict effects on splicing regulation. Tra2β-binding sites have been found in some alternative exons that are not under its direct control, and the

Figure 2 | Models suggesting regulatory interactions of Tra2 proteins with other splicing components that might be important for exon definition

Most animal pre-mRNAs are initially deciphered by interactions across exons; this is called exon definition [60,61]. Protein–protein interactions have been identified between Tra2 and U2AF35 proteins, and these interactions might stabilize spliceosome assembly on weak 3′ splice sites that cannot stably bind U2AF35. No direct interactions have been identified with U1 snRNP components at the 5′ splice site, but Tra2 proteins might interact indirectly with the U1 snRNP components through interactions with SR proteins like SRSF1 that are themselves bound to U1 snRNP.



rules governing this are not well understood [17]. Hence it is important for splicing control elements potentially involving Tra2 proteins to be functionally validated (e.g., by minigene experiments with overexpressed Tra2 proteins).

In principle, Tra2 proteins could activate splicing inclusion by helping binding of spliceosome components at weak 5′ or 3′ splice sites [weak splice sites not fitting the consensus sequences best recognized by U1 snRNP (small nuclear ribonucleoprotein) and U2 snRNP respectively] or by helping to overcome the effect of local splicing silencers. RS domains are often involved in protein–protein interactions. Protein–protein interactions have been identified between Tra2 proteins and the RS domain-containing U2AF35 protein, which is a component of the spliceosome that recruits U2 snRNP to the branchpoint [34]. This suggests the model that Tra2 proteins might activate splicing of weak 3′ splice sites by stabilizing binding of U2AF35 and thereby U2 snRNP to the branchpoint (Figure 2). Consistent with this model, Tra2-binding sites taken from the fruitfly *doublesex* gene can activate the spliceosome to use a β-globin 3′ splice site that is weakened by mutation [35].

Splicing activation of 5′ splice sites is likely to involve other interactions. No direct protein interactions have been detected between Tra2 proteins and the RS-domain containing U170K protein at the 5′ splice site [34]. However, Tra2 proteins interact with members of the core SR protein family including SRSF1 [34]. This suggests the model that Tra2 proteins might indirectly activate 5′ splice sites by stabilizing the associations of SR proteins with exons, with these SR proteins in turn stabilizing the association of U1 snRNP with RNA (shown in Figure 2).

Tra2 proteins activate cassette exon inclusion

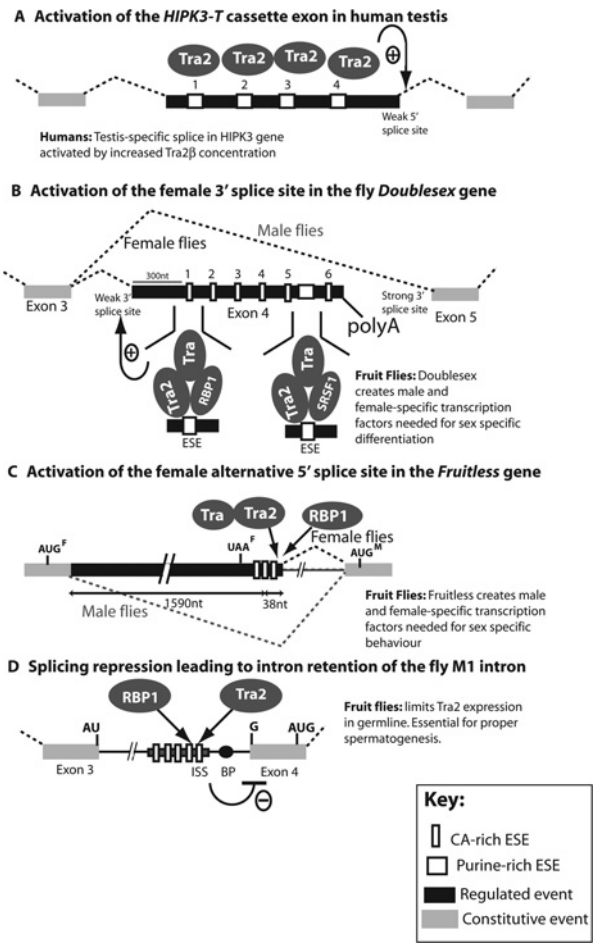
Cassette exons are the most frequent form of alternative splicing in human and mouse cells. A key feature of several Tra2 protein-regulated exons are multiple individual Tra2-binding sites within the target RNA. A requirement for multiple binding sites could fit into models of Tra2 protein function where either (a) multiple Tra2 activator proteins are needed to stabilize spliceosome assembly (this might be if the associated splice sites are very weak, or if there are a lot of associated silencer sequences repressing the exon); or (b) multiple sites available for binding in a regulated exon might increase the probability that at least one of these is occupied at any one time, and so be available to stabilize assembling spliceosomes [17].

The human testis-specific *HIPK3-T* exon has a weak 5′ splice site. *HIPK3-T* contains four ESE sequences activated by Tra2β, and together these ESEs act as a responsive gauge to enable activation of the weak 5′ splice site in response to increased concentrations of Tra2β (Figure 3A) [17]. In the case of *HIPK3-T*, removal of a single Tra2β-responsive ESE from the exon totally blocks splicing activation by Tra2β. This suggests that the four Tra2β-responsive ESEs are likely to provide a threshold of enhancer activity needed to counteract the weak 5′ splice of the *HIPK3-T* exon [17]. *In vitro* gel shift experiments show that Tra2β proteins assemble large complexes on the *HIPKT-T* exon. Another Tra2β-responsive exon that requires multiple Tra2β-responsive ESEs for splicing inclusion is found in the mouse *NASP* gene, where the *NASP-T* exon contains 37 potential Tra2β-binding sites [7]. Experimentally, individual Tra2β-binding sites can be mutated in *NASP-T* without significantly affecting Tra2-mediated splicing activation, but mutation of more than one binding site severely inhibited splicing activation. The *NASP-T* exon also assembles large RNA–protein complexes with Tra2β *in vitro*. Although NASP-T is a direct RNA target of Tra2β protein, a version of Tra2β protein lacking the RRM was able to act as a co-activator in minigene transfection experiments. This is also consistent with multiple Tra2β proteins assembling on the *NASP-T* exon, presumably by complexing with endogenous Tra2 proteins via its RS1 and RS2 domains [7].

Tra2β protein forms stable splicing regulator complexes with other proteins which require all protein components to be present, including hnRNP (heterogeneous nuclear ribonucleoprotein) G and SRp30c [36,37]. A single Tra2β-binding site is found within exon 7 of the *SMN2* gene. Mutation of this exonic Tra2β-binding site, and mutation of flanking candidate-binding sites for SRp30c and hnRNP G negatively affect splicing, consistent with each of these

Figure 3 | Model exons regulated by Tra2

(**A**) Splicing activation of the human *HIPK3T* exon is activated by 4 Tra2-responsive ESEs. (**B**) Activation of the female-specific 3′ splice site in the fruitfly *doublesex* mRNA is activated by a panel of Tra2-responsive ESEs. (**C**) Activation of the female specific 5′ splice site in the fruitfly *fruitless* mRNA is activated by an ESE containing a number of Tra2-binding sites. (**D**) Splicing retention of the M1 intron in the by *Tra2* mRNA is induced by Tra2 protein binding to an intronic silencing sequence.



proteins working together as a complex to regulate this exon [30].

Tra2 proteins activate weak 3′ splice sites

One of the most intensively studied splicing targets of Tra2 proteins is the *doublesex* gene in fruitflies. In fact, Tra2 proteins were first discovered since they play a key role in sex determination in fruitflies, and they are involved in sex determination in other insect species as well [38–40]. In fruitflies alternative splicing choices controlled by Tra2 create different male and female mRNA isoforms of *doublesex* [40]. The male and female *doublesex* mRNA isoforms are translated into distinct male and female transcription factors, which activate different promoters to establish male and female sexual differentiation. *Doublesex* exon 4 has a weak 3′ splice site, and is spliced in females only. *Doublesex* exon 4 is associated with the use of a downstream polyadenylation site, so that female *Doublesex* mRNA contains only exons 1–4. In male fruitflies a default pathway is followed in which *doublesex* exon 4 is skipped. Male *doublesex* mRNA thus contains exons 1–3 spliced to exons 5–6 (Figure 3B).

Tra2 proteins are essential for selection of the weak alternative 3′ splice site of *doublesex* exon 4 female fruitflies. *In vitro* experiments show Tra2 protein stimulates the very earliest stages of spliceosome assembly to activate splice site selection [41].Tra2 proteins bind to *doublesex* exon 4 through seven ESEs. Six of these ESEs are CA-rich Tra2-binding site sequences, and the 7th is a AGAA-rich Tra2-binding site sequence. Cross-linking experiments have shown that the six CA-rich repeat sequences each individually bind a trimeric complex containing Tra2 protein, the SR protein RBP1 and Tra protein. The purine-rich element binds a trimeric complex of Tra2, Tra and another SR called protein SRF1 [42,43]. Each of the three proteins in these trimeric complexes is needed to stably bind to RNA: if just one of the protein components is missing then no stable binding takes place. This is the case in male fruitflies, where the female-specific Tra protein is missing and so splicing follows the default pathway (the weak 3′ splice site of *doublesex* exon 4 is not selected in males).

The Tra2-dependent ESEs are spread over most of *doublesex* exon 4, except within the 300 nucleotides proximal to the weak 3′ splice site. This 300 nucleotide interval distance seems to be important for the ESEs to work properly: a 300 nucleotide gap sequence between the ESEs and the weak 3′ splice site is conserved across different species of fruitflies, and engineered copies of *doublesex* exon 4 that have CA repeat sequences inserted into this gap region lead to Tra2-independent 3′ splice site activation [44]. Tra2 and the other proteins that assemble on the *doublesex* ESEs are able to efficiently activate splicing of the female-specific 3′ splice site over the distance of the 300-nt gap sequence, whereas other RS domain-containing proteins are unable to activate the female-specific splice site [19].

How does Tra2 protein activate splicing of this weak female 3′ splice site in *doublesex*? An obvious mechanism would be through stabilization of U2AF binding at the weak 3′ splice site of *doublesex* exon 4, but deletion of the RS domain of U2AF35 did not affect sex determination in fruitflies [45]. Hence the actual mechanism of splicing activation is still not understood, but might involve a ‘zone of activation’ being created by multiple ESEs bound by Tra2–protein complexes [19]. Experiments in which the number of CA-rich repeat sequences were progressively increased from zero to seven in a *doublesex* minigene resulted in linear rather than synergistic increases in the use of the weak 3′ splice site [46]. This led to the conclusion that each individual copy of the 13 nucleotide CA-rich repeats probably interacts separately with spliceosome complexes at the 3′ splice site. According to this model, in the wild-type *doublesex* exon the presence of seven independent Tra2-binding sites would help ensure that at least one is occupied at any one time.

Tra2 proteins can activate alternative 5′ splice sites

One of the best understood examples of alternative 5′ splice site selection by Tra2 proteins also comes from the fruitfly sex determination pathway, where Tra2 regulates alternative splicing of the *fruitless* mRNA (Figure 3C). The *fruitless* gene is essential for male and female development, and encodes different transcription factor isoforms. In males, the fruitless transcription factor controls the development of a male-specific muscle called the muscle of Lawrence, and is essential for normal sexual behaviour (http://flybase.org/reports/FBgn0004652.html).

Fruitless pre-mRNA is alternatively spliced to produce both male and female splice mRNA isoforms [47]. The Tra2-responsive ESE in *fruitless* is close (within 38 nucleotides) to the weak female 5′ splice site [47] and contains three copies of the CA-rich Tra2-binding site, This ESE locally drives selection of this weaker site when it is bound by Tra and Tra2. In male fruitflies, a stronger 5′ splice site is used instead that is 1590 nt upstream. This means when the male 5′ splice site is used, a much shorter version of the exon is spliced into the *fruitless* mRNA [33,48]. This shorter version encodes a male version of the Fruitless transcription factor from an ORF initiating at a different male-specific AUG.

Tra2 proteins can function as splicing repressors: Tra2 proteins autoregulate their own expression levels through promoting intron retention in fruitflies

Tra2 proteins can also act as splicing repressors. This function has been particularly well characterized in the fruitfly Tra2 protein, where increasing concentrations of Tra2 protein lead to retention of an intron called the M1 intron in the *Tra2* mRNA itself. This M1 intron splits the initiation codon at the start of the ORF that encodes Tra2 protein (Figure 3D). Increases in Tra2 protein concentration blocks splicing of this M1 intron. When the M1 intron is retained in the *Tra2* mRNA, a downstream AUG initiation codon is used for Tra2, encoding a shorter Tra2 protein that lacks the N-terminal RS domain. Retention of the M1 intron in Tra2 mRNAs is tissue-specific, and detected only in the fruitfly testis. More than half of Tra2 mRNAs expressed in the *Drosophila* testis retain intron M1.

To induce splicing repression Tra2 protein binds to intron sequences instead of binding within exons. In the M1 intron, Tra2 protein binds to an ISS (intron splicing silencer) sequence containing five individual copies of the CA-rich Tra2-binding site to cause M1 intron retention [49,50]. This ISS is conserved in different fruitfly species and normally located near the M1 intron branchpoint. This suggests the model that binding of Tra2 protein causes physical obstruction of the branchpoint region from the spliceosome (particularly stearic hindrance of U2 snRNP) and so leads to intron retention. However, this appealing model does not seem to be the case in reality, since the ISS can function just as well if it is moved away from the branchpoint sequence [50]. Binding of Tra2 protein in the M1 intron blocks splicing at the very earliest stages of spliceosome assembly [51]. Tra2 protein concentration is particularly high in fruitfly spermatocytes (these are cells in the testis which are undergoing meiosis), and the ISS sequence might be designed to respond to these higher concentrations of Tra2 protein that only occur in this cell type.

The choice between Tra2 protein functioning as an activator and as a silencer of splicing is dependent on exon versus intron binding. The inhibitory effect of the M1 ISS on splicing depends on an intronic position. This same M1 ISS sequence can function as an ESE when located within an exon. Splicing repression however does not depend on the RS domains of Tra2 protein [51]. Just as Tra2-interacting proteins are required for splicing activation by Tra2 proteins, interacting proteins are also important for splicing repression by Tra2. An RNA-binding protein called half-pint binds to *Tra2* pre-mRNA, and interacts with Tra2 protein to autoregulate the Tra2 expression levels in the testis through M1 intron retention [52]. Although required for M1 intron retention, the half-pint protein is not needed for regulation of *Doublesex* exon 4 splicing, indicating that Tra2 protein interacts with specific protein partners to direct splicing patterns of different target RNAs. Another RNA-binding protein regulating M1 intron retention is RBP1, but unlike for *doublesex* regulation Tra2 and RBP1 bind separately on the M1 intron RNA rather than as a complex.

Why is autoregulation control of Tra2 protein levels important in the fruitfly testis? Tra2 is normally expressed at higher levels in primary spermatocytes than in the rest of the fruitfly body [53]. However, fruitflies engineered to overexpress Tra2 proteins have defective spermatogenesis. The M1 intron ISS might have evolved as a Tra2-responsive switch that is only activated to cause intron retention at the higher levels of Tra2 protein present in primary spermatocytes, preventing expression reaching harmful levels. Transiently increasing Tra2 protein expression in other adult tissues using a heat-shock-activated promoter leads to M1 intron retention, and forced ubiquitous expression using a strong promoter causes embryonic lethality [53]. Hence proper regulation of Tra2 protein levels are important throughout the fruitfly body. Although overexpression of Tra2 protein prevents normal spermatogenesis, normal expression of Tra2 protein is important in the testis and removal of Tra2 also interferes with spermatogenesis. Fruitflies lacking Tra2 protein are infertile and produce sperm with abnormal-shaped heads [2]. Alternative splice forms of three other mRNAs are known to be regulated by Tra2 in the fruitfly testis. These are *exuperantia* (encoding a protein involved in RNA localization and needed for spermatogenesis) [54,55]; *att* (encoding a protein possibly involved in transmembrane transport) [56]; and TAF1 (encoding a transcription factor) [10,57]. Absence of these Tra2-regulated splice forms, particularly *exuperantia*, might lead to the defects in spermatogenesis observed in fruitflies without Tra2 protein.

Tra2 protein levels are regulated through alternative splicing in mammals as well as fruitflies

The mouse and human genes encoding Tra2α and Tra2β proteins also control their expression levels, in this case through splicing control of a poison exon that responds to Tra2 protein concentration in the cell. When Tra2α or Tra2β concentrations reach a certain threshold, poison exons are spliced into the *Tra2a* and *Tra2b* mRNAs to introduce stop codons and block further translation of these mRNAs [7,58]. As in many other examples of Tra2 protein regulated splicing, multiple Tra2 protein–RNA binding sites are found within these regulated poison exons, and these binding sites are highly conserved in evolution. Together the data from mice, humans and fruitflies suggest that there is an important requirement in all animals to tightly control Tra2 protein expression levels.

Conclusions

Tra2 proteins play important roles in animal development and physiology. Current research is uncovering their targets and mechanisms of action.

Funding

This work was supported by the Wellcome Trust [grant number WT089225/Z/09/Z (to D.J.E.)], the Biotechnology and Biological Sciences Research Council [grant numbers BB/I006923/1 and BB/K018957/1 (to D.J.E.)] and the Breast Cancer Campaign (to D.J.E.).

References

1 Baker, B.S. (1989) Sex in flies: the splice of life. Nature **340**, 521–524 CrossRef PubMed
2 Belote, J.M. and Baker, B.S. (1983) The dual functions of a sex determination gene in *Drosophila melanogaster*. Dev. Biol. **95**, 512–517 CrossRef PubMed
3 Dauwalder, B., Amaya-Manzanares, F. and Mattox, W. (1996) A human homologue of the *Drosophila* sex determination factor transformer-2 has conserved splicing regulatory functions. Proc. Natl. Acad. Sci. U.S.A. **93**, 9004–9009 CrossRef PubMed
4 Mende, Y., Jakubik, M., Riessland, M., Schoenen, F., Rossbach, K., Kleinridders, A., Köhler, 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**, 2154–2167 CrossRef PubMed
5 Storbeck, M., Hupperich, K., Gaspar, J.A., Meganathan, K., Martinez Carrera, L., Wirth, R., Sachinidis, A. and Wirth, B. (2014) Neuronal-specific deficiency of the splicing factor tra2b causes apoptosis in neurogenic areas of the developing mouse brain. PLoS ONE **9**, e89020 CrossRef PubMed
6 Roberts, J.M., Ennajdaoui, H., Edmondson, C., Wirth, B., Sanford, J.R. and Chen, B. (2014) Splicing factor TRA2B is required for neural progenitor survival. J. Comp. Neurol. **522**, 372–392 CrossRef PubMed
7 Grellscheid, S., Dalgliesh, C., Storbeck, M., Best, A., Liu, Y., Jakubik, M., Mende, Y., Ehrmann, I., Curk, T., Rossbach, K. et al. (2011) Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2β in development. PLoS Genet. **7**, e1002390 CrossRef PubMed
8 Fu, K., Mende, Y., Bhetwal, B.P., Baker, S., Perrino, B.A., Wirth, B. and Fisher, S.A. (2012) Tra2β protein is required for tissue-specific splicing of a smooth muscle myosin phosphatase targeting subunit alternative exon. J. Biol. Chem. **287**, 16575–16585 CrossRef PubMed
9 Anderson, E.S., Lin, C.H., Xiao, X., Stoilov, P., Burge, C.B. and Black, D.L. (2012) The cardiotonic steroid digitoxin regulates alternative splicing through depletion of the splicing factors SRSF3 and TRA2B. RNA **18**, 1041–1049 CrossRef PubMed
10 Katzenberger, R.J., Marengo, M.S. and Wassarman, D.A. (2009) Control of alternative splicing by signal-dependent degradation of splicing-regulatory proteins. J. Biol. Chem. **284**, 10737–10746 CrossRef PubMed
11 Karunakaran, D.K., Banday, A.R., Wu, Q. and Kanadia, R. (2013) Expression analysis of an evolutionarily conserved alternative splicing factor, Sfrs10, in age-related macular degeneration. PLoS ONE **8**, e75964 CrossRef PubMed
12 Holly, A.C., Melzer, D., Pilling, L.C., Fellows, A.C., Tanaka, T., Furrucci, L. and Harries, L.W. (2013) Changes in splicing factor expression are associated with advancing age in man. Mech. Ageing Dev. **134**, 356–366 CrossRef PubMed
13 Best, A., Dagliesh, C., Ehrmann, I., Kheirollahi-Kouhestani, M., Tyson-Capper, A. and Elliott, D.J. (2013) Expression of Tra2β in cancer cells as a potential contributory factor to neoplasia and metastasis. Int. J. Cell Biol. **2013**, 843781 CrossRef PubMed
14 Hofmann, Y., Lorson, C.L., Stamm, S., Androphy, E.J. and Wirth, B. (2000) Htra2-β1 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**, 9618–9623 CrossRef PubMed
15 Glatz, D.C., Rujescu, D., Tang, Y., Berendt, F.J., Hartmann, A.M., Faltraco, F., Rosenberg, C., Hulette, C., Jellinger, K., Hampel, H. et al. (2006) The alternative splicing of tau exon 10 and its regulatory proteins CLK2 and TRA2-β1 changes in sporadic Alzheimer's disease. J. Neurochem. **96**, 635–644 CrossRef PubMed
16 Jiang, Z., Tang, H., Havlioglu, N., Zhang, X., Stamm, S., Yan, R. and Wu, J.Y. (2003) Mutations in tau gene exon 10 associated with FTDP-17 alter the activity of an exonic splicing enhancer to interact with Tra2β. J. Biol. Chem. **278**, 18997–19007 CrossRef PubMed
17 Grellscheid, S.N., Dalgliesh, C., Rozanska, A., Grellscheid, D., Bourgeois, C.F., Stévenin, D.J. and Elliott, D.J. (2011) Molecular design of a splicing switch responsive to the RNA binding protein Tra2β. Nucleic Acids Res. **39**, 8092–8104 CrossRef PubMed
18 Amrein, H., Hedley, M.L. and Maniatis, T. (1994) The role of specific protein–RNA and protein–protein interactions in positive and negative control of pre-mRNA splicing by transformer 2. Cell **76**, 735–746 CrossRef PubMed
19 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**, 6612–6620 CrossRef PubMed
20 Li, S.J., Qi, Y., Zhao, J.J., Li, Y., Liu, X.Y., Chen, X.H. and Xu, P. (2013) Characterization of nuclear localization signals (NLSs) and function of NLSs and phosphorylation of serine residues in subcellular and subnuclear localization of transformer-2β (Tra2β). J. Biol. Chem. **288**, 8898–8909 CrossRef PubMed
21 Zhou, Z. and Fu, X.D. (2013) Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma **122**, 191–207 CrossRef PubMed
22 Shepard, P.J. and Hertel, K.J. (2009) The SR protein family. Genome Biol. **10**, 242 CrossRef PubMed
23 Long, J.C. and Caceres, J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. Biochem. J. **417**, 15–27 CrossRef PubMed
24 Aubol, B.E., Jamros, M.A., McGlone, M.L. and Adams, J.A. (2013) Splicing kinase SRPK1 conforms to the landscape of its SR protein substrate. Biochemistry **52**, 7595–7605 CrossRef PubMed
25 Rabinow, L. and Samson, M.L. (2010) The role of the *Drosophila* LAMMER protein kinase DOA in somatic sex determination. J. Genet. **89**, 271–277 CrossRef PubMed
26 Du, C., McGuffin, M.E., Dauwalder, B., Rabinow, L. and Mattox, W. (1998) Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in *Drosophila*. Mol. Cell **2**, 741–750 CrossRef PubMed
27 Tacke, R., Tohyama, M., Ogawa, S. and Manley, J.L. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. Cell **93**, 139–148 CrossRef PubMed
28 Inoue, K., Hoshijima, K., Higuchi, I., Sakamoto, H. and Shimura, Y. (1992) Binding of the *Drosophila* transformer and transformer-2 proteins to the regulatory elements of *doublesex* primary transcript for sex-specific RNA processing. Proc. Natl. Acad. Sci. U.S.A. **89**, 8092–8096 CrossRef PubMed

29 Tran, Q., Coleman, T.P. and Roesser, J.R. (2003) Human transformer 2β and SRp55 interact with a calcitonin-specific splice enhancer. Biochim. Biophys. Acta **1625**, 141–152 CrossRef PubMed

30 Clery, A., Jayne, S., Benderska, N., Dominguez, C., Stamm, S. and Allain, F.H. (2011) Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1. Nat. Struct. Mol. Biol. **18**, 443–450 CrossRef PubMed

31 Tsuda, K., Someya, T., Kuwasako, K., Takahashi, M., He, F., Unzai, S., Inoue, M., Harada, T., Watanabe, S., Terada, T. et al. (2011) Structural basis for the dual RNA-recognition modes of human Tra2-β RRM. Nucleic Acids Res. **39**, 1538–1553 CrossRef PubMed

32 Goodier, J.L., Cheung, L.E. and Kazazian, Jr, H.H. (2013) Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. Nucleic Acids Res. **41**, 7401–7419 CrossRef PubMed

33 Lam, B.J., Bakshi, A., Ekinci, F.Y., Webb, J., Graveley, B.R. and Hertel, K.J. (2003) Enhancer-dependent 5′-splice site control of fruitless pre-mRNA splicing. J. Biol. Chem. **278**, 22740–22747 CrossRef PubMed

34 Wu, J.Y. and Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell **75**, 1061–1070 CrossRef PubMed

35 Tian, M. and Maniatis, T. (1992) Positive control of pre-mRNA splicing *in vitro*. Science **256**, 237–240 CrossRef PubMed

36 Hofmann, Y. and Wirth, B. (2002) hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-β1. Hum. Mol. Genet. **11**, 2037–2049 CrossRef PubMed

37 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 Tra2β and affect splicing. Hum. Mol. Genet. **9**, 685–694 CrossRef PubMed

38 Nissen, I., Muller, M. and Beye, M. (2012) The *Am-tra2* gene is an essential regulator of female splice regulation at two levels of the sex determination hierarchy of the honeybee. Genetics **192**, 1015–1026 CrossRef PubMed

39 Schurko, A.M. (2013) To 'bee or not to bee' male or female? An educational primer for use with 'The *Am-tra2* gene is an essential regulator of female splice regulation at two levels of the sex determination hierarchy of the honeybee'. Genetics **193**, 1019–1023 CrossRef PubMed

40 Venables, J.P., Tazi, J. and Juge, F. (2012) Regulated functional alternative splicing in *Drosophila*. Nucleic Acids Res. **40**, 1–10 CrossRef PubMed

41 Tian, M. and Maniatis, T. (1993) A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. Cell **74**, 105–114 CrossRef PubMed

42 Lynch, K.W. and Maniatis, T. (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila doublesex* splicing enhancer. Genes Dev. **10**, 2089–2101 CrossRef PubMed

43 Lynch, K.W. and Maniatis, T. (1995) Synergistic interactions between two distinct elements of a regulated splicing enhancer. Genes Dev. **9**, 284–293 CrossRef PubMed

44 Hertel, K.J., Lynch, K.W., Hsiao, E.C., Liu, E.H. and Maniatis, T. (1996) Structural and functional conservation of the *Drosophila doublesex* splicing enhancer repeat elements. RNA **2**, 969–981 PubMed

45 Rudner, D.Z., Kanaar, R., Breger, K.S. and Rio, D.C. (1998) Interaction between subunits of heterodimeric splicing factor U2AF is essential *in vivo*. Mol. Cell Biol. **18**, 1765–1773 PubMed

46 Hertel, K.J. and Maniatis, T. (1998) The function of multisite splicing enhancers. Mol. Cell **1**, 449–455 CrossRef PubMed

47 Ryner, L.C., Goodwin, S.F., Castrillon, D.H., Anand, A., Villella, A., Baker, B.S., Hall, J.C., Taylor, B.J. and Wasserman, S.A. (1996) Control of male sexual behavior and sexual orientation in *Drosophila* by the *fruitless* gene. Cell **87**, 1079–1089 CrossRef PubMed

48 Heinrichs, V., Ryner, L.C. and Baker, B.S. (1998) Regulation of sex-specific selection of fruitless 5′ splice sites by transformer and transformer-2. Mol. Cell. Biol. **18**, 450–458 PubMed

49 Chandler, D.S., Qi, J. and Mattox, W. (2003) Direct repression of splicing by transformer-2. Mol. Cell. Biol. **23**, 5174–5185 CrossRef PubMed

50 Qi, J., Su, S. and Mattox, W. (2007) The *doublesex* splicing enhancer components Tra2 and Rbp1 also repress splicing through an intronic silencer. Mol. Cell. Biol. **27**, 699–708 CrossRef PubMed

51 Shen, M. and Mattox, W. (2012) Activation and repression functions of an SR splicing regulator depend on exonic versus intronic-binding position. Nucleic Acids Res. **40**, 428–437 CrossRef PubMed

52 Wang, S., Wagner, E.J. and Mattox, W. (2013) Half pint/Puf68 is required for negative regulation of splicing by the SR splicing factor transformer2. RNA Biol. **10**, 1396–1406 CrossRef PubMed

53 Qi, J., Su, S., McGuffin, M.E. and Mattox, W. (2006) Concentration dependent selection of targets by an SR splicing regulator results in tissue-specific RNA processing. Nucleic Acids Res. **34**, 6256–6263 CrossRef PubMed

54 Hazelrigg, T. and Tu, C. (1994) Sex-specific processing of the *Drosophila exuperantia* transcript is regulated in male germ cells by the *tra-2* gene. Proc. Natl. Acad. Sci. U.S.A **91**, 10752–10756 CrossRef PubMed

55 Hazelrigg, T., Watkins, W.S., Marcey, D., Tu, C., Karow, M. and Lin, X.R. (1990) The *exuperantia* gene is required for *Drosophila* spermatogenesis as well as anteroposterior polarity of the developing oocyte, and encodes overlapping sex-specific transcripts. Genetics **126**, 607–617 PubMed

56 Madigan, S.J., Edeen, P., Esnayra, J. and McKeown, M. (1996) *att*, a target for regulation by tra2 in the testes of *Drosophila melanogaster*, encodes alternative RNAs and alternative proteins. Mol. Cell. Biol. **16**, 4222–4230 PubMed

57 Katzenberger, R.J., Marengo, M.S. and Wassarman, D.A. (2006) ATM and ATR pathways signal alternative splicing of *Drosophila TAF1* pre-mRNA in response to DNA damage. Mol. Cell. Biol. **26**, 9256–9267 CrossRef PubMed

58 Stoilov, P., Daoud, R., Nayler, O. and Stamm, S. (2004) Human tra2-β1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. Hum. Mol. Genet. **13**, 509–524 CrossRef PubMed

59 Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K. and Lührmann, B. (2012) Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nat. Commun. **3**, 994 CrossRef PubMed

60 Robberson, B.L., Cote, G.J. and Berget, S.M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. **10**, 84–94 PubMed

61 Black, D.L. (1995) Finding splice sites within a wilderness of RNA. RNA **1**, 763–771 PubMed

Received 1 April 2014
doi:10.1042/BST20140075

Hindawi Publishing Corporation
International Journal of Cell Biology
Volume 2013, Article ID 843781, 9 pages
http://dx.doi.org/10.1155/2013/843781

Review Article

Expression of Tra2β in Cancer Cells as a Potential Contributory Factor to Neoplasia and Metastasis

Andrew Best,[1] Caroline Dagliesh,[1] Ingrid Ehrmann,[1] Mahsa Kheirollahi-Kouhestani,[1] Alison Tyson-Capper,[2] and David J. Elliott[1]

[1] *Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK*
[2] *Institute of Cellular Medicine, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK*

Correspondence should be addressed to David J. Elliott; david.elliott@ncl.ac.uk

Received 2 May 2013; Accepted 9 June 2013

Academic Editor: Claudia Ghigna

The splicing regulator proteins SRSF1 (also known as ASF/SF2) and SRSF3 (also known as SRP20) belong to the SR family of proteins and can be upregulated in cancer. The *SRSF1* gene itself is amplified in some cancer cells, and cancer-associated changes in the expression of *MYC* also increase *SRSF1* gene expression. Increased concentrations of SRSF1 protein promote prooncogenic splicing patterns of a number of key regulators of cell growth. Here, we review the evidence that upregulation of the SR-related Tra2β protein might have a similar role in cancer cells. The *TRA2B* gene encoding Tra2β is amplified in particular tumours including those of the lung, ovary, cervix, stomach, head, and neck. Both *TRA2B* RNA and Tra2β protein levels are upregulated in breast, cervical, ovarian, and colon cancer, and Tra2β expression is associated with cancer cell survival. The *TRA2B* gene is a transcriptional target of the protooncogene ETS-1 which might cause higher levels of expression in some cancer cells which express this transcription factor. Known Tra2β splicing targets have important roles in cancer cells, where they affect metastasis, proliferation, and cell survival. Tra2β protein is also known to interact directly with the RBMY protein which is implicated in liver cancer.

1. Introduction

Cancer is associated with a number of distinctive disease hallmarks [1]. These hallmarks include the ability of cancer cells to continuously divide by maintaining proliferative signalling pathways and to evade growth suppressors, to resist cell death; to induce angiogenesis to ensure a supply of oxygen and nutrition, and to invade other parts of the body (metastasis). These hallmarks of cancer cells occur against other changes including decreasing genome stability and inflammation [1].

Changes in splicing patterns in cancer cells compared to normal cells can contribute to each of these cancer hallmarks through effects on the expression patterns of important protein isoforms which regulate cell behaviour [2–4]. The splicing alterations which occur in cancer cells are partially due to changes in the activity and expression of core spliceosome components [5] and in the RNA binding proteins which regulate alternative exon inclusion [6]. Changes in the splicing environment in cancer cells might have therapeutic implications. Drugs which target the spliceosome are also being developed as potential therapies for treating cancer patients [7].

In this review, we particularly examine the potential role of the splicing regulator Tra2β as a modulator of gene function in cancer cells. Tra2β is part of a larger protein family which contains RNA recognition motifs (RRMs) and extended regions of serine and arginine residues (RS domains, named following the standard 1 letter amino acid code for serine and arginine) [8–10]. Core SR proteins include SRSF1 (previously known as ASF/SF2) and SRSF3 (previously known as SRP20) (Figure 1). Tra2β is considered an SR-like protein rather than a core SR family member because of two features. Firstly, Tra2β contains both an N- and C-terminal RS domains (each of the core members of the SR family has just a single C-terminal RS domain, with the RRM at the N-terminus). Secondly, the core group of SR proteins but not Tra2β can restore splicing activity to S100 extracts [11]. S100 extracts are made from lysed HeLa cells by high-speed

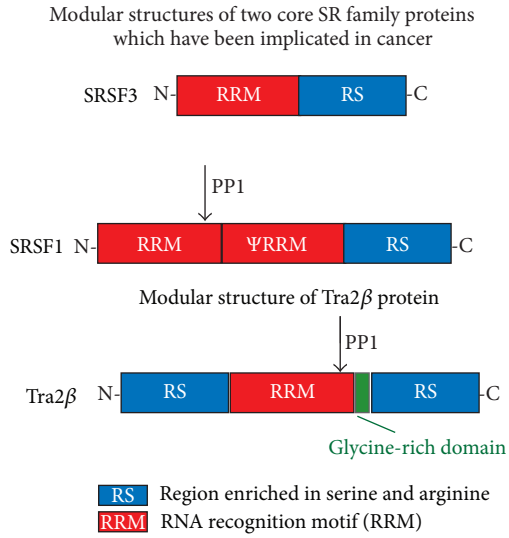


Figure 1: Modular structure of the core SR family proteins SRSF1 (also known as ASF/SF2) and SRSF3 (also known as SRp20) and the SR-like protein Tra2β. The RNA recognition motif (RRM) binds to target RNAs, and the RS region is responsible for protein-protein interactions. SRSF1 has a second RRM, annotated ψRRM. SRSF1 and Tra2β have a PP1 docking site.

ultracentrifugation to remove nuclei but contain most of the core spliceosome components necessary for splicing with the important exception of SR proteins which are insoluble in the magnesium concentrations used [12]. Addition of any single SR protein is sufficient to restore splicing activity to these S100 extracts [13].

Tra2β protein functions as a splicing regulator in the cell nucleus, where it activates the inclusion of alternative exons [14, 15]. Tra2β protein is able to interact with two types of RNA targets through its RRM. Firstly, the major RNA binding site for Tra2β is an AGAA-rich sequence [11, 16, 17]. Although an AGAA RNA sequence works best for Tra2β protein, an NGAA sequence is actually sufficient for binding. However, substituting the first A with either C, G, or T nucleotides in the NGAA target sequence decreases binding efficiency (the Kd value increases 2-fold between AGAA and NGAA) [16]. Secondly, the RRM of Tra2β is able to switch to a second mode of RNA binding, in which it interacts with single-stranded CAA-rich sequences within a stem loop structure [17].

When Tra2β binds to target RNA sites within an exon, it activates splicing inclusion of these bound exons into mRNA [11, 15–17]. Splicing activation by Tra2β protein is concentration dependent: increased Tra2β protein concentration leads to increased levels of target exon splicing inclusion [14, 15]. The RRMs of Tra2β and SRSF1 proteins both contain a docking site for protein phosphatase 1 (PP1), and dephosphorylation of these proteins by PP1 affects alternative splicing regulation [18].

Tra2β protein is encoded by the *TRA2B* gene (also called *SFRS10*) on human chromosome 3. As well as any potential role in cancer cells, Tra2β has important roles in normal development and is essential for normal mouse embryonic and brain development (*TRA2B* knockout mice fail to develop normally) [15, 19]. *TRA2B* has a paralog gene called *TRA2A* on the long arm of human chromosome 7, and this paralog encodes Tra2α protein [20]. Paralogs are additional copies of a gene derived by duplication. *TRA2A* derived by gene duplication from *TRA2B* early in the vertebrate lineage and so is found in all vertebrates.

A number of the SR proteins have been found to have roles in cancer, amongst them, SRSF1 and SRSF3 (Figures 1 and 2). The mechanism of SRSF1 upregulation in cancer cells has been explained at a mechanistic level, and the effects of this upregulation in terms of gene expression control have been mapped onto the pathway of oncogenesis. Here, we review these important principles for SRSF1 and then apply these principles to gauge the likely effect of the Tra2β protein on cancer-specific gene expression.

2. SRSF1 Is Upregulated in Cancer and Is a Target for the Prooncogenic Transcription Factor Myc

SRSF1 upregulation in cancer cells can occur through two distinct mechanisms. Firstly, the *SRSF1* gene itself can become amplified in cancer. The *SRSF1* gene is on a region of chromosome 17q23 which is amplified in some breast cancers, including in tumours with a poor prognostic outlook and in the MCF7 breast cancer cell line [21]. Analysis of the *SRSF1* gene on the cBio Cancer Genomics Portal shows amplification of *SRSF1* mainly in breast cancers (Figure 2) [22, 23]. Secondly, *SRSF1* gene transcription is activated by the prooncogenic transcription factor Myc which is itself activated in some cancers. Myc upregulation in cancer leads to downstream increases in both *SRSF1* mRNA and SRSF1 protein expression [24].

Protein expression analysis using a highly specific monoclonal antibody showed that a number of tumours have increased SRSF1 protein compared to normal tissue [21]. As well as being upregulated in some cancer cells, *SRSF1* operates as a bona fide oncogene. Increased *SRSF1* gene expression can transform rodent fibroblasts in an NIH3T3 assay, and the resulting transformed cells form tumours in nude mice [21]. Tumour formation by these transformed fibroblasts is directly dependent on *SRSF1* expression, since it is blocked by parallel shRNA inhibition of *SRSF1* [21]. Together, these data suggest that upregulation of *SRSF1* gene expression can be one of the initial steps in oncogenesis.

Experiments support an important function for SRSF1 protein in breast cancer cells. Mouse COMMA1-D mammary epithelial cells form tumours more efficiently in mice after transduction with *SRSF1*, and transduction of MF10A cells with *SRSF1* results in increased acinar size and decreased apoptosis in a 3D culture model [25]. A number of splicing targets have been identified which respond to increased levels of *SFRS1* expression in cancer cells (Table 1). These SRSF1-driven splicing changes produce prooncogenic mRNA splice isoforms, which encode proteins which decrease apoptosis and increase cellular survival and proliferation.

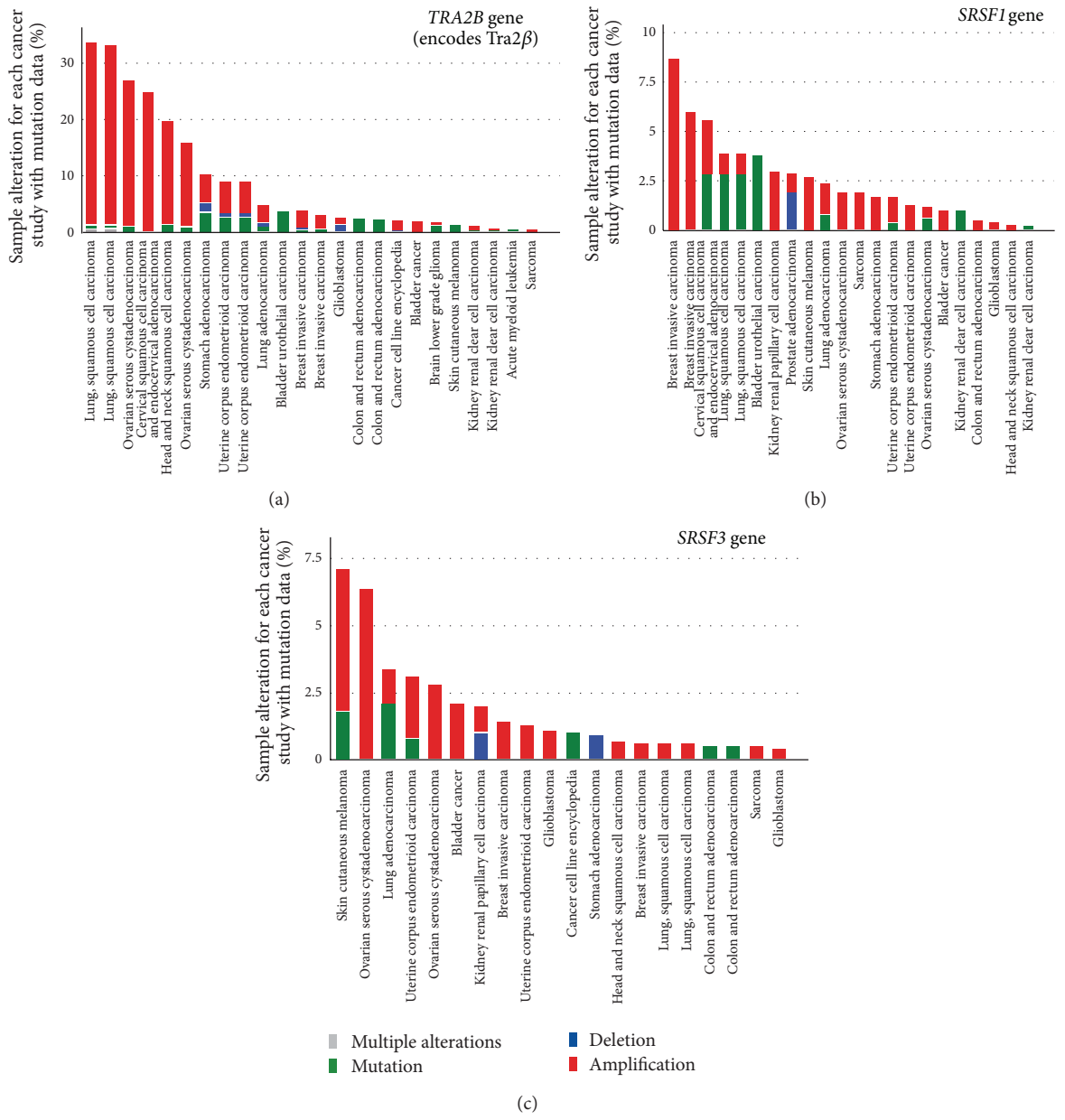


Figure 2: The (a) *TRA2B*, (b) *SRSF1* (also known as ASF/SF2), and (c) *SRSF3* (also known as SRP20) genes are amplified or otherwise mutated in several cancer types. For each of the three genes, data for genetic changes in all cancers were obtained using the cBioPortal database, filtering for percentage of altered cases (studies using mutation data) [22, 23]. The percentage, of cancer samples which showed genetic alterations in large cancer studies are shown on the *Y* axis and the respective type of cancer on the *X* axis. Full details of this kind of analysis are given on the cBioPortal website http://www.cbioportal.org/public-portal/index.do.

3. Increased *SRSF3* Expression Is Also Associated with Cancer

Increased expression of the SR protein SRSF3 is also associated with cancer. The *SRSF3* gene is amplified in some cancers (Figure 2) [22, 23]. Loss of *SRSF3* expression in a number of cancer cell lines increases apoptosis and decreases proliferation, and increased expression of *SRSF3* leads to transformation of rodent fibroblasts and enables them to form tumours in nude mice [26].

Increased *SRSF3* expression levels have been associated with an increased tumour grade in ovarian cancer [27]. Intracellular levels of *SRSF3* mRNA are important for cancer cells: siRNA-mediated downregulation of *SRSF3* leads to cell cycle arrest at G1 in colon cancer cells, and their increased death through apoptosis. The mechanism of increased apoptosis in response to higher levels of SRSF3 protein might include aberrant splicing of the *HIPK2* pre-mRNA (which encodes an important apoptotic regulator related to *HIPK3*, which is a known splicing target of Tra2β), such that

Table 1: Known prooncogenic splicing targets of SFRS1 (previously known as ASF/SF2).

Splicing target	Possible role in cancer cells	Reference
RON	Δexon11 splice isoform increases cell motility and metastasis	[21, 25]
BIN	*BIN12a* splice isoform encodes protein no longer able to bind Myc and acts as tumour suppressor	[21, 25]
MNK2	*MNK2* 13b splice isoform makes kinase which can phosphorylate EIF4E independent of MAP kinase activation	[21, 25]
S6K	Promotes oncogenic isoform of S6kinase which phosphorylates small subunit of ribosome	[21, 25]
MCL-1/BCL-X/CASPASE9	Promotes production of antiapoptotic mRNAs to result in cell survival	[63–65]

a proteasome-resistant form of HIPK2 protein is made after SRSF3 depletion [28].

4. Tra2β Is Amplified in Particular Cancers and Is a Target of the Oncogenic Transcription Factor ETS-1

The *TRA2B* gene which encodes Tra2β becomes amplified in several cancers (Figure 2) and particularly in cancers of the lung, cervix, head and neck, ovary, stomach, and uterus [22, 23]. Upregulation of Tra2β protein expression has also been observed in several cancers, including breast, cervical and ovarian [29–31], and colon [32]. Tra2β upregulation is associated with invasive breast cancer [30], and medium to high Tra2β expression correlates with a poorer prognosis in cervical cancer compared to patients with lower expression levels [29].

Tra2β protein expression has been demonstrated to be important for cancer cell biology. Downregulation of Tra2β inhibits cell growth of a gastric cancer cell line, measured by a corresponding decrease in BrdU incorporation which monitors cells which have entered S phase [33]. Knockdown of Tra2β in colon cancer cells reduced cell viability and increased the level of apoptosis monitored using a TUNEL assay and through measurement of levels of cleaved PARP [32].

As well as *TRA2B* gene amplification, the expression levels of the ETS-1 transcription factor provide a possible mechanism through which Tra2β might be upregulated in cancer cells. Regulated transcription of the *TRA2B* gene in human colon cells is positively controlled by binding of the HSF1 and ETS-1 transcription factors to its promoter proximal region [32]. The ETS-1 protein is itself encoded by a protooncogene. ETS1 expression in metastatic breast cancer correlates with a poor prognosis [34, 35] and is associated with an invasive phenotype [36]. Expression of both ETS-1 [35] and Tra2β [37] might also be under control of estrogen, which is a key driver of estrogen receptor positive breast cancer development. Taken together, these observations suggest that the pathological mechanism of Tra2β upregulation in cancer cells might result from underlying changes in transcription factors in cancer cells. Other positive regulators of cell growth might also stimulate Tra2β expression, since expression of Tra2β is upregulated in response to growth factors in normal smooth muscle cells [38].

Reactive oxygen species made during inflammation provide a further potential mechanism for Tra2β upregulation in cancer cells. Tra2β expression is activated in response to reoxygenation of astrocytes following a period of oxygen deprivation and by ischaemia in rat brains [39]. Expression of Tra2β in smooth muscle cells is similarly induced following reoxygenation of hypoxic cells [38], and is upregulated in response to oxidative stress in human colorectal carcinoma cell line HCT116 [32]. Ischaemia has also been reported to induce cytoplasmic accumulation of Tra2β along with accompanying changes in splice site use [40]. Tra2β translocates into the cytoplasm in gastric cancer cells in response to cell stress induced by sodium arsenate [32], and changes in the nuclear concentration of Tra2β might have downstream effects on the splicing inclusion of target exons.

The increased levels of Tra2β observed in cancer cells mean that the *TRA2B* gene must be able to bypass the normal feedback expression control mechanisms which exist to keep Tra2β protein levels under tight control. An important feedback control mechanism uses an alternatively spliced "poison exon" in the *TRA2B* gene. Poison exons introduce premature stop codons when they are spliced into mRNAs, preventing translation of full-length proteins and often targeting mRNAs for nonsense-mediated decay [41]. Poison exon splicing into the *TRA2B* mRNA is activated by binding of Tra2β itself. Splicing inclusion of this poison exon acts as a brake on production of more Tra2β protein. The predicted outcome is that increased expression of Tra2β protein should lead to increased *TRA2B* poison exon inclusion and so correspondingly less newly translated Tra2β protein through a negative feedback loop [42].

Similarly, the levels of SRSF1 and the other SR proteins are thought to be normally autoregulated through poison exon inclusion [43]; so these other SR proteins must similarly bypass these mechanisms in cancer cells to enable their higher levels of expression to be established.

5. Tra2β Protein Regulates Splicing Patterns Which Are Important to Cancer Cells

How might upregulation of Tra2β affect the biology of cancer cells? Three Tra2β-target exons have been identified in genes known to have important roles in cancer cells (Table 2). For two of these target exons, the actual regulated isoforms have also been demonstrated in cancer cells.

Firstly, strong Tra2β binding to a cancer-associated exon in the *nuclear autoantigenic sperm protein* (abbreviated *NASP*) gene has been detected using HITS-CLIP of endogenous Tra2β protein in the mouse testis [14, 15]. This Tra2β-target exon is abbreviated *NASP-T*. Whilst the somatic

Table 2: Known pro-oncogenic splicing targets of Tra2β.

Splicing target	Possible role in cancer cells	Reference
CD44	Affects cancer cell mobility and metastasis	[30]
Homeodomain-interacting kinase 3 (HipK3)	*HIPK3* increases phosphorylation of cJun and cell proliferation	[57]
Nasp-T	Histone chaperone important for efficient replication Implicated in DNA repair processes	[15]

NASP splice isoform is expressed ubiquitously, the *NASP-T* splicing isoform has a much tighter anatomic distribution and its splicing is associated particularly with cancer cells and embryonic development. While most normal adult tissues do not splice the *NASP-T* exons into their mRNAs, high levels of splicing inclusion are seen in the testis and to a lesser extent the heart, gut, and ovary [15].

Splicing inclusion of the *NASP-T* exon is strongly activated in transfected cells in response to coexpression of Tra2β, and *NASP-T* splicing also decreases in *TRA2B* knockout mouse brains compared to wild type, confirming that the *NASP-T* exon is a bona fide regulated target exon of Tra2β [14, 15]. Tra2β is currently the only known splicing regulator of the *NASP-T* exon. The *NASP-T* exon is unusually long (a 975 nucleotide long cassette exon, while the typical size for a human exon is more like 120 nucleotides), with at least 37 Tra2β protein binding sites within its sequence, making a very responsive target for Tra2β expression. Splicing inclusion of the *NASP-T* exon into the *NASP* mRNA introduces the coding information for an extra 375 amino acids into the encoded NASP protein (Figure 3).

The NASP protein has a strongly biased peptide sequence which contains a high frequency of glutamic acid residues. The negative charges of the glutamic acid residues facilitate interactions with the positively charged histone partner proteins that NASP protein interacts with. NASP proteins also use tetratricopeptide repeats (TPRs) and histone binding motifs to facilitate interactions with protein partners including histones [44]. Both the somatic (sNASP) and NASP-T isoforms of the NASP protein contain the same TPRs involved in protein-protein interactions and seem to be functionally interchangeable in cells [45]. However, the longer NASP-T protein isoform has an additional histone binding motif and a longer stretch of the glutamic-acid-enriched sequence, suggesting that it might more efficiently interact with histones (Figure 3(a)). The NASP-T peptide cassette also adds a number of potentially phosphorylated serine and threonine residues to the NASP protein [44, 46]. Splicing inclusion of the *NASP-T* exon is likely to be important in cancer cells. The specific siRNA-mediated downregulation of *NASP* mRNAs containing the *NASP-T* exon leads to a block in proliferation and increased levels of apoptosis in cancer cells [47, 48].

Isoforms of the NASP protein with and without the peptide cassette inserted by the NASP-T exon are molecular chaperones which import histone H1 into the nucleus [49]. NASP protein isoforms also stably maintain the soluble pools of H3 and H4 histones needed for assembly of chromatin at times of high replication activity and are part of the complexes which load these into chromatin [45]. The *NASP* gene is critical for cell cycle progression in cultured cells and for mouse embryogenesis [50].

Why might NASP protein be important for cancer cells? *NASP* belongs to a network of genes important for cell survival [51], and NASP protein is a tumour-associated antigen in ovarian cancer [52]. NASP is highly expressed in S phase of the cell cycle [49], when chromatin needs to be reassembled after replication. Higher levels of NASP protein expression might be needed by cancer cells to enable their higher rates of replication to be achieved. NASP protein also has other roles related to chromatin stability. NASP protein is phosphorylated by the ATM and ATR kinases in response to ionising radiation and implicated in the repair of DNA double strand breaks [53]. One of the protein partners of NASP protein is the DNA repair protein Ku, and the yeast homologue of NASP is present at double strand breaks suggesting an important role in DNA repair (reviewed in [44]).

The second known splicing target of Tra2β with likely important functions in cancer cells is within the *CD44* pre-mRNA. *CD44* encodes an important transmembrane protein partly displayed on the cell surface as the CD44 antigen (Figure 3(b)). CD44 protein acts as a receptor for hyaluronic acid and possibly other molecules and controls interactions with other cells, the extracellular matrix, and cellular motility through modulation of intracellular signalling cascades [54].

The N- and C-termini of the CD44 protein are encoded by constitutive exons, but the *CD44* gene also contains an internal block of 10 consecutive internal alternative exons which are differentially regulated during development and in cancer [55]. These alternative exons encode portions of the extracellular domain of the protein (Figure 3(b)). *CD44* variable exons show variant splicing inclusion in breast cancer cells [30]. In particular, two *CD44* internal variable exons, *CD44*v4 and *CD44*v5, increase their splicing inclusion in transfected HeLa cells in response to increased Tra2β protein expression [30], suggesting that Tra2β might also increase their inclusion in breast tumours with elevated Tra2β expression. Although expression of variant CD44 exons has historically been associated with cancer metastasis, the picture regarding *CD44* alternative splicing in cancer is complex. Very recent data suggest that the standard isoform of *CD44* mRNA (without splicing inclusion of its variable exons) might in fact play a key role in metastatic breast cancer, particularly in enabling an epithelial-mesenchyme transition of breast cancer cells [56].

The third known Tra2β-target exon which might be potentially relevant in cancer cells is in the *HIPK3* gene, which encodes a serine/threonine kinase involved in transcriptional regulation and negative control of apoptosis. High cellular levels of Tra2β stimulate splicing inclusion of a poison exon called HIPK3-T into the HIPK3 mRNA [57]. Normal HIPK3 protein is concentrated in subnuclear structures called promyelocytic leukemia bodies (PML bodies). The shorter

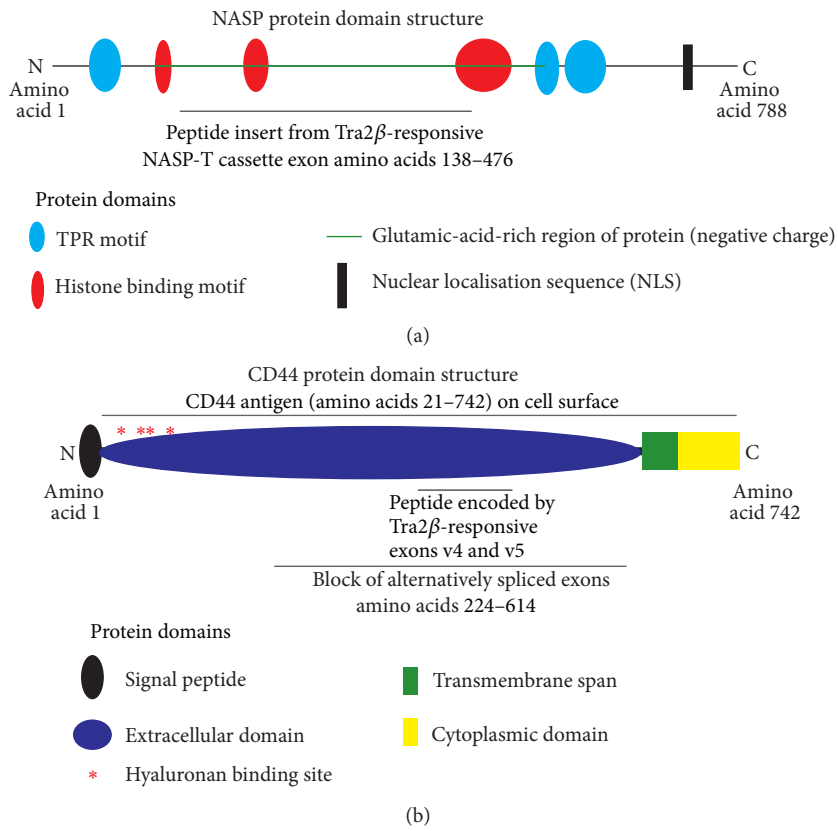


Figure 3: Protein domain architecture of known Tra2β splicing targets which are expressed in cancer cells. (a) Modular structure of NASP protein assembled from the UniProt database (http://www.uniprot.org/uniprot/P49321) [46], showing the position of the peptide insert encoded by the Tra2β-target exon NASP-T. (b) Modular structure of CD44 protein assembled using information from the UniProt database (http://www.uniprot.org/uniprot/P16070#P16070-6) [46], showing the position of peptide sequences encoded by the Tra2β-target exons CD44 v4 and v5. The CD44 antigen is displayed on the cell surface, and the protein is anchored on the cell surface by a single trans-membrane domain. Alternative isoforms are made through alternative splicing of 10 exons out of 19 encoding amino acids in the extracellular domain and also 2 exons which encode peptide sequence in the cytoplasmic domain. The two exons reported CD44 v4 and v5 exons correspond to amino acids 386–428 and 429–472, respectively, in the encoded protein. The protein domain structures are not drawn to scale.

HIPK3 protein isoform made under control of Tra2β fails to localise in PML bodies and lacks regions of the protein predicted to bind the androgen receptor, homeodomains, Fas, and p53 [57]. HIPK3-T is not confirmed as a splicing target of Tra2β in cancer, since splicing of the HIPK3-T exon has only been observed thus far in human testis and has not been directly reported from cancer cells [57].

6. Tra2β Is Involved in Protein Interaction Networks with Partner Proteins Involved in Cancer

Some of the proteins which are known to interact either directly or indirectly with Tra2β have themselves been implicated with roles in cancer cells. Tra2β directly interacts with members of the hnRNP G family of proteins which includes the prototypic member hnRNP G (encoded by the *RBMX* gene located on the X chromosome); RBMY protein (which is encoded by a multigene family on the Y chromosome); and a number of retrogene-derived proteins. Of these retrogene-derived proteins, one called HNRNP G-T is both highly conserved in mammals and specifically expressed in meiosis. The interaction between Tra2β and hnRNP G family members likely buffers the splicing activity of Tra2β [58, 59], although they might also coregulate some target exons [60]. Expression of the RBMY protein has been directly implicated in liver cancer biology, where it may contribute to the male specificity of this cancer [61]. RBMY protein also interacts with SRSF3 protein [62].

7. Summary

The splicing regulator Tra2β is upregulated in some human cancers. Possible mechanisms for this upregulation include changes in oncogenic transcription factor expression and oxygen free radical concentrations in neoplastic tissue, both of which affect *TRA2B* gene expression (Figure 4). We do not currently know whether the *TRA2B* gene can function as an oncogene in its own right until experiments to test transformation of NIH3T3 cells are performed or the behaviour of such transformed cells in nude mice is tested. However, we

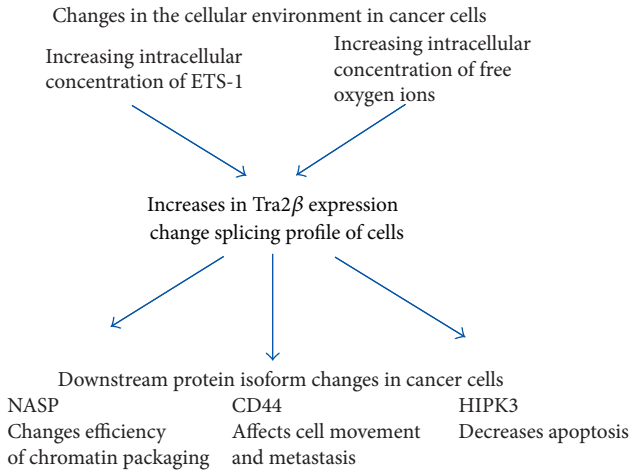


Figure 4: Hypothetical model suggesting how changes in the cellular environment may influence the expression of Tra2β and lead to downstream changes in mRNA splice isoform production.

do know that some of the known splicing targets of Tra2β identified in normal tissues are important for cancer cell biology and are particularly implicated in cell division and motility. Tra2β is essential during embryonic development, and many embryonic developmental pathways involved in cell growth and motility which are turned off in adult cells often become reactivated in cancer cells. Future analysis of the role of Tra2β in cancer cells will require the detailed identification of its endogenous splicing targets in cancer cells and the elucidation of their physiological roles.

Acknowledgments

This work was funded by the Breast Cancer Campaign, the Wellcome Trust (Grant numbers WT080368MA and WT089225/Z/09/Z), and the BBSRC (grant numbers BB/D013917/1 and BB/I006923/1).

References

[1] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: the next generation," *Cell*, vol. 144, no. 5, pp. 646–674, 2011.

[2] P. Rajan, D. J. Elliott, C. N. Robson, and H. Y. Leung, "Alternative splicing and biological heterogeneity in prostate cancer," *Nature Reviews Urology*, vol. 6, no. 8, pp. 454–460, 2009.

[3] J. P. Venables, "Aberrant and alternative splicing in cancer," *Cancer Research*, vol. 64, no. 21, pp. 7647–7654, 2004.

[4] J. P. Venables, "Unbalanced alternative splicing and its significance in cancer," *BioEssays*, vol. 28, no. 4, pp. 378–386, 2006.

[5] V. Quesada, L. Conde, N. Villamor et al., "Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia," *Nature Genetics*, vol. 44, no. 1, pp. 47–52, 2012.

[6] A. R. Grosso, S. Martins, and M. Carmo-Fonseca, "The emerging role of splicing factors in cancer," *EMBO Reports*, vol. 9, no. 11, pp. 1087–1093, 2008.

[7] S. Bonnal, L. Vigevani, and J. Valcarcel, "The spliceosome as a target of novel antitumour drugs," *Nature Reviews Drug Discovery*, vol. 11, pp. 847–859, 2012.

[8] J. C. Long and J. F. Caceres, "The SR protein family of splicing factors: master regulators of gene expression," *Biochemical Journal*, vol. 417, no. 1, pp. 15–27, 2009.

[9] P. J. Shepard and K. J. Hertel, "The SR protein family," *Genome Biology*, vol. 10, no. 10, p. 242, 2009.

[10] Z. Zhou and X. D. Fu, "Regulation of splicing by SR proteins and SR protein-specific kinases," *Chromosoma*, vol. 122, no. 3, pp. 191–207, 2013.

[11] R. Tacke, M. Tohyama, S. Ogawa, and J. L. Manley, "Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing," *Cell*, vol. 93, no. 1, pp. 139–148, 1998.

[12] A. R. Krainer and T. Maniatis, "Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for Pre-mRNA splicing in vitro," *Cell*, vol. 42, no. 3, pp. 725–736, 1985.

[13] A. M. Zahler, K. M. Neugebauer, W. S. Lane, and M. B. Roth, "Distinct functions of SR proteins in alternative pre-mRNA splicing," *Science*, vol. 260, no. 5105, pp. 219–222, 1993.

[14] D. J. Elliott, A. Best, C. Dalgliesh, I. Ehrmann, and S. Grellscheid, "How does Tra2beta protein regulate tissue-specific RNA splicing?" *Biochemical Society Transactions*, vol. 40, pp. 784–788, 2012.

[15] S. Grellscheid, C. Dalgliesh, M. Storbeck et al., "Identification of evolutionarily conserved exons as regulated targets for the splicing activator Tra2β in development," *PLoS Genetics*, vol. 7, no. 12, Article ID e1002390, 2011.

[16] A. Cléry, S. Jayne, N. Benderska, C. Dominguez, S. Stamm, and F. H.-T. Allain, "Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1," *Nature Structural and Molecular Biology*, vol. 18, no. 4, pp. 443–451, 2011.

[17] K. Tsuda, T. Someya, K. Kuwasako et al., "Structural basis for the dual RNA-recognition modes of human Tra2-β RRM," *Nucleic Acids Research*, vol. 39, no. 4, pp. 1538–1553, 2011.

[18] T. Novoyatleva, B. Heinrich, Y. Tang et al., "Protein phosphatase 1 binds to the RNA recognition motif of several splicing factors and regulates alternative pre-mRNA processing," *Human Molecular Genetics*, vol. 17, no. 1, pp. 52–70, 2008.

[19] Y. Mende, M. Jakubik, M. Riessland et al., "Deficiency of the splicing factor Sfrs10 results in early embryonic lethality in mice and has no impact on full-length SMN /Smn splicing," *Human Molecular Genetics*, vol. 19, no. 11, Article ID ddq094, pp. 2154–2167, 2010.

[20] L. R. Meyer, A. S. Zweig, A. S. Hinrichs et al., "The UCSC Genome Browser database: extensions and updates 2013," *Nucleic Acids Research*, vol. 41, pp. D64–D69, 2013.

[21] R. Karni, E. De Stanchina, S. W. Lowe, R. Sinha, D. Mu, and A. R. Krainer, "The gene encoding the splicing factor SF2/ASF is a proto-oncogene," *Nature Structural and Molecular Biology*, vol. 14, no. 3, pp. 185–193, 2007.

[22] E. Cerami, J. Gao, U. Dogrusoz et al., "The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data," *Cancer Discovery*, vol. 2, pp. 401–404, 2012.

[23] J. Gao, B. A. Aksoy, U. Dogrusoz et al., "Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal," *Science Signaling*, vol. 6, no. 269, p. pl1, 2013.

[24] S. Das, O. Anczukow, M. Akerman, and A. R. Krainer, "Oncogenic splicing factor SRSF1 is a critical transcriptional target of MYC," *Cell Reports*, vol. 1, pp. 110–117, 2012.

[25] O. Anczuków, A. Z. Rosenberg, M. Akerman et al., "The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation," *Nature Structural and Molecular Biology*, vol. 19, no. 2, pp. 220–228, 2012.

[26] R. Jia, C. Li, J. P. McCoy, C.-X. Deng, and Z.-M. Zheng, "SRp20 is a proto-oncogene critical for cell proliferation and tumor induction and maintenance," *International Journal of Biological Sciences*, vol. 6, no. 7, pp. 806–826, 2010.

[27] X. He, A. D. Arslan, M. D. Pool et al., "Knockdown of splicing factor SRp20 causes apoptosis in ovarian cancer cells and its expression is associated with malignancy of epithelial ovarian cancer," *Oncogene*, vol. 30, no. 3, pp. 356–365, 2011.

[28] K. Kurokawa, Y. Akaike, K. Masuda et al., "Downregulation of serine/arginine-rich splicing factor 3 induces G1 cell cycle arrest and apoptosis in colon cancer cells," *Oncogene*, 2013.

[29] B. Gabriel, A. Z. Hausen, J. Bouda et al., "Significance of nuclear hTra2-beta1 expression in cervical cancer," *Acta Obstetricia et Gynecologica Scandinavica*, vol. 88, no. 2, pp. 216–221, 2009.

[30] D. O. Watermann, Y. Tang, A. Z. Hausen, M. Jäger, S. Stamm, and E. Stickeler, "Splicing factor Tra2-β1 is specifically induced in breast cancer and regulates alternative splicing of the CD44 gene," *Cancer Research*, vol. 66, no. 9, pp. 4774–4780, 2006.

[31] D.-C. Fischer, K. Noack, I. B. Runnebaum et al., "Expression of splicing factors in human ovarian cancer," *Oncology Reports*, vol. 11, no. 5, pp. 1085–1090, 2004.

[32] K. Kajita, Y. Kuwano, N. Kitamura et al., "Ets1 and heat shock factor 1 regulate transcription of the *Transformer* 2β gene in human colon cancer cells," *Journal of Gastroenterology*, 2013.

[33] K. Takeo, T. Kawai, K. Nishida et al., "Oxidative stress-induced alternative splicing of transformer 2β (SFRS10) and CD44 pre-mRNAs in gastric epithelial cells," *American Journal of Physiology*, vol. 297, no. 2, pp. C330–C338, 2009.

[34] D. W. Lincoln II and K. Bove, "The transcription factor Ets-1 in breast cancer," *Frontiers in Bioscience*, vol. 10, pp. 506–511, 2005.

[35] D. W. Lincoln II, P. G. Phillips, and K. Bove, "Estrogen-induced Ets-1 promotes capillary formation in an in vitro tumor angiogenesis model," *Breast Cancer Research and Treatment*, vol. 78, no. 2, pp. 167–178, 2003.

[36] J. Dittmer, "The biology of the Ets1 proto-oncogene," *Molecular Cancer*, vol. 2, article 29, 2003.

[37] X. Zhang, A. N. Moor, K. A. Merkler, Q. Liu, and M. P. McLean, "Regulation of alternative splicing of liver scavenger receptor class B gene by estrogen and the involved regulatory splicing factors," *Endocrinology*, vol. 148, no. 11, pp. 5295–5304, 2007.

[38] Y. Tsukamoto, N. Matsuo, K. Ozawa et al., "Expression of a novel RNA-splicing factor, RA301/Tra2β, in vascular lesions and its role in smooth muscle cell proliferation," *American Journal of Pathology*, vol. 158, no. 5, pp. 1685–1694, 2001.

[39] N. Matsuo, S. Ogawa, Y. Imai et al., "Cloning of a novel RNA binding polypeptide (RA301) induced by hypoxia/reoxygenation," *Journal of Biological Chemistry*, vol. 270, no. 47, pp. 28216–28222, 1995.

[40] R. Daoud, G. Mies, A. Smialowska, L. Oláh, K.-A. Hossmann, and S. Stamm, "Ischemia induces a translocation of the splicing factor tra2-β1 and changes alternative splicing patterns in the brain," *Journal of Neuroscience*, vol. 22, no. 14, pp. 5889–5899, 2002.

[41] N. J. McGlincy and C. W. J. Smith, "Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense?" *Trends in Biochemical Sciences*, vol. 33, no. 8, pp. 385–393, 2008.

[42] P. Stoilov, R. Dauod, O. Nayler, and S. Stamm, "Human tra2-beta1 autoregulates its protein concentration by infuencing alternative splicing of its pre-mRNA," *Human Molecular Genetics*, vol. 13, no. 5, pp. 509–524, 2004.

[43] L. F. Lareau, M. Inada, R. E. Green, J. C. Wengrod, and S. E. Brenner, "Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements," *Nature*, vol. 446, no. 7138, pp. 926–929, 2007.

[44] R. M. Finn, K. Ellard, J. M. Eirin-Lopez, and J. Ausio, "Vertebrate nucleoplasmin and NASP: egg histone storage proteins with multiple chaperone activities," *The FASEB Journal*, vol. 26, pp. 4788–4804, 2012.

[45] A. J. L. Cook, Z. A. Gurard-Levin, I. Vassias, and G. Almouzni, "A specific function for the histone chaperone NASP to fine-tune a reservoir of soluble H3-H4 in the histone supply chain," *Molecular Cell*, vol. 44, no. 6, pp. 918–927, 2011.

[46] UniProt Consortium, "Reorganizing the protein space at the Universal Protein Resource (UniProt)," *Nucleic Acids Research*, vol. 40, pp. D71–D75, 2012.

[47] O. M. Alekseev, R. T. Richardson, J. K. Tsuruta, and M. G. O'Rand, "Depletion of the histone chaperone tNASP inhibits proliferation and induces apoptosis in prostate cancer PC-3 cells," *Reproductive Biology and Endocrinology*, vol. 9, p. 50, 2011.

[48] W. Ma, S. Xie, M. Ni et al., "MicroRNA-29a inhibited epididymal epithelial cell proliferation by targeting nuclear autoantigenic sperm protein (NASP)," *Journal of Biological Chemistry*, vol. 287, no. 13, pp. 10189–10199, 2012.

[49] R. T. Richardson, I. N. Batova, E. E. Widgren et al., "Characterization of the histone H1-binding protein, NASP, as a cell cycle-regulated somatic protein," *Journal of Biological Chemistry*, vol. 275, no. 39, pp. 30378–30386, 2000.

[50] R. T. Richardson, O. M. Alekseev, G. Grossman et al., "Nuclear autoantigenic sperm protein (NASP), a linker histone chaperone that is required for cell proliferation," *Journal of Biological Chemistry*, vol. 281, no. 30, pp. 21526–21534, 2006.

[51] O. M. Alekseev, R. T. Richardson, O. Alekseev, and M. G. O'Rand, "Analysis of gene expression profiles in HeLa cells in response to overexpression or siRNA-mediated depletion of NASP," *Reproductive Biology and Endocrinology*, vol. 7, article 45, 2009.

[52] R. Ali-Fehmi, M. Chatterjee, A. Ionan et al., "Analysis of the expression of human tumor antigens in ovarian cancer tissues," *Cancer Biomarkers*, vol. 6, no. 1, pp. 33–48, 2009.

[53] S. Matsuoka, B. A. Ballif, A. Smogorzewska et al., "ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage," *Science*, vol. 316, no. 5828, pp. 1160–1166, 2007.

[54] H. Ponta, L. Sherman, and P. A. Herrlich, "CD44: from adhesion molecules to signalling regulators," *Nature Reviews Molecular Cell Biology*, vol. 4, no. 1, pp. 33–45, 2003.

[55] G. R. Screaton, M. V. Bell, D. G. Jackson, F. B. Cornelis, U. Gerth, and J. I. Bell, "Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 24, pp. 12160–12164, 1992.

[56] R. L. Brown, L. M. Reinke, M. S. Damerow et al., "CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression," *Journal of Clinical Investigation*, vol. 121, no. 3, pp. 1064–1074, 2011.

[57] J. P. Venables, C. F. Bourgeois, C. Dalgliesh, L. Kister, J. Stevenin, and D. J. Elliott, "Up-regulation of the ubiquitous alternative splicing factor Tra2β causes inclusion of a germ cell-specific exon," *Human Molecular Genetics*, vol. 14, no. 16, pp. 2289–2303, 2005.

[58] Y. Liu, C. F. Bourgeois, S. Pang et al., "The germ cell nuclear proteins hnRNP G-T and RBMY activate a testis-specific exon," *PLoS Genetics*, vol. 5, no. 11, Article ID e1000707, 2009.

[59] M. T. Nasim, T. K. Chernova, H. M. Chowdhury, B.-G. Yue, and I. C. Eperon, "HnRNP G and Tra2β: opposite effects on splicing matched by antagonism in RNA binding," *Human Molecular Genetics*, vol. 12, no. 11, pp. 1337–1348, 2003.

[60] Y. Hofmann, C. L. Lorson, S. Stamm, E. J. Androphy, and B. Wirth, "Htra2-β1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 17, pp. 9618–9623, 2000.

[61] D.-J. Tsuei, P.-H. Lee, H.-Y. Peng et al., "Male germ cell-specific RNA binding protein RBMY: a new oncogene explaining male predominance in liver cancer," *PLoS ONE*, vol. 6, no. 11, Article ID e26948, 2011.

[62] D. J. Elliott, C. F. Bourgeois, A. Klink, J. Stévenin, and H. J. Cooke, "A mammalian germ cell-specific RNA-binding protein interacts with ubiquitously expressed proteins involved in splice site selection," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 11, pp. 5717–5722, 2000.

[63] M. J. Moore, Q. Wang, C. J. Kennedy, and P. A. Silver, "An alternative splicing network links cell-cycle control to apoptosis," *Cell*, vol. 142, no. 4, pp. 625–636, 2010.

[64] M. P. Paronetto, T. Achsel, A. Massiello, C. E. Chalfant, and C. Sette, "The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x," *Journal of Cell Biology*, vol. 176, no. 7, pp. 929–939, 2007.

[65] H. L. Gautrey and A. J. Tyson-Capper, "Regulation of Mcl-1 by SRSF1 and SRSF5 in cancer cells," *PLoS ONE*, vol. 7, Article ID e51497, 2012.

How does Tra2β protein regulate tissue-specific RNA splicing?

David J. Elliott[1], Andrew Best, Caroline Dalgliesh, Ingrid Ehrmann and Sushma Grellscheid
Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle upon Tyne NE1 3BZ, U.K.

Abstract

The splicing regulator protein Tra2β is conserved between humans and insects and is essential for mouse development. Recent identification of physiological RNA targets has started to uncover molecular targets and mechanisms of action of Tra2β. At a transcriptome-wide level, Tra2β protein binds a matrix of AGAA-rich sequences mapping frequently to exons. Particular tissue-specific alternatively spliced exons contain high concentrations of high scoring Tra2β-binding sites and bind Tra2β strongly *in vitro*. These top exons were also activated for splicing inclusion *in cellulo* by co-expression of Tra2β protein and were significantly down-regulated after genetic depletion of Tra2β. Tra2β itself seems to be fairly evenly expressed across several different mouse tissues. In the present paper, we review the properties of Tra2β and its regulated target exons, and mechanisms through which this fairly evenly expressed alternative splicing regulator might drive tissue-specific splicing patterns.

Tra2β protein is a splicing activator conserved between fruitflies and mice

Alternative splicing introduces new coding information into mRNAs, and so plays a pivotal role in expanding genome capacity to encode more proteins than just the ~23 000 that would be expected if each gene encoded a single protein [1]. Alternative splice events are controlled in part by a number of different RNA-binding proteins attaching to pre-mRNAs, although links with transcription and epigenetic modification of the template chromatin are also important. A large group of splicing regulator proteins contain domains that are enriched in arginine and serine residues (so-called RS domains, based on the one letter amino acid code) [2,3]. These include Tra2 proteins, which have a modular organization comprising a single central RRM (RNA recognition motif) flanked either side by RS domains [4,5].

A single Tra2 protein is found in fruitflies, where Tra2 is one of the classical splicing regulators controlling sexual differentiation as well as being essential for spermatogenesis [6–8]. The fruitfly gene is the source of the acronym Tra2, which is short for transformer 2, since mutations in this gene transform sexual phenotype. The *Tra2* gene has duplicated in vertebrates, resulting in two mammalian Tra2 proteins with 63% amino acid identity (aligned at http://www.ebi.ac.uk/Tools/emboss/align/). These proteins are called Tra2α (encoded by the *Tra2a* gene on mouse chromosome 6) and Tra2β (encoded by the *Sfrs10* gene on mouse chromosome 16). Tra2β binds to exons to regulate their alternative splicing inclusion. For example, Tra2β binds to the testis-specific T exon in the *homeodomain interacting protein kinase 3* gene to regulate its splicing inclusion in the testis [9,10]. Recently, the details of exactly how Tra2β protein binds to both AGAA and CAA target RNA sequences have been revealed at atomic resolution, and involve protein–RNA interactions with both the RRM and flanking regions [11,12].

Key words: development, gene expression, high-throughput sequencing of RNAs isolated by cross-linking immunoprecipitation (HITS-CLIP), RNA-binding proteins, RNA splicing, Tra2β, transcriptome.

Abbreviations used: HITS-CLIP, high-throughput sequencing of RNAs isolated by cross-linking immunoprecipitation; RRM, RNA recognition motif; SMA, spinal muscular atrophy.

[1]To whom correspondence should be addressed (email David.Elliott@ncl.ac.uk).

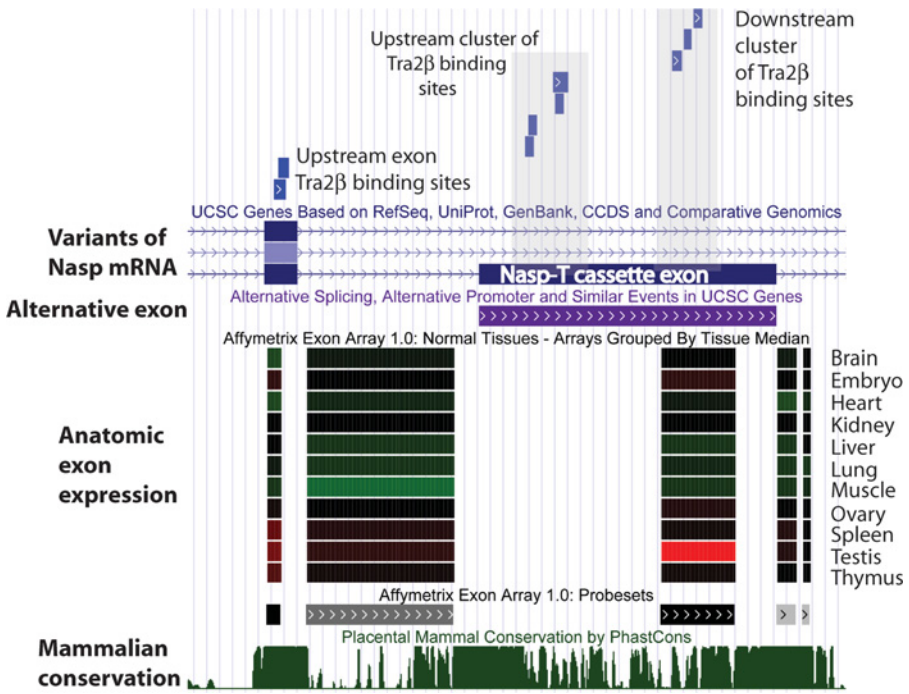
Transcriptome-wide identification of splicing targets for Tra2β

Despite similar amino acid sequences and RNA-binding specificities between Tra2α and Tra2β [13], genetic deletion of the *Sfrs10* gene still results in embryonic lethality even though the *Tra2a* gene remains intact [14]. *Sfrs10*$^{-/-}$ mice die at approximately 12 days gestation. This indicates either non-redundancy with Tra2α at this stage of development or that expression levels achieved from both genes are needed for embryonic development [14]. Similarly, Tra2β is essential in the embryonic brain [15].

Because of the known functions for Tra2β protein in splicing, it is likely that defects in splicing regulation are a major contributing factor to the embryonic death of *Sfrs10*$^{-/-}$ mice. Such defects would probably lead to downstream changes in mRNA and protein isoforms impacting on development. Previously there were just a handful of known splicing targets for Tra2β, and these were not clearly mis-regulated in the absence of Tra2β [10,16]. Recently HITS-CLIP (high-throughput sequencing of RNAs isolated by cross-linking immunoprecipitation) has been used to comprehensively identify endogenous RNA targets for Tra2β during mouse germ cell development [15]. In this procedure, endogenous target RNAs are cross-linked by UV radiation, then short fragments bound to Tra2β are rigorously purified, amplified, deep sequenced and mapped

Figure 1 | **HITS-CLIP physically identifies multiple Tra2β-binding sites on the *Nasp-T* exon**

This screenshot shows the location of *in vivo*-binding sites for Tra2β within a region of the mouse transcriptome corresponding to the *Nasp-T* exon and flanking intron sequences. The position of physical RNA–Tra2β protein interactions identified by HITS-CLIP are annotated (blue boxes). Clusters of Tra2β-binding sites mapped to the upstream and downstream regions of the *Nasp-T* exon. Additional sites of Tra2β binding were observed in the upstream constitutive exon. Additional tracks shown are alternative events (highlighting the *Nasp-T* exon, shown in purple); the Affymetrix exon array dataset corresponding to this position in the transcriptome, indicating tissue-specific RNA expression levels (notice high levels of *Nasp-T* exon inclusion are seen in the testis, whereas the upstream exon has similar levels of inclusion in other tissues, including spleen and thymus); and conservation levels in mammals. This screenshot is modified from the UCSC mouse genome browser [19].



on to the genome. As an illustration of the resolution of this technique in identifying binding sites *in vivo*, the CLIP tags mapping to one of these identified target exons (*Nasp-T*) is shown in Figure 1 [17].

Target exons that depend on Tra2β for splicing inclusion should be mis-spliced in the absence of the Tra2β protein. Splicing analyses in the brains of mice, which contain *Sfrs10*$^{-/-}$ neurons indicated splicing inclusion of at least two of the identified target exons (within the *Nasp* and *Tra2a* genes) identified from the HITS-CLIP screen, were indeed strongly down-regulated in the absence of Tra2β protein, thus identifying these as physiologically regulated target exons [15].

The identification of these target exons reveals for the first time biological functions operating downstream of the Tra2β protein in mouse development. The *Nasp* gene itself is essential for mouse development [18] and encodes a protein that mediates histone import into the nucleus and assembly of chromatin after replication. The Tra2β-regulated *Nasp-T* exon is very long at 975 nt [19], making this single exon almost as long as the rest of the *Nasp* mRNA. This 975 nt exon length is divisible by three, and inserts the ORF (open reading frame) for a peptide cassette into the *Nasp* mRNA in those tissues where it is spliced. In adult mice, the *Nasp-T* exon is most actively spliced in the testis (it is also spliced in the embryo). In the testis, the longer encoded Nasp protein isoform is associated with meiotic chromosomes where it forms part of the complex that monitors the completion of double-strand break repair [18,20–22]. Hence, by regulating splicing of the *NaspT* exon, Tra2β operates as one of the key upstream regulators for this crucial physiological process in mouse male germ cells.

Tra2β also physiologically regulates splicing inclusion of a 'poison' exon into the *Tra2a* mRNA [9,23]. This poison exon inserts in-frame stop codons into the *Tra2a* mRNA, thereby targeting it for nonsense-mediated decay and preventing subsequent production of Tra2α protein. The identification of this regulated poison exon has thus revealed a novel pathway of feedback control, which operates between the vertebrate Tra2 proteins: overexpression of Tra2β protein leads to increased splicing inclusion of the *Tra2a* poison exon and down-regulation of Tra2α expression.

Both the *Nasp-T* and *Tra2a* exons are found in all available vertebrate genome sequences, indicating strong selective pressure [24–28]. High conservation of the *Nasp-T* and *Tra2a* exons is particularly remarkable in the context of the

testis, since many alternative exons included in this tissue are not conserved between species [24]. Functionally important exons also tend to be frequently included into mRNAs in at least some tissues. In mice, the *Nasp-T* and *Tra2a* exons are spliced into mRNAs at high levels in testis, and also spliced into mRNAs in other adult tissues and in the embryo. The defects in *Nasp-T* and *Tra2a* splicing may therefore contribute to the phenotype of the *Sfrs10*$^{-/-}$ mice. Since RNA-binding proteins are global regulators, defects resulting from their absence are likely to be widespread. Tra2β can bind and potentially regulate thousands of exons in parallel across the transcriptome, many of which might contribute to the embryonic or brain phenotype when Tra2β is missing [14,15].

Multiple binding sites are needed for Tra2β-mediated splicing activation of *Nasp-T* and *Tra2a*

Both the *Nasp-T* and *Tra2a* poison exons have two key features relating to their particular dependence on Tra2β for splicing inclusion, as discussed below.

A high frequency of Tra2β-binding sites

Both the *Nasp-T* and *Tra2a* exons contain multiple Tra2β-binding sites inferred from transcriptome-wide analysis [15]. Within the transcriptome-wide dataset of RNA targets, particular 6-mer sequences were identified as enriched compared with either the genomic- or testis-specific transcriptomic backgrounds. Each of these most frequently recovered 6-mers represent binding sites for Tra2β recognized *in vivo* within the mouse transcriptome, and were subtle variations of AGAA-rich sequences, very similar to the known Tra2β-binding site [11,12].

These frequently recovered 6-mers were used as a Tra2β-binding site matrix, with the most frequently recovered 6-mers representing the best physiological binding sites. When analysed according to preferred physiological binding sites, the *Nasp-T* exon contains approximately 37 Tra2β-binding sites (defined as containing sequences from the 25 most frequently recovered 6-mers in the transcriptome-wide dataset, with particular importance being placed on the top five recovered 6-mers); and the *Tra2a* exon contains approximately 12 binding sites [15]. Note that these numbers are approximations: some predicted Tra2β-binding sites overlap, making it difficult to be precise. Tra2β-binding sites clustered in the upstream and downstream portions of the *Nasp-T* exon identified by CLIP are shown in Figure 1.

Highly efficient Tra2β binding

Consistent with their high concentration in binding sites, both the *Nasp-T* and *Tra2a* exons bound Tra2β protein very efficiently in gel shift experiments, and less stable complexes were formed after binding site mutagenesis [15].

By looking at the exon sequences of *Nasp-T* and *Tra2a*, it is possible to estimate when splicing regulation by Tra2β might have evolved. Superimposition of binding sites and detailed phylogenetic comparisons of exon sequence show that the *Tra2a* poison exon is likely to be controlled by Tra2β across all vertebrates (all of which contain multiple binding sites), whereas Tra2β control of *Nasp-T* is more likely to have evolved in mammals [15].

Models for physiological splicing control by Tra2β

The *Sfrs10* gene is expressed fairly evenly in different mouse tissues [15]. This raises the important question: how does Tra2β regulate tissue-specific splicing patterns? Two conceptually different (but not mutually exclusive) models might explain how Tra2β protein operates as a tissue-specific splicing factor.

Tissue-specific patterns of splicing may be controlled by differences in the cellular concentration of Tra2β (Figure 2A).

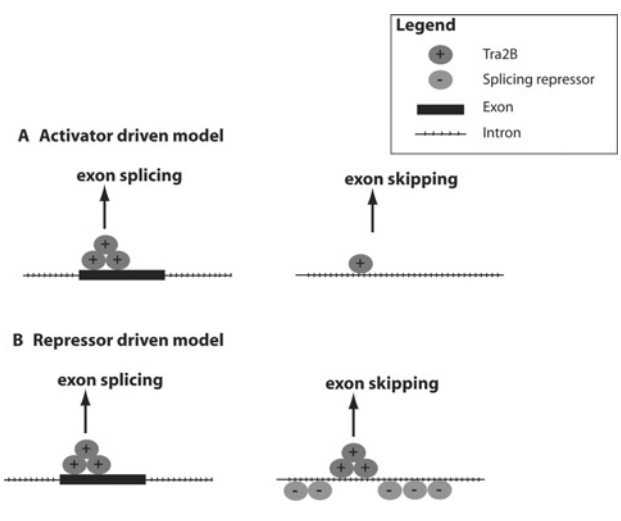
Subtle but important differences in Tra2β protein concentration might occur both between and within tissues to drive tissue-specific splicing patterns. For example, within the mouse testis, Tra2β protein is low in spermatogonia (a cell population including the stem cells), but expression is higher in spermatocytes (the meiotic cells) [15]. Tra2β protein levels then fall again in round spermatids (the post-meiotic haploid cells). The splicing inclusion of exons such as *Nasp-T* and *Tra2a* thus might peak in spermatocytes, which have the highest Tra2β protein concentration to activate splicing inclusion. When compared between different tissues, total RNA and protein preparations would conceal these intrinsic gradients of Tra2β protein concentration. Another piece of evidence argues for this kind of differential splicing control within tissues. Tra2β protein regulates its own level through regulating splicing inclusion of a poison exon into its own mRNA [29], and splicing levels of this poison exon are highest in the testis, possibly also in meiotic cells expressing maximum levels of Tra2β protein. Muscle also shows high levels of inclusion of this poison exon [15].

Tra2β might regulate specific exons that are de-repressed in particular tissues or cell types and thus available for activation (Figure 2B).

In this model, alternative splicing changes might be driven not directly by changes in Tra2β protein concentration, but rather in changes in the concentration of splicing repressors which antagonize Tra2β. Most exons are under combinatorial control, meaning they are influenced by a combination of both activating and repressive activities of splicing factors [30]. Particular tissue-specific exons that contain Tra2β protein-binding sites might be strongly repressed in some tissues due to high concentrations of these splicing repressors, but more weakly repressed in other tissues or cell types where the expression of these repressors is weaker. In tissues containing lower levels of splicing repressors, Tra2β might

Figure 2 | Local levels of splicing activation and repression may co-ordinately control tissue-specific *Nasp-T* splicing inclusion

(**A**) Activator-driven model of alternative splicing inclusion, in which changes in concentration of Tra2β protein between cells determine the splicing outcome. On the left-hand side, a high concentration of Tra2β binds to the RNA and activates splicing of the target exon (exon splicing). On the right-hand side, a decreased cellular concentration is not sufficient to activate exon splicing (so the exon is skipped). (**B**) Repressor-driven model of alternative splicing activation. In this model, tissues might have similar expression levels of Tra2β, but vary in expression of splicing repressor proteins, which repress Tra2β-target exons. In the left-hand side example, binding of Tra2β protein to the exon would activate splicing inclusion (no repressor is present). In the right-hand side example, there is a higher concentration of splicing repressor protein present and bound around the target exon. In this case, the same amount of Tra2β still binds to the target RNA, but cannot overcome the repressive effects of locally bound negatively acting splicing repressors (usually heterogeneous nuclear ribonucleoproteins). As a result of this, the exon is skipped.



operate as a tissue-specific activator of exons which contain Tra2β-binding sites, even though its actual expression level may not be different compared with the former tissues. For example, hnRNPA1 protein (a splicing repressor) is down-regulated over the course of germ cell development [31], so de-repressed exons that contain Tra2β-binding sites might be co-ordinately activated at this time.

Irrespective of the actual model that explains tissue-specific splicing of individual exons by Tra2β (referred to as 'activator driven' or 'repressor driven' in Figure 2), experimental evidence directly shows multiple adjacent Tra2β-binding sites are important to activate tissue-specific exons (Figure 2) [15]. Since RNA–protein interactions are likely to be dynamic in the nucleus and have both on and off rates, multiple adjacent binding sites might simply act to increase the probability of occupancy by a single Tra2β protein at any one time to activate exon splicing. It is, however, difficult to see how mutation of just two binding sites out of a total of ~37 potential sites in *Nasp-T* could explain the ~80% reduction in splicing inclusion using this kinetic probability model [15].

A more likely scenario than the probability model is that adjacent RNA-binding sites might act as a platform to assemble multiple Tra2β proteins into a splicing activator complex [9]. Gel shifts show large protein complexes assembling on these exons *in vitro* [9,15]. For some exons such as *Nasp-T*, a version of Tra2β unable to directly bind RNA by itself can still co-activate splicing *in cellulo*, possibly by attaching to splicing complexes already nucleated by wild-type endogenous Tra2β protein on the regulated exons [9,15]. Binding site mutagenesis of *Nasp-T* also resulted in large differences in splicing inclusion levels between individual double mutants, indicating that the organization of available Tra2β-binding sites is also important as well as the number of sites available.

Overexpressed Tra2β protein might also play a key role in disease

The above studies are starting to address the role of Tra2β protein in normal development. Tra2β might have additional roles in situations when it is expressed above its normal cellular concentration, perhaps by enabling it to activate weaker target exons than would be normally regulated. One such important target exon is *SMN2* exon 7.

Splicing of *SMN2* exon 7 was not obviously affected in the *Sfrs10*$^{-/-}$ mice [14], but was activated by overexpression of Tra2β in transfected cells [32]. Regulation of *SMN2* makes overexpression of Tra2β a potential therapeutic option to treat the developmental disease SMA (spinal muscular atrophy). SMA is caused by deletion of the *SMN1* gene. Expression of the adjacent *SMN2* gene is normally very low, since exon 7 is poorly spliced into the *SMN2* mRNA, resulting in an unstable protein product. By improving exon 7 splicing, Tra2β could improve expression from the *SMN2* gene and ameliorate disease in some patients. In addition to potentially useful roles in therapy, Tra2β has also been implicated as a more sinister modifier of other diseases, including breast cancer [33] where overexpression leads to increased inclusion of *CD44* alternative exons associated with metastasis.

Conclusions

Recent research has identified new physiological target RNAs for Tra2β in development, and revealed it as an important developmental splicing regulator. Gradients in Tra2β protein concentrations may drive exon inclusion patterns over development, leading to the regulated production of developmentally important protein isoforms. Alternatively, Tra2β might provide a level of intrinsic background activation which is enough to activate available de-repressed exons that contain Tra2β-binding sites and generate tissue-specific splicing profiles. Future work will continue to elucidate the role of Tra2β both in splicing regulation, as well as in the metabolism of other kinds of cellular RNA target.

Funding

This work was supported by the Wellcome Trust [grant numbers WT080368MA and WT089225/Z/09/Z (to D.J.E.)], the Biotechnology and Biological Sciences Research Council [grant numbers BB/D013917/1 and BB/I006923/1 (to D.J.E.)] and the Breast Cancer Campaign (to D.J.E.).

References

1 Stamm, S. (2002) Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. Hum. Mol. Genet. **11**, 2409–2416
2 Shepard, P.J. and Hertel, K.J. (2009) The SR protein family. Genome Biol. **10**, 242
3 Long, J.C. and Caceres, J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. Biochem. J. **417**, 15–27
4 Beil, B., Screaton, G. and Stamm, S. (1997) Molecular cloning of htra2-β-1 and htra2-β-2, two human homologs of tra-2 generated by alternative splicing. DNA Cell Biol. **16**, 679–690
5 Dauwalder, B., Amaya-Manzanares, F. and Mattox, W. (1996) A human homologue of the *Drosophila* sex determination factor transformer-2 has conserved splicing regulatory functions. Proc. Natl. Acad. Sci. U.S.A. **93**, 9004–9009
6 Baker, B.S. (1989) Sex in flies: the splice of life. Nature **340**, 521–524
7 Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H. and Shimura, Y. (1991) Control of doublesex alternative splicing by transformer and transformer-2 in *Drosophila*. Science **252**, 833–836
8 Belote, J.M. and Baker, B.S. (1983) The dual functions of a sex determination gene in *Drosophila melanogaster*. Dev. Biol. **95**, 512–517
9 Grellscheid, S.N., Dalgliesh, C., Rozanska, A., Grellscheid, D., Bourgeois, C.F., Stévenin, J. and Elliott, D.J. (2011) Molecular design of a splicing switch responsive to the RNA binding protein Tra2β. Nucleic Acids Res. **39**, 8092–8104
10 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 Tra2β causes inclusion of a germ cell-specific exon. Hum. Mol. Genet. **14**, 2289–2303
11 Tsuda, K., Someya, T., Kuwasako, K., Takahashi, M., He, F., Unzai, S., Inoue, M., Harada, T., Watanabe, S., Terada, T. et al. (2011) Structural basis for the dual RNA-recognition modes of human Tra2β RRM. Nucleic Acids Res. **39**, 1538–1553
12 Clery, A., Jayne, S., Benderska, N., Dominguez, C., Stamm, S. and Allain, F.H. (2011) Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1. Nat. Struct. Mol. Biol. **18**, 443–450
13 Tacke, R., Tohyama, M., Ogawa, S. and Manley, J.L. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. Cell **93**, 139–148
14 Mende, Y., Jakubik, M., Riessland, M., Schoenen, F., Rossbach, K., Kleinridders, A., Köhler, 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**, 2154–2167
15 Grellscheid, S., Dalgliesh, C., Storbeck, M., Best, A., Liu, Y., Jakubik, M., Mende, Y., Ehrmann, I., Curk, T., Rossbach, K. et al. (2011) Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2β in development. PLoS Genet. **7**, e1002390
16 Gabut, M., Chaudhry, S. and Blencowe, B.J. (2008) SnapShot: the splicing regulatory machinery. Cell **133**, 192.e1
17 Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X. et al. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature **456**, 464–469
18 Richardson, R.T., Alekseev, O.M., Grossman, G., Widgren, E.E., Thresher, R., Wagner, E.J., Sullivan, K.D., Marzluff, W.F. and O'Rand, M.G. (2006) Nuclear autoantigenic sperm protein (NASP), a linker histone chaperone that is required for cell proliferation. J. Biol. Chem. **281**, 21526–21534
19 Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A. et al. (2011) The UCSC genome browser database: update 2011. Nucleic Acids Res. **39**, D876–D882
20 Alekseev, O.M., Richardson, R.T. and O'Rand, M.G. (2009) Linker histones stimulate HSPA2 ATPase activity through NASP binding and inhibit CDC2/cyclin B1 complex formation during meiosis in the mouse. Biol. Reprod. **81**, 739–748
21 Osakabe, A., Tachiwana, H., Matsunaga, T., Shiga, T., Nozawa, R.S., Obuse, C. and Kurumizaka, H. (2010) Nucleosome formation activity of human somatic nuclear autoantigenic sperm protein (sNASP). J. Biol. Chem. **285**, 11913–11921
22 Richardson, R.T., Batova, I.N., Widgren, E.E., Zheng, L.X., Whitfield, M., Marzluff, W.F. and O'Rand, M.G. (2000) Characterization of the histone H1-binding protein, NASP, as a cell cycle-regulated somatic protein. J. Biol. Chem. **275**, 30378–30386
23 Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C. and Brenner, S.E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature **446**, 926–929
24 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**, 5659–5666
25 Lareau, L.F., Green, R.E., Bhatnagar, R.S. and Brenner, S.E. (2004) The evolving roles of alternative splicing. Curr. Opin. Struct. Biol. **14**, 273–282
26 Sorek, R., Shamir., R and Ast, G. (2004) How prevalent is functional alternative splicing in the human genome? Trends Genet. **20**, 68–71
27 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**, 2850–2855
28 Keren, H., Lev-Maor, G. and Ast, G. (2010) Alternative splicing and evolution, diversification, exon definition and function. Nat. Rev. Genet. **11**, 345–355
29 Stoilov, P., Daoud, R., Nayler, O. and Stamm, S. (2004) Human tra2-β1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. Hum. Mol. Genet. **13**, 509–524
30 Smith, C.W. and Valcarcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem. Sci. **25**, 381–388
31 Kamma, H., Portman, D.S. and Dreyfuss, G. (1995) Cell type-specific expression of hnRNP proteins. Exp. Cell Res. **221**, 187–196
32 Hofmann, Y., Lorson, C.L., Stamm, S., Androphy, E.J. and Wirth, B. (2000) Htra2-β1 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**, 9618–9623
33 Watermann, D.O., Tang, Y., Zur Hausen, A., Jager, M., Stamm, S. and Stickeler, E. (2006) Splicing factor Tra2-β1 is specifically induced in breast cancer and regulates alternative splicing of the *CD44* gene. Cancer Res. **66**, 4774–4780

Received 17 February 2012
doi:10.1042/BST20120036

OPEN ACCESS Freely available online

PLoS GENETICS

Identification of Evolutionarily Conserved Exons as Regulated Targets for the Splicing Activator Tra2β in Development

Sushma Grellscheid[1☯*], Caroline Dalgliesh[1☯], Markus Storbeck[2,3,4☯], Andrew Best[1], Yilei Liu[1¤a], Miriam Jakubik[2,3,4], Ylva Mende[2,3,4], Ingrid Ehrmann[1], Tomaz Curk[5], Kristina Rossbach[2,3,4], Cyril F. Bourgeois[6¤b], James Stévenin[6], David Grellscheid[7], Michael S. Jackson[1], Brunhilde Wirth[2,3,4], David J. Elliott[1*]

1 Institute of Genetic Medicine, Newcastle University, Newcastle, United Kingdom, **2** Institute of Human Genetics, University of Cologne, Cologne, Germany, **3** Institute of Genetics, University of Cologne, Cologne, Germany, **4** Center for Molecular Medicine, University of Cologne, Cologne, Germany, **5** University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia, **6** Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U 964, CNRS UMR 7104, Université de Strasbourg, Illkirch, France, **7** Institute for Particle Physics Phenomenology, Durham University, Durham, United Kingdom

Abstract

Alternative splicing amplifies the information content of the genome, creating multiple mRNA isoforms from single genes. The evolutionarily conserved splicing activator Tra2β (Sfrs10) is essential for mouse embryogenesis and implicated in spermatogenesis. Here we find that Tra2β is up-regulated as the mitotic stem cell containing population of male germ cells differentiate into meiotic and post-meiotic cells. Using CLIP coupled to deep sequencing, we found that Tra2β binds a high frequency of exons and identified specific G/A rich motifs as frequent targets. Significantly, for the first time we have analysed the splicing effect of *Sfrs10* depletion *in vivo* by generating a conditional neuronal-specific *Sfrs10* knock-out mouse (*Sfrs10*$^{fl/fl}$; *Nestin-Cre*$^{tg/+}$). This mouse has defects in brain development and allowed correlation of genuine physiologically Tra2β regulated exons. These belonged to a novel class which were longer than average size and importantly needed multiple cooperative Tra2β binding sites for efficient splicing activation, thus explaining the observed splicing defects in the knockout mice. Regulated exons included a cassette exon which produces a meiotic isoform of the Nasp histone chaperone that helps monitor DNA double-strand breaks. We also found a previously uncharacterised poison exon identifying a new pathway of feedback control between vertebrate Tra2 proteins. Both *Nasp-T* and the *Tra2a* poison exon are evolutionarily conserved, suggesting they might control fundamental developmental processes. Tra2β protein isoforms lacking the RRM were able to activate specific target exons indicating an additional functional role as a splicing co-activator. Significantly the N-terminal RS1 domain conserved between flies and humans was essential for the splicing activator function of Tra2β. Versions of Tra2β lacking this N-terminal RS1 domain potently repressed the same target exons activated by full-length Tra2β protein.

Citation: Grellscheid S, Dalgliesh C, Storbeck M, Best A, Liu Y, et al. (2011) Identification of Evolutionarily Conserved Exons as Regulated Targets for the Splicing Activator Tra2β in Development. PLoS Genet 7(12): e1002390. doi:10.1371/journal.pgen.1002390

Editor: Wendy A. Bickmore, Medical Research Council Human Genetics Unit, United Kingdom

Received May 11, 2011; **Accepted** October 5, 2011; **Published** December 15, 2011

Funding: This work was supported by the Wellcome Trust (Grant numbers WT080368MA and WT089225/Z/09/Z to DJE), the BBSRC (Grant numbers BB/D013917/1 and BB/I006923/1 to DJE), the Breast Cancer Campaign (to DJE), the Center of Molecular Medicine Cologne (CMMC) to BW [D7], the Deutsche Forschungsgemeinschaft to BW [Wi 945/12-3], European Commission FP7 EURASNET (to JS), and the Slovenian Research Agency to TC (grant number Z7-3665). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: David.Elliott@ncl.ac.uk (DJE); sushma@cantab.net (SG)

¤a Current address: Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States of America
¤b Current address: INSERM 1052 / CNRS UMR 5286, Centre de Recherche en Cancérologie de Lyon, Centre Léon Bérard, Lyon, France

☯ These authors contributed equally to this work.

Introduction

Almost all transcripts from genes encoding multiple exons are alternatively spliced, and correct patterns of alternative splicing are important for health and normal development [1,2,3]. Alternative splicing introduces new coding information into mRNAs, thereby increasing genome capacity to encode an expanded number of mRNAs and proteins from a finite number of genes [3]. Poison exons which introduce premature stop codons can also be alternatively spliced to target mRNAs for degradation through Nonsense Mediated Decay (NMD) [4,5,6,7,8].

Alternative splice events are controlled in part by *trans*- acting RNA binding proteins which help establish patterns of alternative splicing through deciphering a splicing code embedded within the pre-mRNA sequence [9,10,11]. Tra2 proteins bind directly to target exons thereby activating splicing inclusion [12], and have a modular organisation comprising a single central RNA recognition motif (RRM) which binds to target RNA sequences, flanked by

Author Summary

Alternative splicing amplifies the informational content of the genome, making multiple mRNA isoforms from single genes. Tra2 proteins bind and activate alternative exons, and in mice Tra2β is essential for embryonic development through unknown target RNAs. Here we report the first target exons that are physiologically regulated by Tra2β in developing mice. Normal activation of these regulated exons depends on multiple Tra2β binding sites, and significant mis-regulation of these exons is observed during mouse development when Tra2β is removed. As expected, Tra2β activates splicing of some target exons through direct RNA binding via its RNA Recognition Motif. Surprisingly, for some exons Tra2β can also activate splicing independent of direct RNA binding through two domains enriched in arginine and serine residues (called RS domains). The N-terminal RS1 domain of Tra2β is absolutely essential for splicing activation of physiological target exons, explaining why this domain is conserved between vertebrates and invertebrates. Surprisingly, Tra2β proteins without RS1 operate as splicing repressors, suggesting the possibility that endogenous Tra2β protein isoforms may differentially regulate the same target exons.

arginine-serine rich (RS1 and RS2) domains [13,14]. The N-terminal Tra2 RS1 domain is longer and contains more RS dipeptides than RS2. The reason for this unique modular organisation is unknown, but is conserved in vertebrate and invertebrate Tra2 proteins and different from the classical SR super-family which have a single C-terminal RS domain [15]. Also unlike classical SR proteins, Tra2 proteins do not restore splicing activity to S100 extracts [12].

A single Tra2 protein is conserved in fruit flies, where it is essential for spermatogenesis and sex determination [16]. There are two mammalian Tra2 proteins called Tra2α (encoded by the *Tra2a* gene on mouse chromosome 6) and Tra2β (encoded by the *Sfrs10* gene on mouse chromosome 16) which share 63% amino acid identity and similar RNA binding specificities [12]. NMR analyses have recently shown that the optimal core RNA target sequence for binding full length Tra2β protein is an AGAA motif, with each of the nucleotide residues being specifically recognized by the Tra2β RRM [17,18].

A key priority to understand the biological functions of Tra2β is to identify target RNAs which are functionally regulated within animal cells, and associated pathways of gene activity. Mice with ubiquitous deficiency of the *Sfrs10* gene die at around 7.5 to 8.5 days of gestation [19]. Splicing of some Tra2β candidate target exons have been investigated using minigenes, but recently a well known regulated splice target exon (*SMN2* exon 7) was found to have the same splicing pattern within wild type mice and $Smn^{-/-}$; $SMN2^{tg/tg}$; $Sfrs10^{-/-}$ mouse cells which do not express Tra2β protein [19]. These data suggest Tra2β is not the key protein regulating physiological inclusion of *SMN2* exon 7 within animal cells.

The *Sfrs10* gene itself is alternatively spliced to five mRNA isoforms encoding at least 2 protein isoforms [20,21,22]. The major isoform encodes full length Tra2β protein. Full length Tra2β protein regulates its own levels through activating splicing inclusion of a poison exon (exon 2) into a second mRNA isoform, preventing protein translation (Figure 1A) [22]. A third mRNA isoform encodes just the C-terminus of the protein (containing the RRM, glycine linker and the RS2 domain) giving rise to the protein isoform Tra2beta-3 or Tra2βΔRS1 [20,21,22]. No distinct function has been assigned to the Tra2βΔRS1 isoform compared to full length Tra2β [17], although this isoform is conserved in invertebrates so likely important. Tra2βΔRS1 expression is tissue specific in both flies and mammals, and is up-regulated by expression of Clk kinases and neural stimulation [20,21,22,23].

Male germ cell development is one of the few developmental pathways to continue into the adult. In the fly testis, Tra2 regulates splicing of *Exuperentia* and *Att* pre-mRNAs in male germ cells, as well as its own alternative splicing pathway [24,25]. Tra2β has been implicated in mammalian spermatogenesis through interaction with RBMY protein which is genetically deleted in some infertile men [26,27], and regulates the splicing of the human testis-specific HIPK3-T exon through a switch-like mechanism [28,29]. Given its important role in *Drosophila* spermatogenesis and established interactions with proteins implicated in human male fertility we predicted that Tra2β-regulated alternative splicing events would control fundamental pathways in mammalian male germ cell development. We have tested this prediction here using a transcriptome-wide approach.

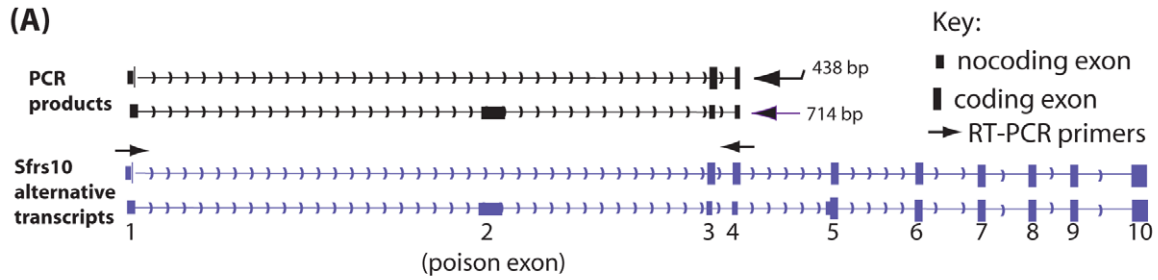
Results

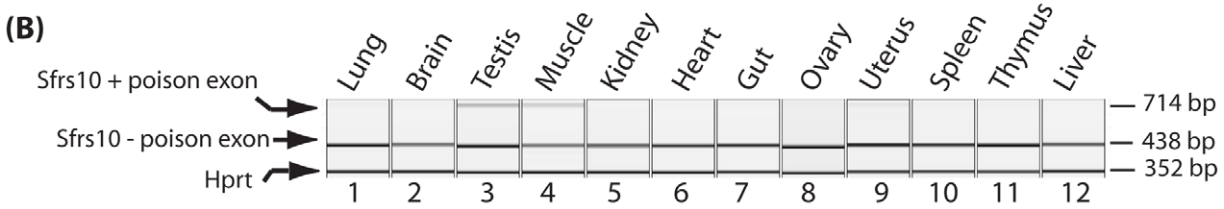
Tra2β is ubiquitously expressed but up-regulated at the onset of meiosis in male germ cells

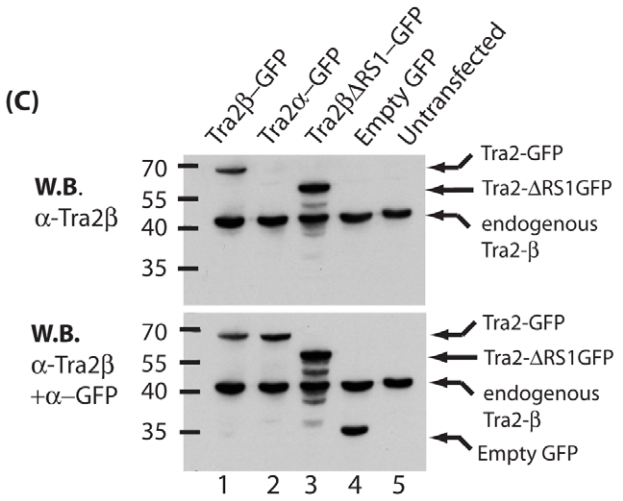
We analysed the expression of *Sfrs10* mRNA in different adult mouse (*Mus musculus*) tissues by RT-PCR using primers in exons 1 and 4. An RT-PCR product derived from *Sfrs10* mRNA in which exons 1 and 3 were directly spliced (skipping poison exon 2) was seen in every tissue indicating the *Sfrs10* gene is ubiquitously expressed (Figure 1A and 1B). A larger *Sfrs10* RT-PCR product made from mRNAs including poison exon 2 was detected at high levels in just two tissues, testis and muscle, indicating that expression of Tra2β is particularly tightly controlled in these tissues [22]. Similar levels of expression of *Hprt* mRNA were observed in each tissue by multiplex RT-PCR.

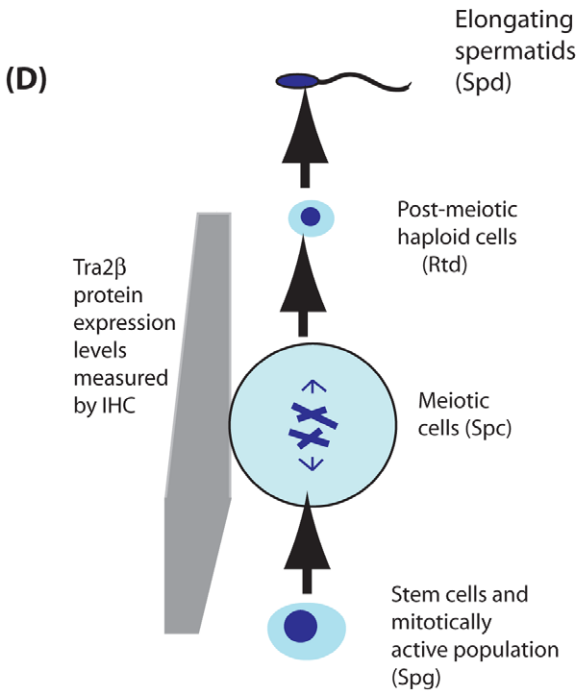
A polyclonal antiserum raised to Tra2β protein identified a single endogenous protein of around 40 KDa in both transfected and untransfected HEK293 cells corresponding in size to endogenous Tra2β (Figure 1C). A Tra2β-GFP fusion protein was additionally detected within transfected cells, but no cross-reaction was detected with a Tra2α-GFP fusion indicating high specificity of the antiserum. We were also able to detect a GFP-fusion protein containing Tra2βΔRS1, but not endogenous Tra2βΔRS1 protein suggesting that this particular isoform is expressed at low levels in these cells. Further probing of the same filter indicated that all the GFP fusion proteins were expressed at similar levels (Figure 1C, lower panel).

We used indirect immunohistochemistry to determine the cell type distribution of full length Tra2β in the adult testis (Figure 1D and 1E). Tra2β was detected as a nuclear protein (Figure 1E upper panel), and all staining was prevented by pre-incubation of the antisera with the immunising peptide (Figure 1E lower panel). Tra2β was most highly expressed during mouse male germ cell development at the meiotic stage in spermatocytes (abbreviated Spc), and afterwards in round spermatids (abbreviated Rtd). Less intense Tra2β staining was detected within spermatogonia which contain the mitotically active stem cell population. No immuno-staining was detected in elongating spermatids (abbreviated Spd). This regulated expression pattern predicts that Tra2β might play a role in regulating meiotic and post-meiotic exon inclusion during male germ cell development. Outside the germline, Tra2β protein expression was detected in Sertoli cells (indicated by red arrows on Figure 1E).


(A)
Key:
nocoding exon
coding exon
RT-PCR primers
PCR products
438 bp
714 bp
Sfrs10 alternative transcripts
1
2
3 4
5
6
7
8
9
10
(poison exon)


(B)
Lung
Brain
Testis
Muscle
Kidney
Heart
Gut
Ovary
Uterus
Spleen
Thymus
Liver
Sfrs10 + poison exon
Sfrs10 - poison exon
Hprt
714 bp
438 bp
352 bp
1 2 3 4 5 6 7 8 9 10 11 12


(C)
Tra2β–GFP
Tra2α–GFP
Tra2βΔRS1–GFP
Empty GFP
Untransfected
W.B.
α-Tra2β
70
55
40
35
Tra2-GFP
Tra2-ΔRS1GFP
endogenous Tra2-β
W.B.
α-Tra2β
+α–GFP
Empty GFP
1 2 3 4 5


(D)
Elongating spermatids (Spd)
Post-meiotic haploid cells (Rtd)
Tra2β protein expression levels measured by IHC
Meiotic cells (Spc)
Stem cells and mitotically active population (Spg)

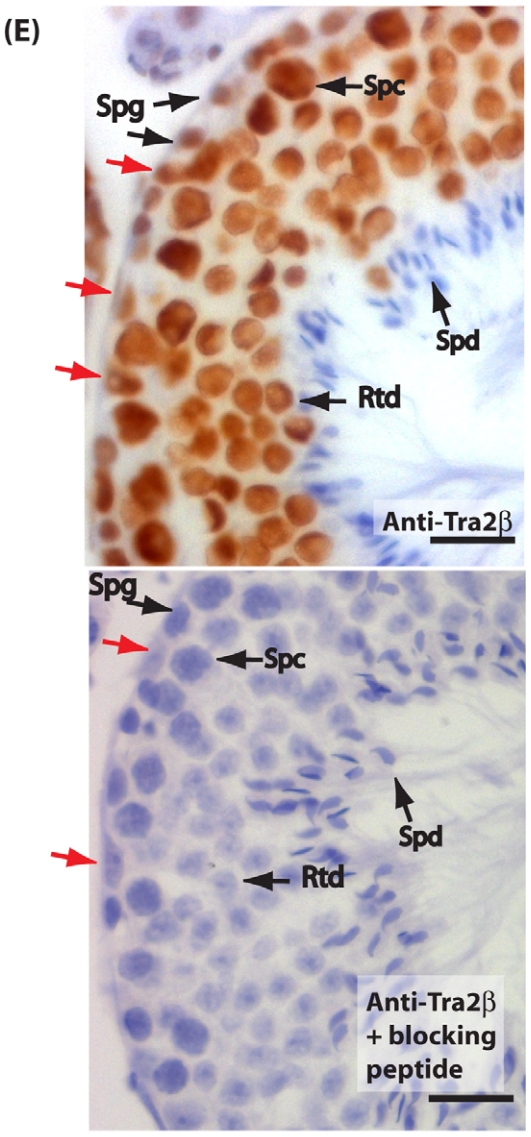

(E)
Spg
Spc
Spd
Rtd
Anti-Tra2β
Spg
Spc
Spd
Rtd
Anti-Tra2β + blocking peptide

Figure 1. Tra2β is a nuclear protein highly expressed in mouse germ cells. (A) Diagram showing *in silico* PCR of the mouse *Sfrs10* mRNA redrawn from the UCSC mouse genome browser [69]. Two different RT-PCR products are amplified using primers in exons 1 and 4. The smaller product (438 nucleotides) represents the amplified product when exon 1 is directly spliced to exon 3 and then exon 3 to exon 4 (upper *Sfrs10* mRNA isoform). The larger product (714 nucleotides) represents when the poison exon 2 is spliced resulting in the non-translated isoform Tra2β4 (lower *Sfrs10* mRNA isoform). (B) Capillary gel electrophoresis image showing levels levels of *Sfrs10* mRNA assayed by multiplex RT-PCR using RNA purified from adult mouse tissues. Primers were used for amplification complementary to exons 1 and 4 as described in (A) above. Within a multiplex RT-PCR, primers were included to detect *Hprt* as a parallel loading control to ensure equivalent amounts of RNA were used in each lane. (C) Immunoblotting experiment to confirm the specificity of the polyclonal antisera used for immunohistochemistry. HEK293 cells were transfected with plasmids expressing the indicated proteins. Proteins were then isolated and analysed by SDS-PAGE and Western blotting. The same blot was probed sequentially with an affinity purified antisera ab31353 raised against Tra2β (top panel) and then with a polyclonal specific for GFP to detect overall expression of each of the fusion proteins (lower panel). The ab31353 α-Tra2β antisera detected a single band in HEK293 cells corresponding to endogenous Tra2β protein, and in transfected cells additionally detected the Tra2β-GFP fusion protein and Tra2βΔRS1-GFP. No cross reaction with Tra2α-GFP was observed, indicating that this purified antisera is highly specific to Tra2β. (D) Flow chart summarising major developmental stages in male germ cell development. (E) Tra2β is a nuclear protein expressed during and after meiosis. Paraffin embedded adult mouse testis sections were stained with an affinity purified antibody raised against Tra2β (brown staining), and counterstained with haematoxylin (blue). Abbreviations: Spg –spermatogonia (mitotically active population which includes stem cells); Spc –spermatocyte (meiotic cells); Rtd –round spermatid (post-meiotic haploid cell); Spd –elongating spermatid (differentiating haploid cell with condensed nuclei). The scale bar is equivalent to 20 μm. The red arrows indicate Sertoli Cells. Based on these immunohistochemistry results, the Tra2β protein expression levels during mouse germ cell development are summarised also on the flow chart in part (D).
doi:10.1371/journal.pgen.1002390.g001

Tra2β primarily binds AGAA-rich target sequences in mouse germ cells

To identify endogenous cellular RNA targets for Tra2β we carried out high throughput sequencing cross linking immunoprecipitation (HITS-CLIP) [30]. Adult mouse testis cells were used according to published procedures (see methods for details) to retrieve an average tag length of 40 nucleotides. These recovered CLIP tags correspond to specific RNA sequences bound and subsequently cross-linked to endogenous Tra2β protein within the testis.

To identify frequent physiological Tra2β binding sites in mouse testis we searched for frequently occurring 6-mers in the retrieved CLIP tags, and normalised these to their background occurrence in the mouse genome and transcriptome using custom-written Python scripts (Table S1 and Table S2). Each of the most frequently recovered 6-mers was significantly enriched in the CLIP dataset compared to their representation in the mouse genome or mouse testis transcriptome. Strikingly, purine-rich sequences were preferentially recovered in our CLIP tags. In fact, 14 hexamers out of the top 30 recovered genome corrected hexamers in Table S1 have only purine residues, and 13 have only one pyrimidine. More specifically and consistent with the known RNA binding site for Tra2β [17,18], GAA-containing sequences were frequently observed. The distribution of GAA-containing 6-mers in the overall population of CLIP tags was visualised by plotting the genomic ranking of 6-mer recovery (X axis) against their representation in the CLIP population (Y axis) (Figure 2A: GAA-containing 6-mers are shown in red, with all other 6-mer sequences in blue). Of the 30 most frequently recovered 6-mers, 27 had a core GAA motif and the other 3 an AGA motif. The most frequent 6-mer (the AGAAGA motif, 10° on the X axis of Figure 2A -equivalent to 1) was found in almost 20% of the recovered CLIP tags. The ten most frequently recovered 6-mers were found in more than 40% of the CLIP tags.

Next we aligned full length CLIP tags to generate a transcriptome-wide consensus sequence. We anchored this line-up between CLIP tags using the trinucleotide GAA from the core binding motif which is essential for efficient RNA protein interactions [17] (Figure 2B). Within this consensus alignment, an A residue followed by a T residue (and less frequently a G residue) was usually found upstream of the GAA motif (position 1 in Figure 2B), consistent with reported *in vitro* RNA-protein binding data between the RRM of Tra2β and synthetic oligonucleotides [17]. Furthermore, a G residue (and less frequently an A residue) was preferentially selected at the position downstream of the GAA motif (position 5), and an A at the next nucleotide position downstream (position 6). This results in an extended AGAAGA consensus, in agreement with the sequence of the 3 top hexamers. Interestingly, when only a GAA triplet but not an AGAA core is present within a CLIP tag, 89% of the tags have a G residue immediately downstream (GAAG), consistent with the important contribution of the G5 residue for efficient binding of Tra2β to its natural RNA targets. No further strong sequence bias was noticed in the sequences upstream and downstream of the AGAAGA hexamer. A similar consensus was obtained previously for SRSF1 protein [31]. However since SRSF1 has 2 RRMs with different RNA binding capacities and only one RS domain, it is most likely that its global specificity of RNA recognition and binding are broader than that for Tra2β and also depends on other ESEs within its individual target exons.

Tra2β binds a high frequency of exonic sequences

To identify specific endogenous target transcripts CLIP tags were mapped onto the mouse genome sequence (a full bed file of Tra2β CLIP tags is provided as Dataset S1) [32]. Overall, the distribution of Tra2β CLIP tags was predominantly intragenic: Around 69% of Tra2β binding sites were located within protein coding genes, even though genes contribute just 25% of the genome (Figure 2C). Network analyses indicated the main functional properties associated withTra2β target transcripts were post-translational modification, the cell cycle, gene expression, RNA post-transcriptional modification and cell death (Figure 2D). Top physiological systems associated with Tra2β target transcripts included reproductive system and nervous system development, and there was significant enrichment of signalling pathways in the top detected pathways (Table S3). Most intragenic CLIP tags mapped to transcripts in the sense orientation, but 7.5% of retrieved CLIP tags were antisense to known annotated genes.

Only 1.3% of the mouse genome encodes exons (5′ UTR, ORF and 3′ UTR, based on mm9 annotation version ensembl59). For Tra2β some 29% of Tra2β CLIP tags mapped within exons of protein coding genes (Figure 2C) which indicates the presence of numerous Tra2β-specific target exons. Similar CLIP-based transcriptome-wide analyses found that the SR protein SRSF1 also frequently binds to exonic sequences, while Nova and PTB target sites are mainly intronic in distribution [30,31,33].

Non-exonic Tra2β binding sites were found within deep intronic regions, within locations annotated as intergenic and within noncoding RNAs (ncRNAs) [34]. Within ncRNAs Tra2β binding sites were found within the small subunit rRNA (also

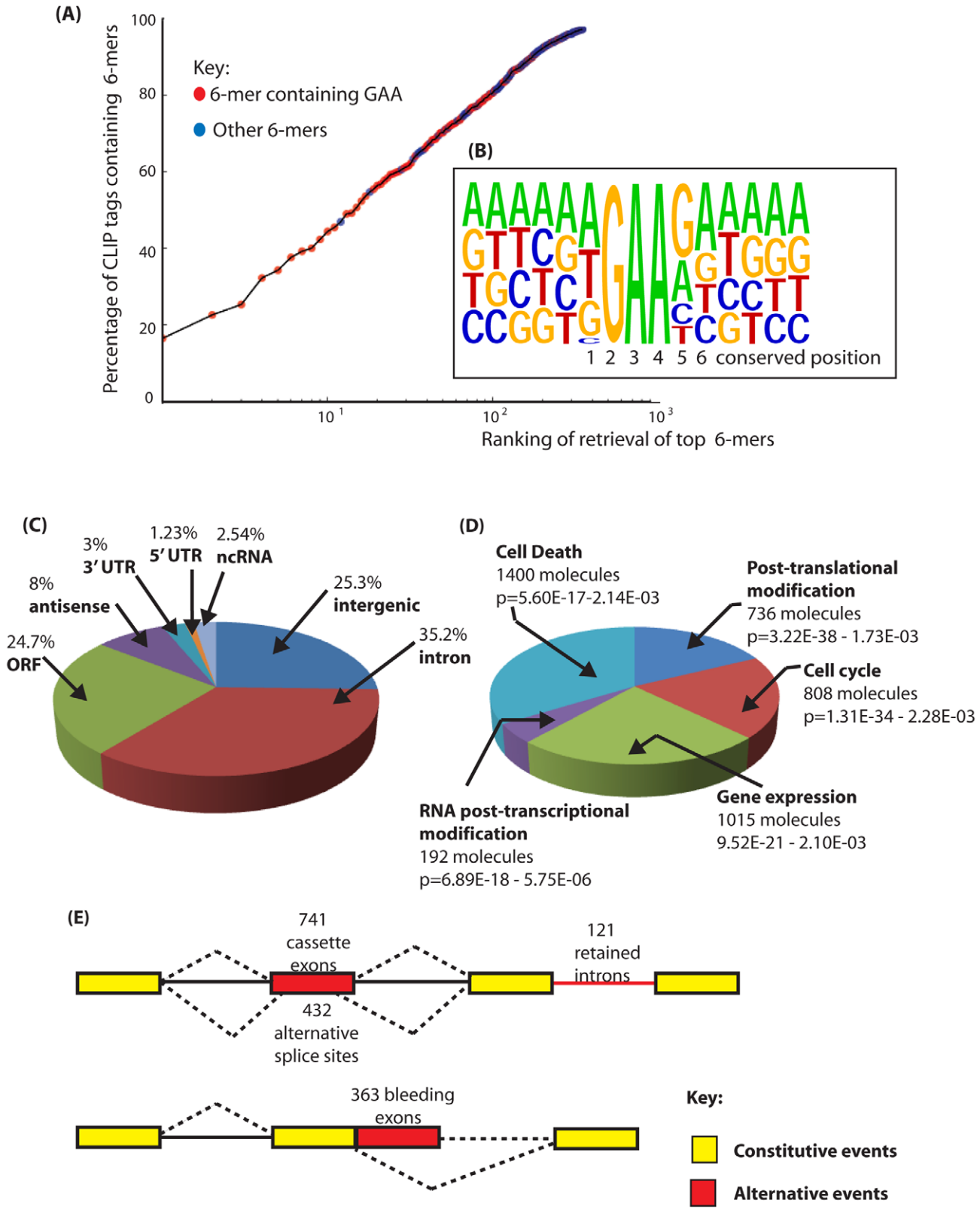


Figure 2. Identification of binding sites for Tra2β in the mouse transcriptome. (A) Nucleotide sequences enriched in the Tra2β CLIP tags are enriched in the core motif GAA. The percentage of CLIP tags was plotted against the order of retrieval of individual 6-mers on a logarithmic scale to identify the most frequently occurring 6-mer sequences within the CLIP tags. CLIP tag sequences which contain GAA are indicated in red. All other CLIP tags are shown in blue. (B) Consensus binding site for Tra2β derived from alignment of full length CLIP tags. The consensus was constructed by anchoring CLIP tags around GAA and then performing an alignment. The positions 1–6 which are particularly conserved are shown underneath and discussed in the main text. (C) Pie chart showing percentage of retrieved CLIP tags mapping to different inter- and intragenic locations within the

mouse transcriptome. (D) Summary of the top 5 molecular and cellular functions for Tra2β determined by Ingenuity Pathway Analysis. (E) Distribution of Tra2β binding sites relative to the different classes of alternative events annotated on the mouse genome. Alternative events are shown in red, and the constitutive events as yellow boxes (exons) or black lines (introns). Alternative events are annotated according to the UCSC genome browser track Alternative Events (URL: http://genome.ucsc.edu/cgi-bin/hgTrackUi?g = knownAlt&hgsid = 212031267).
doi:10.1371/journal.pgen.1002390.g002

identified as a binding site for SRSF1 [31]) and 7SK RNA. There were also Tra2β binding sites within the ncRNA *Malat1* which is known to be localised in nuclear splicing speckles enriched in pre-mRNA splicing components (*Malat1* is also bound by SRSF1 [31]), and within microRNAs. These identified targets suggest that Tra2β might in fact be a somewhat multifunctional post-transcriptional regulator. Similarly diverse classes of target RNA (including both coding and ncRNAs) have been identified for a number of other RNA binding proteins by HITS-CLIP [30,31,33,35,36].

Analysis of endogenous target exons indicate that isoforms of Tra2β can activate, co-activate, and repress exon inclusion

Tra2β bound to both constitutive and alternative exons and also to each different class of alternative events annotated on the mouse genome browser at UCSC. In particular, Tra2β binding sites mapped preferentially to cassette exons (this is also the most frequent class of alternative splicing event in metazoans [37]) (Figure 2E). To test for splicing regulation of these identified target exons by Tra2β, a panel of seven cassette exons with high numbers of mapped CLIP tags, together with flanking intronic sequences, were cloned into an exon trap vector (see Materials and Methods). The resulting minigenes were then transfected into HEK293 cells with expression constructs encoding either GFP, Tra2β-GFP, or GFP-tagged Tra2β deletion variants. Western blots indicated each of the GFP-fusion proteins were efficiently expressed in HEK293 cells (Figure 3A), although the fusion protein without the RS1 domain was expressed at higher levels.

Splicing patterns of pre-mRNAs were analysed using RT-PCR. We observed particularly strong splicing activation of a poison exon in the *Tra2a* gene in response to co-expression of Tra2β-GFP (Figure 3B). Ectopic expression of both Tra2α and Tra2β were equally able to activate splicing of the *Tra2a* poison exon indicating that these two proteins are functionally equivalent in this assay (Figure 3B, lanes 2 and 3). No splicing activation of the *Tra2a* poison exon was observed with either Tra2βΔRRM-GFP or GFP alone, indicating a requirement for RRM-dependent binding by full length Tra2β proteins for splicing activation (Figure 3B, lanes 1 and 4).

Full length Tra2β also mediated statistically significant splicing activation of a cassette exon annotated *Nasp-T* in the *Nasp* gene. Surprisingly, equally strong and highly statistically significant *Nasp-T* exon splicing activation was also observed in response to ectopic expression of Tra2βΔRRM-GFP protein (Figure 3C, lanes 2 and 3). Because of the high levels of splicing inclusion observed for the wild type *Nasp-T* exon at endogenous cellular concentrations of Tra2β (Figure 3C), we also repeated these experiments using a mutated exon which is less efficiently spliced (mutant M3+M4 –see below) and again observed significant splicing activation by Tra2βΔRRM-GFP protein (Figure 3D –in this case the effect of Tra2βΔRRM-GFP is clearer because of the lower levels of splicing inclusion of this mutated exon at endogenous cellular Tra2β protein concentrations). Together these data indicate that for some exons including Nasp-T, Tra2β can activate splicing through RRM independent interactions as well as being a direct splicing activator as previously described.

The *Sfrs10* locus encodes a second endogenous protein isoform called Tra2βΔRS1 [20,21,22] which lacks the RS1 domain. Surprisingly, after co-expression of a Tra2β-GFPΔRS1 protein isoform we observed significant splicing repression of both the *Tra2a* poison exon and *Nasp-T* exon (Figure 3B–3D) indicating that this protein isoform behaves as a potent splicing repressor, and of the same target exons recognised by full length Tra2β protein.

Two further exons, *Creb* exon 2 and *Lin28b* exon 2, did not detectably respond to ectopic expression of full length Tra2β or any of its derivatives (Figure 3G and 3H) and were already included at high levels in the absence of ectopically expressed Tra2β protein. No strong splicing repression of *Creb* exon 2 and *Lin28b* exon 2 was observed on co-expression of Tra2β-GFPΔRS1. Full length Tra2β weakly but significantly activated splicing of two other target exons, *Krba1* exon 9 and *Pank2* exon 3 (Figure 3E and 3F) and splicing of these exons was also not significantly repressed by Tra2β-GFPΔRS1 (compare lanes 1 and 3: notice slight repression which was not statistically significant). We also looked at two other exons which are spliced in the testis and which we independently characterised as being regulated by Tra2β. Minigene experiments indicated both the *Creb*γ and *Fabp9* exons [38,39] were moderately activated by Tra2β, and were also co-ordinately moderately repressed by the Tra2βΔRS1 isoform (Figure 3I and 3J, lanes 1 and 4). Taken together these data are consistent with full length Tra2β protein activating specific target exons, and the Tra2βΔRS1 protein isoform specifically repressing exons which are at least moderately to strongly activated by full length Tra2β, but not acting as a general repressor of cellular splicing.

Tra2β directly binds to target transcripts identified by CLIP, and binding efficiency correlates with splicing activity

We carried out further *in silico* and molecular analyses to correlate Tra2β binding with the observed patterns of exon regulation. We firstly looked for the occurrence of over-represented transcriptome-wide enriched 6-mer sequences (k-mers) [40] to identify putative Tra2β binding sites in the analysed target exons *in silico* (Figure S1). Both the *Nasp-T* and *Tra2a* poison exon had a high predicted content of 6-mers corresponding to putative Tra2β binding sites and consistent with their strong Tra2β regulation observed *in vitro*.

We then directly measured Tra2β binding affinities using Electromobility Shift Assays (EMSAs) (Figure 4: the positions of predicted binding sites within the RNA probes are shaded as in Table S1. Notice the dark green corresponds to the top 5 most frequently recovered 6-mers, and lighter shades of green correspond to less frequently recovered 6-mers). Both *Nasp-T* and *Tra2a* poison exon probes were very efficiently shifted by even very low concentrations of Tra2β protein (the *Nasp-T* probe was shifted into the well by only 50 ng of added Tra2β protein indicating formation of very large Tra2β protein-RNA complexes, and increasing molecular weight *Tra2a* RNA-protein complexes were observed with increasing concentrations of full length Tra2β protein).

A series of increased molecular weight complexes also formed on the *Creb*γ exon RNA probe (corresponding exon regulated *in*

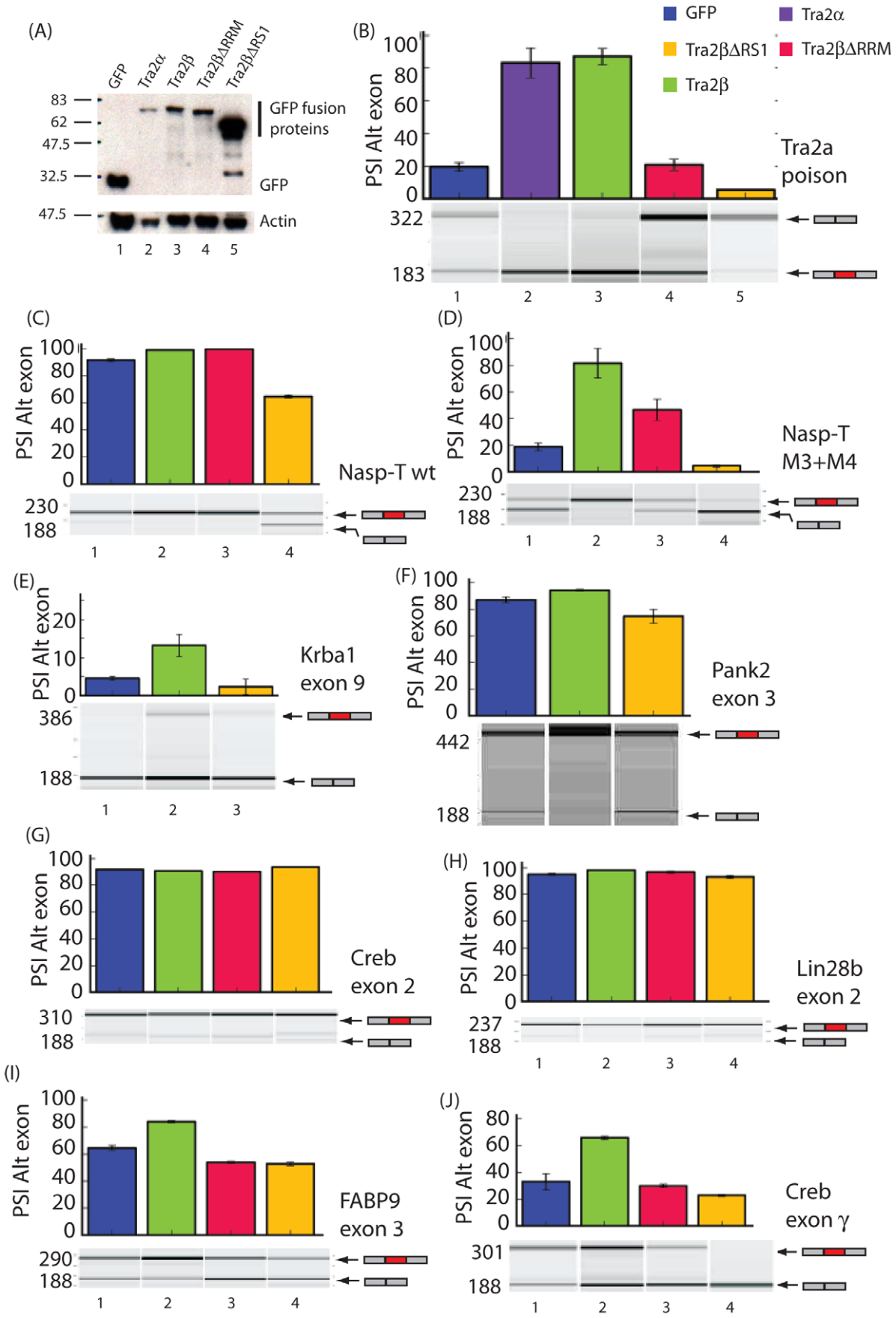

(A)
GFP
Tra2α
Tra2β
Tra2βΔRRM
Tra2βΔRS1
83
62
47.5
32.5
GFP fusion proteins
GFP
47.5
Actin
1 2 3 4 5
(B)
GFP
Tra2α
Tra2βΔRS1
Tra2βΔRRM
Tra2β
PSI Alt exon
Tra2a poison
322
183
(C)
Nasp-T wt
230
188
(D)
Nasp-T M3+M4
(E)
Krba1 exon 9
386
188
(F)
Pank2 exon 3
442
188
(G)
Creb exon 2
310
188
(H)
Lin28b exon 2
237
188
(I)
FABP9 exon 3
290
188
(J)
Creb exon γ
301
188

Figure 3. Different protein isoforms of Tra2β can act as specific splicing activators, co-activators, and repressors of a target exons identified by HITS-CLIP. (A) Efficient protein expression levels of different GFP fusion proteins used in these experiments (upper panel). Levels of actin were measured in parallel (lower panel). (B)–(J). Upper panels: Bar charts showing percentage splicing inclusion (PSI) of a panel of exons identified through HITS-CLIP in response to GFP and Tra2β-GFP fusion proteins. All data used to make the bar charts was from at least 3 biological replicates, and the error bars are shown as standard errors. Lower panels: Representative capillary gel electrophoresis image from each RT-PCR analysis. Probability (p) values were calculated using an independent two-sample T-test between the PSI levels for cells co-transfected with GFP and each of the different Tra2β-GFP constructs (* $p \leq 0.05$, ** $p \leq 0.01$).
doi:10.1371/journal.pgen.1002390.g003

cellulo by Tra2β) and on the *Krba1* RNA probe (weakly responsive *in cellulo* to Tra2β splicing activation). A single higher molecular weight complex formed on the *Lin28* probe (exon splicing not activated *in vitro* by Tra2β, and contains a single predicted Tra2β binding site). Much less efficient binding was observed for the non Tra2β-responsive *Creb* exon 2 (which formed a single molecular weight complex only with 200 ng added Tra2β protein, compared with 50 ng for the *Creb*γ probe).

The *Tra2a* poison and *Nasp-T* cassette exons are phylogenetically conserved and show high levels of splicing inclusion in mouse testis

An important measure of the functional importance of individual alternative splice events is evolutionary conservation [1,2,37,41,42]. Although many testis-specific exons are species-specific, phastcons analysis (which measures phylogenetic conservation of sequences on a scale of 0 to 1, with 1 being most conserved) indicated very high levels of phylogenetic conservation for the *Tra2a* poison exon along with flanking intronic sequences (Figure 5A–5C). Similar high levels of nucleotide conservation have been reported for poison exons in other genes encoding splicing regulator proteins including *Sfrs10* itself [4,5,22].

The *Tra2a* poison exon, which is 306 nucleotides long, introduces stop codons into the reading frame of the *Tra2a* mRNA which encodes Tra2α protein. Despite the lack of protein coding capacity, 48% of nucleotides within the *Tra2a* poison exon are in fact conserved in all vertebrates (Figure S2A: the nucleotide positions universally conserved in sequenced vertebrate genomes are shown in red). As a group, the 24 top most frequently recovered 6-mers from the entire transcriptome-wide screen were enriched in the nucleotide positions conserved between all vertebrates at levels much higher than would be expected by chance (Figure S2A, $p = 0.0075$, Fisher exact test: $p = 0.0003$, Chi Squared test). These data are consistent with maintenance of multiple Tra2β-binding sites within the *Tra2a* poison exon since the radiation of vertebrates. When analysed by RT-PCR, the *Tra2a* poison exon was found to be particularly strongly alternatively spliced in the testis, with zero or much lower levels in other adult tissues (Figure 5A–5C).

Phastcons analyses also showed the *Nasp-T* cassette exon, which is also particularly long at 975 nucleotides, has been conserved since the last common ancestor of all vertebrates (Figure 5D–5F). However neither the nucleotide or the peptide sequence encoded by *Nasp-T* are particularly highly conserved over the full length of the exon (Figure 5E).The *Nasp* gene encodes a histone chaperone essential for mouse development [43], and the *Nasp-T* exon introduces a peptide-encoding cassette exon generating a longer version of the Nasp protein. Similar to the *Tra2a* poison exon, 6-mers predicting Tra2β binding site sequences were found throughout the *Nasp-T* exon, and high frequency 6-mers mapped closely adjacent to CLIP tags (Figure S2B). Within mammalian *Nasp-T* exons multiple Tra2β binding sites have been conserved. Extremely high levels of *Nasp-T* exon inclusion were detected by RT-PCR in the testis and heart. In gut, muscle and ovary, the *Nasp-T* exon inclusion isoform was also preferentially included but in other tissues it was frequently skipped (Figure 5F).

Efficient splicing activation of the testis-specific Nasp-T by Tra2β depends on multiple Tra2β binding sites

To experimentally address the function of multiple Tra2β binding sites in *Nasp-T* we used a combination of *in silico* and experimental analyses, and focused on an upstream portion of the exon (from positions 117 to 271). Using octamers predictive of splicing enhancers and silencers [44,45,46], we firstly identified 3 strong putative ESEs (Exonic Splicing Enhancers, ESE1 to ESE3) which we selected for further analysis, as well as other putative moderate ESEs (Z score around 4) of which only one designated ESE4 was further studied (Figure 6A). Each of these putative ESEs directly overlapped with Tra2β binding sites initially identified through 6-mers derived from the transcriptome-wide CLIP analysis.

To experimentally test the need for individual Tra2β binding sites in splicing regulation, individual sites were mutated within the minigenes without creating Exonic Splicing Silencer (ESS) sequences (Figure 6A) [28], and the splicing effect monitored. Mutation of single Tra2β binding sites had only a minor effect on *Nasp-T* splicing inclusion at endogenous cellular concentrations of Tra2β. However, pre-mRNAs containing double mutations affecting Tra2β binding sites (M2+M3, M1+M2 and M3+M4) had strongly reduced *Nasp-T* exon splicing inclusion compared to their wild type counterparts at normal endogenous cellular concentrations of Tra2β (Figure 6B). Mutation of different Tra2β binding sites within *Nasp-T* also had distinct outcomes on exon inclusion, indicating underlying combinatorial effects between different patterns of Tra2β binding. In particular, mutant M3+M4 reduced exon inclusion levels to 20% of wild type at endogenous cellular levels of Tra2β, whereas double mutations comprising M2 and M3 reduced *Nasp-T* exon inclusion to just below 60% (Figure 6B).

Although they showed decreased exon inclusion at normal cellular concentrations of Tra2β, each of the double mutated Nasp-T exons gave at least 80% splicing inclusion after Tra2β protein was ectopically expressed. This suggested a requirement for higher levels of ectopic Tra2β protein for splicing inclusion. To test this, we co-transfected cells with minigenes containing either wild type *Nasp-T* exon or the M3+M4 mutant derivative, and a concentration gradient of Tra2β (Figure 6C). Splicing inclusion of the wild type *Nasp-T* exon was already 90% without over-expression of Tra2β and was maximal after co-transfection of no more than 30 ng Tra2β expressing plasmid. In contrast, levels of inclusion of the M3+M4 *NaspT* exon derivative increased more slowly over the whole concentration gradient, indicating decreased splicing sensitivity to Tra2β after removal of just two binding sites. This is particularly striking since the M3+M4 *NaspT* exon retains multiple other Tra2β binding sites (both experimentally confirmed sites in the case of ESEs 1–4, and further predicted sites throughout the exon shown in Figure S1). We used EMSAs to directly analyse RNA-protein interactions using both wild type and mutated versions of the *Nasp-T* RNA probe

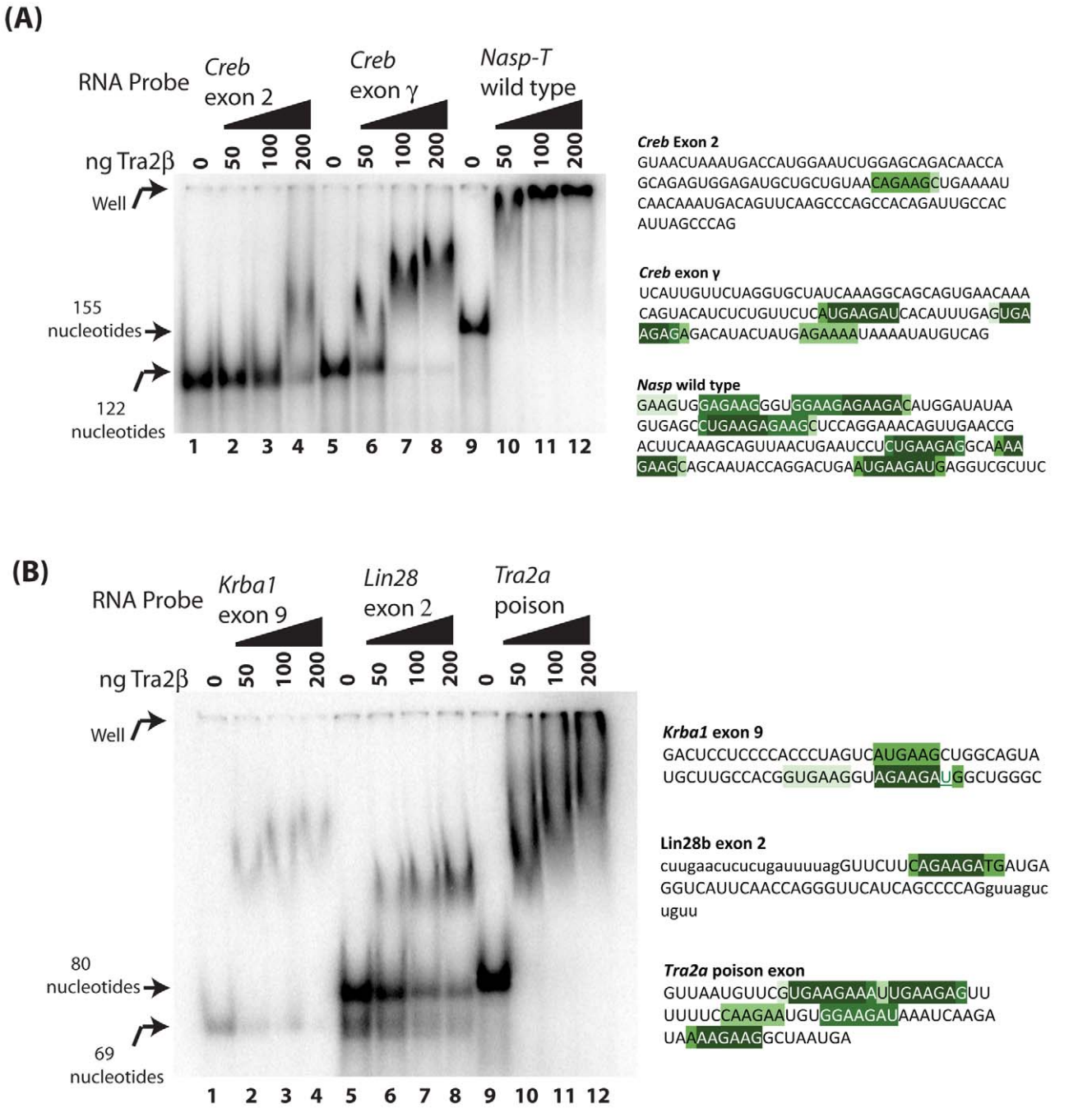


Figure 4. Tra2β CLIP targets bind to full length Tra2β protein. (A) EMSAs of *Creb* exon 2, *Creb* exon γ and the wild type *Nasp-T* exon. (B) EMSAs of *Krba1* exon 9, *Lin28* exon 2 and the *Tra2a* poison exon. Electrophoretic Mobility Shift Assays (EMSAs) were carried out with full length Tra2β protein and short radioactive RNA probes from pre-mRNAs identified by CLIP and which contained predicted Tra2β protein binding sites from the transcriptome-wide 6-mer analysis. The RNA probes are shown to the right of the gel panels, and the sequences are highlighted for different categories of 6-mers as in Table S1. Exon sequences are shown in upper case, and any flanking intron sequence in lower case (the *Lin28b* exon is very short).
doi:10.1371/journal.pgen.1002390.g004

(Figure 7). While wild type *Nasp-T* and the single mutant M2 RNA were efficiently shifted, the average size of the M3+M4 RNA-protein complex was only slightly smaller (the average size of the shifted complexes is indicated by a red asterisk on Figure 7). Hence even a moderate change in *in vitro* RNA-protein interactions translates to a detectable change in splicing inclusion within cells.

Levels of neuronal Tra2β protein are depleted in a Nestin-Cre mouse model and are functionally buffered by the *Sfrs10* poison exon

Mice with clearly reduced expression levels of *Sfrs10* would be a prerequisite to enable detection of altered splicing patterns in Tra2β- targeted transcripts identified by CLIP. Since ubiquitous *Sfrs10* deletion leads to embryonic lethality [19], we generated a

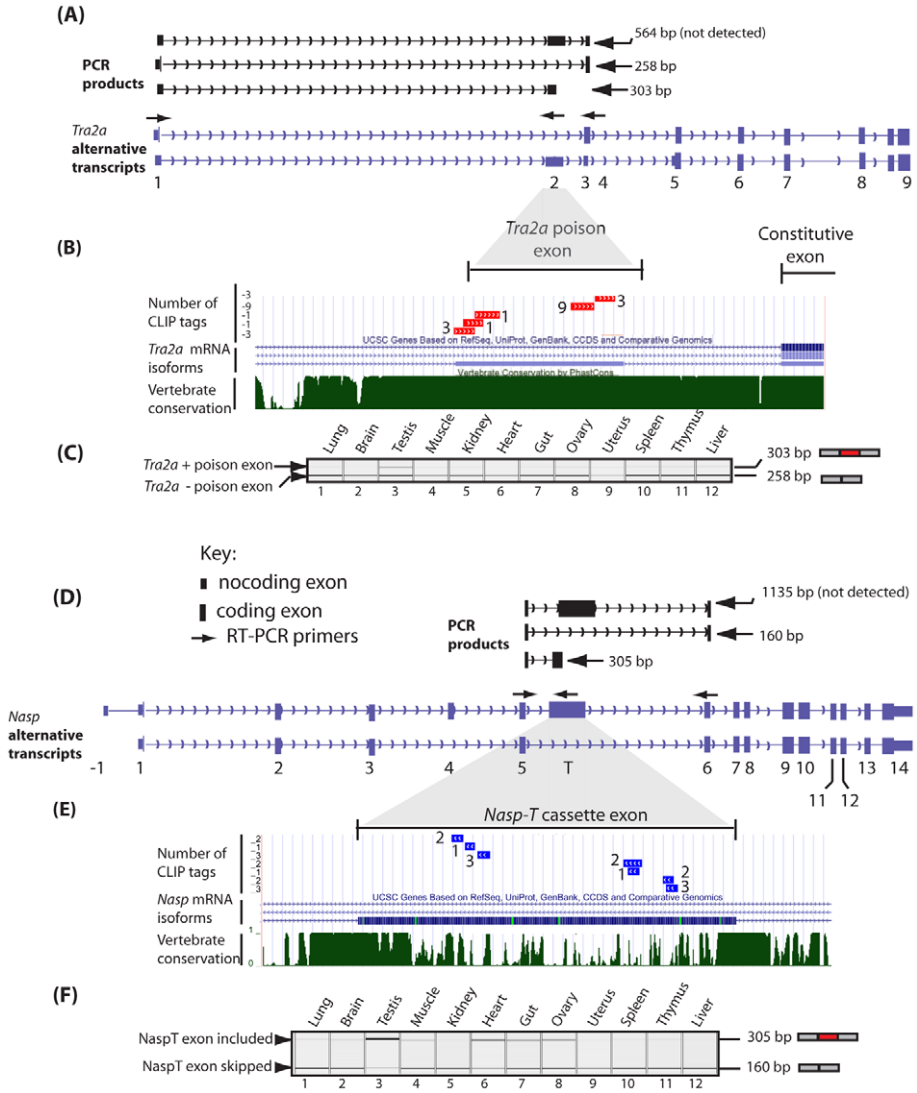


Figure 5. The *Tra2a* poison exon and *Nasp-T* cassette exon are conserved in vertebrates and spliced at high levels of inclusion in the mouse testis. (A) The structure of annotated alternative *Tra2a* transcripts (purple) and predicted PCR products (black) are shown above. (B) Comparative genomic analysis with supporting EST information confirm splicing inclusion of these *Tra2a* poison exons indicate they are found in vertebrates as distantly related as humans, mice, zebrafish and frog. (C) Expression of the *Tra2a* poison exon in different mouse tissues was monitored using RT-PCR (primers in exons 1 and 4) followed by capillary gel electrophoresis, and a representative capillary gel electrophoresis image is shown. (D) Multiple Tra2β CLIP tags mapped to a poison exon in the *Nasp-T* gene. The structure of annotated alternative *Nasp* transcripts (purple) and predicted PCR products (black) are shown above. (E) Underneath the Phastcons alignment of the *Nasp-T* exon from multiple vertebrates is shown. (F) Incorporation of the *Nasp-T* exon was monitored by RT-PCR and capillary gel electrophoresis. High levels of splicing inclusion were detected in the mouse testis, and lower levels of inclusion in other tissues. Multiple CLIP tags mapped to an evolutionarily conserved cassette exon in the *Nasp* gene. The Phastcons alignment of the *Nasp-T* exon from multiple vertebrates is shown. Phastcons analyses in parts (B) and (E) are shown as downloads from UCSC [69]. The key for both parts (A) and (D) are indicated in (D).
doi:10.1371/journal.pgen.1002390.g005

neuronal specific *Sfrs10*-depleted mouse by crossbreeding *Sfrs10*$^{fl/fl}$ mice with *Sfrs10*$^{fl/+}$ mice carrying the Nestin-Cre transgene (Nestin-Cre$^{tg/+}$). In *Sfrs10*$^{fl/fl}$; *Nestin-Cre*$^{tg/+}$ offspring the Cre recombinase would be specifically activated in neuronal and glial precursor cells from embryonic day 11 [47] to generate animals with a homozygous *Sfrs10* knockout in the developing central nervous system (CNS).

Homozygous neuronal *Sfrs10* mice died immediately after birth at postnatal day 1 (PND1) whereas heterozygote mice had normal lifespans. Neuronal specific *Sfrs10*-depleted embryos showed severe malformations of the brain including strong dilation of the third and lateral ventricles as well as degeneration of cortical structures (Figure 8A, right panel and data not shown) whereas heterozygous knockout mouse embryos (Sfrs10$^{fl/wt}$; Nestin-Cre^{tg}) had normal brain morphology (Figure 8A, left panel). This indicates Tra2β protein is functionally very important for brain development in the mouse. As the liquid filled ventricles make up the majority of the whole brain volume, the brain morphology is heavily altered and the proportion of intact tissue is heavily reduced. Immunohistochemical analysis of whole brain paraffin-embedded cross-sections showed strongly decreased expression of Tra2-β with some Tra2-β positive cell areas in the cortical plate zone (Figure 8A, right panel). These residual Tra2-β positive cells likely do not express *Cre* from the *Nestin* promoter and are likely of

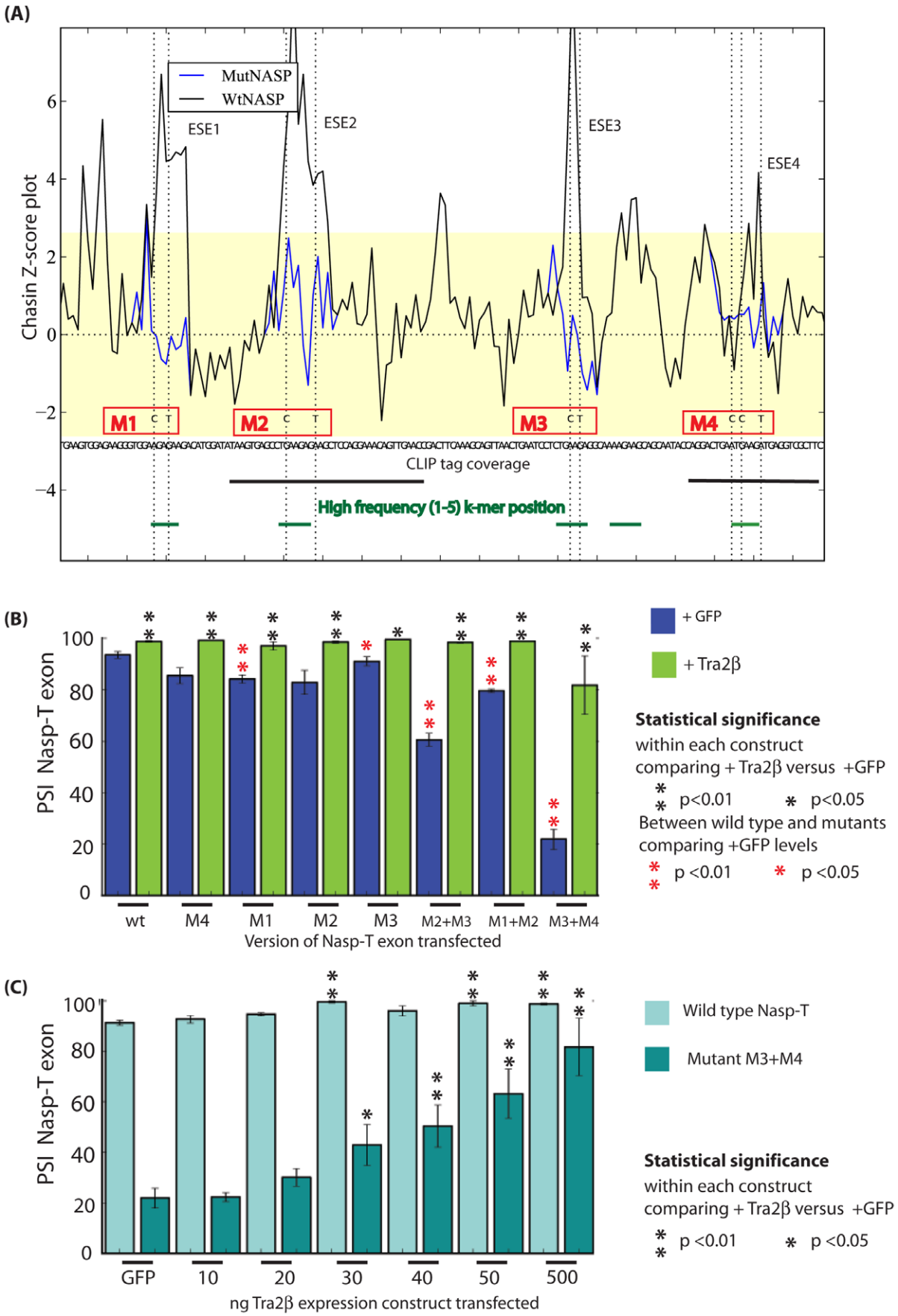

(A)
MutNASP
WtNASP
ESE1
ESE2
ESE3
ESE4
Chasin Z-score plot
M1
M2
M3
M4
CLIP tag coverage
High frequency (1-5) k-mer position
(B)
PSI Nasp-T exon
wt
M4
M1
M2
M3
M2+M3
M1+M2
M3+M4
Version of Nasp-T exon transfected
+ GFP
+ Tra2β
Statistical significance
within each construct
comparing + Tra2β versus +GFP
** p<0.01 * p<0.05
Between wild type and mutants
comparing +GFP levels
** p <0.01 * p <0.05
(C)
PSI Nasp-T exon
GFP
10
20
30
40
50
500
ng Tra2β expression construct transfected
Wild type Nasp-T
Mutant M3+M4
Statistical significance
within each construct
comparing + Tra2β versus +GFP
** p <0.01 * p <0.05

Figure 6. The splicing response to Tra2β is mediated through binding to four independent sites. (A) z-score plot predicting the splicing control sequences according to [45] in the upstream portion of the *Nasp-T* cassette exon. Investigated exonic regions with z-scores above the threshold value for exonic splicing enhancers are labelled ESE1–4. The z-score plots of the wild type *Nasp* exon is shown in black, superimposed with z-score plots for each of the point mutants which affected individual ESEs (shown as blue coloured lines, with the changed nucleotide indicated as a broken line). Individual mutants are shown as M1–M4. Local CLIP tag coverage is shown as black lines, and the relative positions of local 6-mers identified at a high frequency in the CLIP screen as green lines. (B) Effect of Tra2β on splicing inclusion of different *Nasp-T* cassette exons (wild type and mutants) co-expressed in HEK293 cells in the presence of endogenous Tra2β or with constant levels of Tra2β (500 ng, ectopically expressed). (C) Percentage exon inclusion of the wild type and *Nasp-T* exon derivative M3+M4 obtained after transfection of increasing levels of each of Tra2β. Error bars are shown as the standard error of the mean. Probability (p) values were calculated using an independent two-sample T-test between the PSI levels for cells co-transfected with GFP and Tra2β-GFP (black asterisks), or between endogenous PSI for each of the Nasp-T constructs at endogenous Tra2β concentrations (just transfected with GFP, red asterisks). P value scores are indicated as * p≤0.05 and **p≤0.01.
doi:10.1371/journal.pgen.1002390.g006

non-neuronal origin, or may represent mosaicism of *Nestin-Cre* expression. Furthermore, Western blots from whole brain also demonstrated a clear down-regulation of Tra2-β in neuronal specific *Sfrs10*-depleted embryos compared to controls and heterozygous knockout animals at 16.5 dpc (Figure 8B). In control animals the *Sfrs10* mRNA levels remained largely unchanged during development (16.5 dpc, 18.5 dpc and PND1) ($Sfrs10^{fl/fl}$ n = 10; $Sfrs10^{fl/+}$ n = 6; data not shown).

Expression analysis of whole brain RNA from neuronal *Sfrs10*-depleted embryos at 16.5 dpc and 18.5 dpc and mice at PND1 showed clearly reduced *Sfrs10* mRNA levels compared with brains of control littermates ($Sfrs10^{fl/fl}$, $Sfrs10^{fl/+}$ or $Sfrs10^{fl/+}$; $Nestin\text{-}Cre^{tg/+}$) (Figure 8C). Regardless of the developmental stage the majority of $Sfrs10^{fl/fl}$ pups exhibited somewhat reduced *Sfrs10* expression levels compared with heterozygously floxed mice, which suggested that the integration of the floxed allele has a slightly negative influence on *Sfrs10* expression. Therefore for statistical analysis the expression levels of splice isoforms of $Sfrs10^{fl/fl}$; $Nestin\text{-}Cre^{tg/+}$ mice were always compared with $Sfrs10^{fl/+}$ and not $Sfrs10^{fl/fl}$ mice.

Tra2-β regulates its own expression level via alternative splice regulation in an autoregulatory feedback-loop. Inclusion of poison exon 2 into *Sfrs10* transcripts introduces a premature stop codon which leads to a non-functional protein and thus a reduction in Tra2-β levels [22]. Isoform specific qRT-PCR indicated a highly significant down-regulation of both individual mRNA splice isoforms and total length *Sfrs10* mRNA in neuronal specific *Sfrs10*-depleted mice $Sfrs10^{fl/fl}Nestin\text{-}Cre^{tg/+}$) compared to controls at 16.5 dpc (Figure 8C). In contrast, in heterozygous knockout animals ($Sfrs10^{fl/+}Nestin\text{-}Cre^{tg/+}$) down-regulation of the functional isoform (− exon 2) was less effective than for the non-functional (+ exon 2) isoform indicating the involvement of the autoregulatory feedback loop which counteracts any decrease in functional Tra2β protein in neuronal cells.

Tra2β physiologically regulates splicing inclusion of the *Tra2a* poison and *Nasp-T* cassette exons in mouse brain development

We next set out to determine whether the *Tra2a* poison exon and *Nasp-T* cassette exon were true physiological target exons regulated by Tra2β *in vivo*. Correlating with an important regulatory role for Tra2β protein, splicing inclusion of the poison exon into the *Tra2a* mRNA was reduced 3-fold in neuronal *Sfrs10*-depleted mouse brains compared to controls at 16.5 dpc (Figure 8E). Surprisingly, this decrease in poison exon inclusion could not be detected at later developmental stages like 18.5 dpc or PND1 (data not shown).

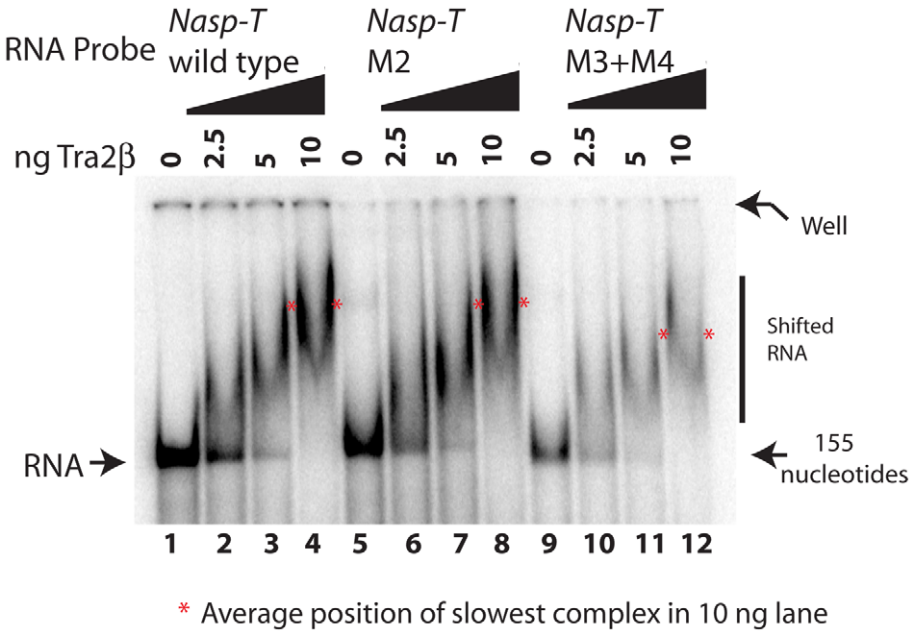
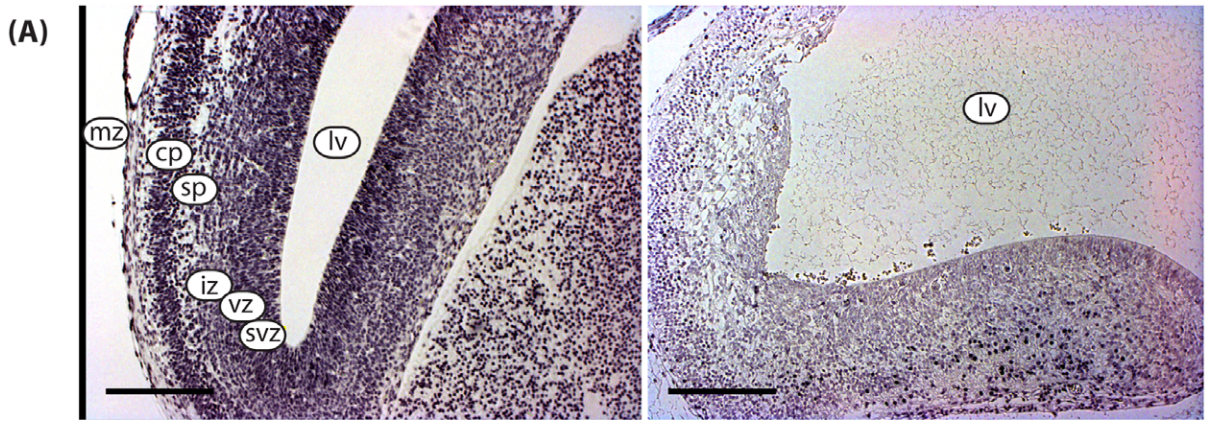
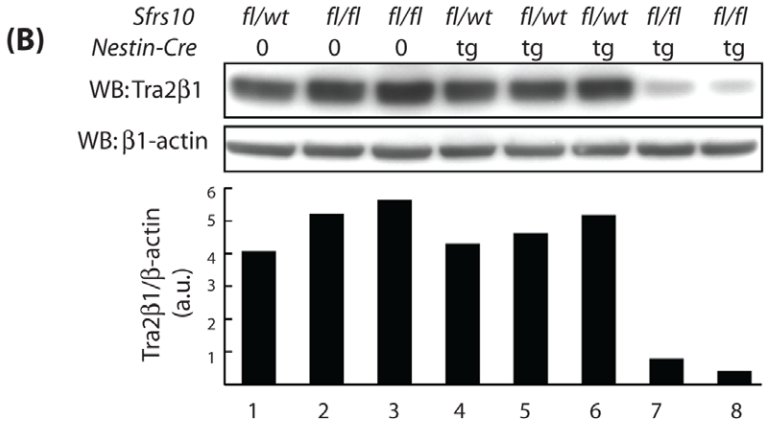
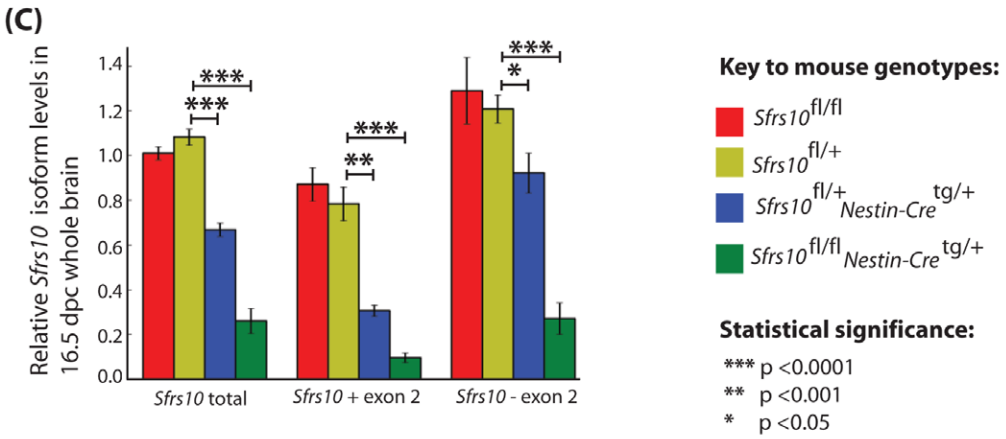
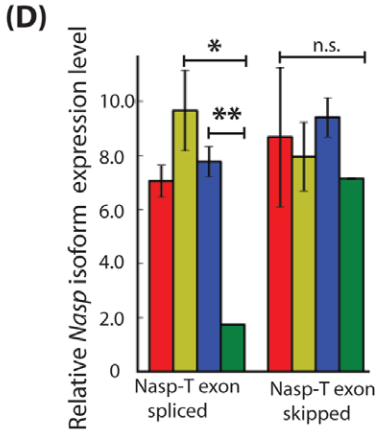


Figure 7. Point mutants in the Nasp-T exon within candidate Tra2β binding sites are still able to bind to Tra2β. RNA-protein interactions were monitored by EMSAs. The average position of the slowest migrating complex in the lane containing 10 ng of added Tra2β protein is indicated by an asterisk, and the RNA probes used were as in Figure 4 but containing the appropriate point mutation.
doi:10.1371/journal.pgen.1002390.g007


(A)
mz
cp
sp
iz
vz
svz
lv
lv


(B)
Sfrs10
Nestin-Cre
fl/wt 0
fl/fl 0
fl/fl 0
fl/wt tg
fl/wt tg
fl/wt tg
fl/fl tg
fl/fl tg
WB: Tra2β1
WB: β1-actin
Tra2β1/β-actin (a.u.)
1 2 3 4 5 6 7 8


(C)
Relative Sfrs10 isoform levels in 16.5 dpc whole brain
Sfrs10 total
Sfrs10 + exon 2
Sfrs10 - exon 2
Key to mouse genotypes:
Sfrs10fl/fl
Sfrs10fl/+
Sfrs10fl/+ Nestin-Cretg/+
Sfrs10fl/fl Nestin-Cretg/+
Statistical significance:
*** p <0.0001
** p <0.001
* p <0.05


(D)
Relative Nasp isoform expression level
Nasp-T exon spliced
Nasp-T exon skipped
n.s.

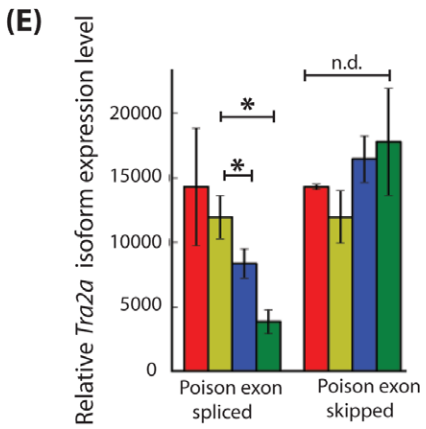

(E)
Relative Tra2a isoform expression level
Poison exon spliced
Poison exon skipped
n.d.

Figure 8. Tra2β protein levels are drastically reduced in the brains of neuronal specific *Sfrs10* knockout mice and correlate with defects in splicing of the *Nasp-T* cassette and *Tra2a* poison exon. (A) Whole brain sections derived from 16.5 dpc *Sfrs10*$^{fl/wt}$; Nestin-Cre^{tg} (left panel) and *Sfrs10*$^{fl/fl}$; Nestin-Cre^{tg} (right panel) stained with antibodies against Tra2β. Brains of heterozygous knockout animals (left panel) appear normal and Sfrs10 is expressed throughout all cortical layers. Brains of neuronal specific knockout animals (right panel) show a vast dilation of the lateral ventricles and disturbed cortical patterning. Tra2β expression is not detectable in the majority of intact tissue areas but is clearly retained in some cells of the cortical plate region. Scale bars represent 200 μm. Abbreviations are mz: marginal zone; cp: cortical plate zone; sp: subplate zone; iz: intermediate zone; svz: subventricular zone; vz: ventricular zone; lv: lateral ventricle. (B) Western blot analysis indicates that Tra2β expression is reduced in neuronal specific knockout mice. Proteins were isolated from whole brains of 16.5 dpc embryos and Tra2β was specifically detected by Western blotting. The Tra2β protein level is drastically reduced in *Sfrs10*$^{fl/fl}$; Nestin-Cre^{tg} animals compared to controls or heterozygous knockout animals. β-actin was used as a loading control. The relative levels are shown underneath as a bar chart (a.u. = arbitrary units). (C) Expression of the *Sfrs10* mRNA in different mouse genotypes used in this study. Levels of the *Sfrs10* mRNA isoforms in different mouse genotypes were independently measured by qRT-PCR from whole brain RNA isolated at 16.5 dpc (*Sfrs10*$^{fl/fl}$, n = 4; *Sfrs10*$^{fl/+}$, n = 5; *Sfrs10*$^{fl/+}$*; Nestin-Cre*$^{tg/+}$, n = 4; *Sfrs10*$^{fl/fl}$*; Nestin-Cre*$^{tg/+}$, n = 4). Levels of *Sfrs10* mRNA isoforms are consistent with use of the poison exon for autoregulation of transcript levels in vivo at 16.5 dpc. Isoform-specific qRT -PCR for *Sfrs10* on whole brain RNA revealed a coordinate downregulation of both the functional (−78%) and the non-functional (−88%) isoform in neuronal specific knockout animals at a highly significant level. The decrease of *Sfrs10* transcripts was also detectable in heterozygous knockout animals, in which the functional and non-functional isoform were decreased by 24% and 61%, respectively. (D) Splicing of the *Nasp-T* cassette exon is misregulated in *Sfrs10*$^{fl/fl}$*; Nestin-Cre*$^{tg/+}$,mice. Levels of the different mRNA isoforms were measured by qRT-PCR from brain RNA samples isolated at 16.5 dpc (*Sfrs10*$^{fl/fl}$, n = 2; *Sfrs10*$^{fl/+}$, n = 3; *Sfrs10*$^{fl/+}$*; Nestin*$^{-}$*Cre*$^{tg/+}$, n = 5; *Sfrs10*$^{fl/fl}$*; Nestin-Cre*$^{tg/+}$, n = 2). (E) Splicing of the Tra2a poison exon is misregulated in *Sfrs10*$^{fl/fl}$*; Nestin-Cre*$^{tg/+}$mice. Levels of the different mRNA isoforms were measured by qRT-PCR from brain RNA samples (*Sfrs10*$^{fl/fl}$, n = 2; *Sfrs10*$^{fl/+}$, n = 3; *Sfrs10*$^{fl/+}$*; Nestin-Cre*$^{tg/+}$, n = 5; *Sfrs10*$^{fl/fl}$*; Nestin-Cre*$^{tg/+}$, n = 2). (C–E) Error bars represent the s.e.m. Statistical significance was monitored using the T-test, and the significance values are as indicated.
doi:10.1371/journal.pgen.1002390.g008

To determine whether low Tra2β levels directly affect the splicing of the *Nasp-T* exon, qRT-PCR was carried out on whole brain RNA of 16.5 dpc and PND1 pups. The levels of the T-exon isoform of *Nasp* mRNA (*Nasp-T*) were 4-fold reduced in brains of neuronal *Sfrs10*-depleted mice compared to controls at 16.5 dpc (Figure 8D) and PND1 (data not shown). Given the 4-fold reduction of the *Nasp-T* isoform in *Sfrs10*-depleted tissue, we conclude that Tra2β protein is likely to be an important *in vivo* activator of *Nasp-T* exon inclusion during mouse development.

These data correlate a defect in splicing regulation of *Nasp-T* and *Tra2a* with *Sfrs10* depletion but do not necessarily imply a causal relationship, because of the differences in cell types present after *Sfrs10* depletion which result from the physiological importance of Tra2β for brain development. To address this further we compared overall patterns of expression of the *Nasp* and *Tra2a* genes in wild type and knockout mice, by quantifying levels of the somatic *Nasp* and *Tra2a* mRNA isoforms. Consistent with no significant changes in overall *Tra2a* gene expression resulting from changes in the cell type population of the knockout brains, no statistically significant changes in functional *Tra2a* or *Nasp* expression were seen when comparing brain RNA of *Sfrs10*$^{fl/+}$ mice with RNA of *Sfrs10*$^{fl/fl}$*; Nestin-Cre*$^{tg/+}$ mice (Figure 8D and 8E). These results are consistent with essentially similar patterns of *Nasp* and *Tra2a* gene expression in the mutant and wild type brains despite any differences in cellular composition, while in contrast the Tra2β-regulated splice isoforms from these same genes are very different between the wild type and mutant mice.

Discussion

Here we have identified (for the first time to the best of our knowledge) physiological target exons regulated by Tra2β during mouse development. Identification is based on the criteria of *in vivo* cross-linking of endogenous RNAs and proteins, *in cellulo* experiments using transfected minigenes and proteins, RNA-protein interaction assays and genetic analysis using a newly derived conditional mouse strain which does not express Tra2β protein in neurons and has significant abnormalities in brain development. Our analyses reveal important pathways regulated by Tra2β protein *in vivo* which likely contribute both to prenatal death in *Sfrs10*$^{-/-}$ embryos and also to normal germ cell development [19]. Nasp protein is a histone chaperone required for nuclear import of histones at the G1-S phase transition of the cell cycle, and is essential for cell proliferation and embryonic survival [43]. Nasp functions in chromatin remodelling after DNA repair, and links chromatin remodelling to the cell cycle machinery after S phase [48]. The T exon is also spliced in embryos, and within the testis alternative splicing inclusion of the *Nasp-T* cassette exon generates the testis-enriched tNASP protein isoform. Timing of tNASP protein expression during male adult germ cell development [48,49] exactly parallels the expression of Tra2β protein. The tNASP protein isoform localises to the synaptonemal complex of meiotic chromosomes where it may help monitor double strand DNA break repair [43,48,50].

Tra2α and Tra2β are very similar proteins, and are interchangeable in our *in cellulo* splicing assays. Tra2β protein helps regulate overall Tra2 protein levels through both activating splicing inclusion of a poison exon into its own *Sfrs10* mRNA, and also activating splicing inclusion of a poison exon into *Tra2a* mRNA which encodes Tra2α protein. *In vivo* experiments described here show that reduced inclusion of the poison exon does indeed help buffer the effect of decreased gene dosage in *Sfrs10* heterozygote mice. However, down-regulation of *Tra2a* poison exon inclusion in *Sfrs10*$^{-/-}$ cells does not lead to an increase in *Tra2a* mRNA levels sufficient to restore splicing patterns of Tra2β target exons, perhaps suggestive of unique functions for the Tra2α and Tra2β proteins. In flies, auto-regulation of splicing by Tra2 protein of its own pre-mRNA has been shown to be critical for spermatogenesis, indicating that it might be a highly conserved feature for germ cells to tightly maintain expression levels of this class of splicing regulator [24,25,51]. Since Tra2α regulates *Tra2a* poison exon *in cellulo*, it is likely that it also autoregulates its own mRNA levels *in vivo* through activation of this same poison exon.

An important current question is how RNA binding proteins like Tra2β achieve sequence specificity in target sequence selection despite having fairly short target sequences [15]. Here we have found a short consensus binding motif for Tra2β (AGAAGA, Figure 2A) which matches perfectly with specific motifs obtained both by classical SELEX analysis [12] and from identification of Tra2β specific ESEs in various genes [22,29,52,53,54,55,56,57]. Parallel genome-wide mapping showed that Tra2β primarily binds to exonic sequences. An explanation for exonic enrichment despite the short binding site would be if Tra2β binds to exons cooperatively with adjacent exonic RNA binding proteins. In the case of *SMN2* exon 7, the Tra2β binding site is flanked by cooperative binding sites for SRp30c and hnRNP G [17,53,58]. For *Nasp-T* and *Tra2a* there are instead arrays of exonic Tra2β

binding sites. Removal of more than one binding site negatively affects exon activation by Tra2β, indicating for *Nasp-T* and *Tra2a* adjacent binding and assembly of homotypic Tra2β protein activation complexes play important roles in splicing activation.

A model of splicing activation for the *Nasp-T* and the *Tra2a* poison exon which depends largely on sole binding of Tra2β protein might explain why these exons are particularly sensitive to depletion of Tra2β *in vivo* compared with *SMN2* exon 7 (splicing of which is not affected after deletion of *Sfrs10*, and which has a single Tra2β binding site, Figure S1). The human testis-specific HIPK3-T exon [50] also requires multiple Tra2β binding sites to enable splicing activation of a weak 5′ splice site *in vitro* [28], and the *Sfrs10* poison exon also has multiple Tra2β binding sites [22]. Other than *Tra2a* and *Nasp-T*, the remaining target exons we analysed using minigenes here have less dense coverage of Tra2β binding sites (Figure S1). These remaining exons also responded less robustly to Tra2β protein expression *in vitro* in transfected cells, and it is likely that RNA binding proteins other than Tra2β might also be more important for their splicing regulation *in vivo*.

We also found that full lengthTra2β protein activates splicing of the *Nasp-T* exon at a lower level through its RS1 and RS2 domains only (i.e. without the RRM and so without direct RNA binding). Mechanistically the RS domains of Tra2β might activate splicing by helping assemble other RS-domain containing splicing regulators and components of the spliceosome into functional splicing complexes. Although both RS domains could co-activate splicing when present together, removal of the RS1 domain completely disabled Tra2β-mediated splicing activation of the physiological target exons identified here. The observed functional importance of RS1 provides a mechanistic explanation why this N-terminal RS domain structure is maintained for Tra2 proteins in both vertebrates and invertebrates. Surprisingly Tra2β molecules without the RS1 domain were not just neutral for splicing inclusion *in cellulo*, but for some exons actually functioned as potent splicing repressors. Since the Tra2βΔRS1 isoform contains a functional RRM sequence, splicing repression could be due to competitive inhibition through this shorter Tra2β protein binding to the same RNA targets, but then being unable to assemble functional splicing complexes with other Tra2β proteins in the absence of the RS1 domain. Detection of such a competitive inhibitory function might have been helped by the increased levels of the Tra2βΔRS1 isoform expressed in our experiments. *In vivo*, the Tra2β-3 protein which lacks the N-terminal RS1 domain might also operate as a natural splicing repressor isoform [20,21,22], depending on its level of expression being enough in specific cell types or tissues. Tra2βΔRS1 actually activates *SMN2* exon 7 rather than being a repressor as seen for the physiological target exons we describe in this report [17]. Although the biology of *SMN2* exon 7 has been an area of controversy in the literature [59,60], a possible mechanistic explanation for this difference might be if Tra2β binding to *SMN2* exon 7 blocked the action of an adjacent Exonic Splicing Silencer, rather than directly activating splicing by itself.

Our analysis shows that the RNA targets identified for Tra2β in developing adult germ cells can predict patterns of splicing regulation by Tra2β in the developing brain. However, our data further suggest that splicing regulation by Tra2β is temporally restricted during development and also differentially regulated between various Tra2β targets. This is highlighted by *Tra2a* poison-exon splicing, which is affected by neuronal specific *Sfrs10* knockout only at a defined developmental stage, while *Nasp-T* exon inclusion is perturbed by *Sfrs10* knockout in all analyzed situations. Both the *Nasp-T* and the *Tra2a* poison exon are biologically important: they are conserved in all vertebrates for which genome sequences are available; have known functional roles; and like other phylogenetically conserved exons are spliced at high levels in at least some tissues [4,37,41]. The tNASP protein has been identified immunologically after the leptotene stage of meiosis in both rabbits and mice, indicating that this exon is meiotically expressed in both species [48,49]. In addition, although a high frequency of alternative splicing events in the testis are species-specific [61], the high conservation of binding sites in the *Tra2a* poison-exon suggests regulation by Tra2β has been conserved since the radiation of vertebrates. Overall our data indicate maintenance of ancient patterns of splicing regulation controlled by this RNA binding protein, consistent with its observed key role in development [19].

Materials and Methods

Detection of RNA and proteins in different mouse tissues

mRNA levels were detected in total RNA isolated from different mouse tissues using RT-PCR and standard conditions. RT-PCR products were analysed both by normal agarose gel electrophoresis (not shown) and capillary gel electrophoresis [62,63]. *Sfrs10* primers were specific to sequences in exons 1 and 4 respectively (5′-GAGCTCCTCGCAAAAGTGTG-3′ and 5′-CAACAT-GACGCCTTCGAGTA-3′). Tra2β protein was detected using immunohistochemistry in the mouse brain as previously described [64] and in the mouse testis using Abcam polyclonal Tra2β antibody ab31353 [28] as previously described [26].

Different *Tra2a* mRNA isoforms mRNA were detected by multiplex RT-PCR using Tra2aF (5′-GTTGTAGCCGTCGC-CTTC T-3′), Tra2aB (5′-TGGGATTCAGAATGTTTGGA-3′) and Tra2a poison (5′-TTCAAGTGCTTCTATCTGACCAA-3′). Different *Nasp-T* mRNA isoforms were detected by RT-PCR using Nasp-TF (5′-AATGGAGTGTTGGGAAATGC-3′), Nasp-TB (5′-TTGGTGTTTCTTCAGCCTTG-3′) and Nasp-TC (5′-TGCTTTGAAGTCGGTTCAACT-3′).

Hprt expression was detected using primers HrptF (5′-CCTGCTGGATTACATTAAAGCACTG-3′) and HprtR (5′-GTCAAGGGCATATCCAACAACAAAC-3′).

HITS-CLIP

HITS-CLIP was performed as previously described [30] using an antibody specific to Tra2β [65]. The specificity of the antibody to Tra2β was confirmed by the experiment shown in Figure S3, as well as the additional characterization already described [65]. In short, for the CLIP analysis mouse testis was sheared in PBS and UV crosslinked. After lysis, the whole lysate was treated with DNase and RNase, followed by radiolabelling and linker ligation. After immunoprecipitation with purified antisera specific to Tra2β [65], RNA bound Tra2β was separated on SDS-PAGE. A thin band at the size of 55 kDa (Tra2β migrates at around 40 kDa and MW of 50 nt RNA is about 15 kDa) was cut out and subject to protein digestion. RNA was recovered and subject to sequencing which was carried out on the Newcastle University Roche 454 GS-FLX platform. Mapping was done with Bowtie [66], allowing for two mismatches (parameter –v 2). 297070 reads were processed, of which 177457 (59.74%) aligned successfully to the mouse genome (Mm9). 74476 (25.07%) failed to align, and 45137 (15.19%) reads were suppressed due to multiple hits on the mouse genome. K-mer analysis was carried out using custom written scripts in Python. Briefly, we calculated the frequency of occurance of each possible 6-mer sequence in the following: our CLIP dataset, the mouse genome (mm9) and in the mouse testis transcriptome (http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc = GSM475281). The genome and transcriptome corrected frequencies were obtained

by subtracting the background (genome and transcriptome frequencies respectively) from the signal (frequency in CLIP dataset). CLIP reads were filtered to remove duplicates including overlapping reads. Statistical significance was determined using a Chi-squared test. The weblogo was derived from tags containing a GAA sequence by analysing the sequence composition surrounding the fixed sequence, using custom written scripts to generate an input for the freely available program weblogo (http://weblogo.berkeley.edu/).

Generation of neuronal specific *Sfrs10* knock-out mice for in *vivo* splicing analysis

In our *in vivo* splicing study we utilized a previously established *Sfrs10* mouse model on pure C57BL/6 background as described [19]. Genotyping was performed using tail DNA according to established protocols [19]. To induce a conditional *Sfrs10* knock-out in the central nervous system we crossbred *Sfrs10*$^{fl/fl}$ mice with a *Nestin-Cre*$^{tg/+}$ mouse line. These mice express Cre recombinase under control of the rat nestin (Nes) promoter and enhancer [47]. Therefore Cre recombinase is expressed in neuronal and glia cell precursors from embryonic day 11 as well as in neurogenic areas of the adult brain [47,67]. For our analyses the presence of the *Nestin* transgene was determined by a standard PCR using the oligonucleotides 5′–CGCTTCCGCTGGGTCACTGTCG-3′ (forward) and 5′–TCGTTGCATCGACCGGTAATGCAGGC-3′ (reverse) at an annealing temperature of 58°C producing a 300 bp amplicon.

Quantitative analysis of *Sfrs10* expression and Tra2β targeted transcripts

Whole brain RNA was isolated from 16.5 dpc, 18.5 dpc and PND1 mice using the RNeasy Lipid Tissue Mini Kit (Qiagen, Hilden, Germany). RNA concentration was determined by Quant-iT RiboGreen RNA Reagent and Kit (Invitrogen, Darmstadt, Germany) and equal amounts of RNA were used for first strand cDNA synthesis utilizing the QuantiTect reverse Transcription Kit (Qiagen, Hilden, Germany). Quantitative real-time PCR was carried out using the Roche LC FastStart DNA Master SYBR green Kit (Roche, Mannheim, Germany) on the Roche LightCycler 1.5. For realtime quantification total *Sfrs10* transcripts were amplified using the oligonucleotides 5′-TAGAAGGCATTATACAAG-3′ (forward) and 5′′-CTCAACCCAAACACGC-3′ (reverse) at 3 mM $MgCl_2$ and an annealing temperature of 63°C producing a 186 bp bp amplicon. To quantify *Sfrs10* isoforms specifically we used the oligonucleotides 5′-AGAACTACGGCGAGCGGGAATC-3′ (forward) and 5′-CCTTGTATAATGCCTTCTAGAACTTCTTC-3′ (reverse) for the functional isoform and 5′-GAACTACGGCGAGCGGGTTAATG-3′ (forward) and 5′-CAAGTGGGACTTCTGGTCTGATAATTAGC-3′ (reverse) for the non-functional isoform. Both were run at annealing temperatures of 64°C resulting in amplicons of 191 bp and 161 bp, respectively. For the quantification of different target splice variants single isoforms were amplified separately. For the *Nasp-T* exon containing isoform the oligonucleotides 5′-GGAGTGCATGTAGAAGAGG-3′ (forward) and 5′-CGTCATAAACCTGTTCTCTC-3′ (reverse) were used at 1 mM $MgCl_2$ and annealing at 65°C producing a 115 bp amplicon. The somatic isoform of *Nasp* was amplified using 5′-AATGGAGTGTTGGGAAATGC-3′ (forward) and 5′-CTGAGCCTTCAGTTTCATCTAC-3′ (reverse) at 3 mM $MgCl_2$, 62°C annealing while producing a product of 118 bp length. The functional *Tra2a* transcript was amplified using the oligonucleotides 5′-GTTGTAGCCGTCGCCTTCT-3′ (forward) and 5′-GAGACTCTCTGCCCTCGAAG-3′ (reverse) at 3 mM $MgCl_2$ and 66°C annealing resulting in a 155 bp product. For the poison exon-containing isoform we used the same forward oligonucleotide as for the functional isoform and 5′-CTTGATTTATCTTCCACATTCTTGG-3′ (reverse) at 3 mM $MgCl_2$ and 64°C annealing producing a 206 bp amplicon. All quantification data was normalized against *Gapdh*. Amplification was performed using the oligonucleotides 5′-GGCTGCCCAGAACATCATCC-3′ (forward) and 5′-GTCATCATACTTGGCAGGTTTCTC-3′ (reverse) at 3 mM $MgCl_2$ and 63°C annealing producing a 169 bp amplicon. Agarose gel electrophoresis and basic melting curve analysis was performed to confirm PCR product specificity. For quantification a dilution series of cDNA was used to generate a standard curve for each isoform. Therefore the cycle threshold was plotted versus the logarithm of the concentration and the standard curve was determined by linear regression. This curve was then utilized to calculate the template concentration of unknown samples. All samples were measured in duplicates. Individuals of a genotype were averaged using the arithmetic mean. Fluctuations are displayed by the standard error of the mean, and these are indicated on the bar charts by error bars. The significance of differences between genotypes was verified using student's t-test.

Minigene splicing experiments

Candidate alternatively spliced exons identified by HITS-CLIP and approximately 240 nucleotides of intronic flanking region at each end were amplified from mouse genomic DNA with the primer sequences given below. PCR products were digested with the appropriate restriction enzyme and cloned into the *Mfe*1 site in pXJ41 [68], which is exactly midway through the 757 nucleotide rabbit β-globin intron 2. PCR products were made using the following primers:

Krba1L: 5′-AAAAAAAAGAATTCtggggatcctagcaggtaca -3′

Krba1R: 5′-AAAAAAAAGAATTCccaaggatgtgataagcagga -3′

CREB2U: 5′-AAAAAAAACAATTGgggaccattcctcatttcct -3′

CREB2D: 5′-AAAAAAAACAATTGaaggcagttgtcatcattgc -3′

LIN28F: 5′-AAAAAAAAGAATTCccagcctggtctttaagagagt -3′

LIN28B: 5′-AAAAAAAAGAATTCcatacagtgaattatttgaaaacacc -3′

PankF: 5′-AAAAAAAAGAATTCcacatctgtgggtgcacttt -3′

PANKR: 5′-AAAAAAAAGAATTCttcaaaggactatttggttaacagc -3′

FABP9F 5′-AAAAAAAACAATTGtggcattcctttctcaccttt -3′

FABP9R 5′-AAAAAAAACAATTGgagccttcctgtgtgggtat -3′

CREBGammaF: 5′-AAAAAAAACAATTGcaaacttctagatggtagaatgatagc -3′

CREBGammaR: 5′-AAAAAAAACAATTGtagccagagaacggaaccac -3′

NaspTF: 5′-AAAAAAAACAATTGtccttggaggacttctgttttc-3′

NaspTR: 5′-AAAAAAAACAATTGggcatgcctgcttaagtgta-3′

Tra2aF: 5′-AAAAAAAAGAATTCattagggactaggatggaacatga -3′

Tra2aR: 5′-AAAAAAAAGAATTCgcatgatggcacatgacttt-3′

ESE mutations within Nasp-T were made by overlap PCR with the additional primers NASPM1-S (5′-GGGTGGACGATAAGACAT GG-3′) and its complementary primer (5′-CCATGTCTTATCGTCCAC CC-3′); NASPM2-S (5′-GTGAGCCTCAAGAGTAGCTCC-3′) and its complementary primer 5′-GGAGCTACTCTTGAGGCTCAC-3′; NASPM3-S (5′-GAATCCTCTGCATAGGCAAAAG-3′) and its complementary primer (5′-CTTTTGCCTATGCAGAGGATT C-3′); NASPM4-S (5′-GGACTGACTCAAGTTGAGGTCGC-3′) and its complementary primer (5′-GCGACCTCAACTTGAGTCAGTCC-3′).

Analysis of splicing of pre-mRNAs transcribed from minigenes was carried out in HEK293 cells as previously described using primers within the β-globin exons of pXJ41 [29]. Because of the length of the regulated exons, additional internal primers were included in multiplex to detect inclusion of the Nasp-T cassette exon (5′-TGCTTTGAAGTCGGTTCAACT-3′) and Tra2a poison exon (5′-TTCAAGTGCTTCTATCTGACCAA-3′).

EMSAs

EMSAs were carried out as previously described [28] using full length Tra2β protein and in vitro translated RNA probes made from constructs containing amplified regions of the mouse genome cloned into pBluescript. Regions of the mouse genome were amplified using the following primers:

Nasp1TraGSF 5′-AAAAAAAAGGTACCGAAGTGGAGAAGGGTGGAAG-3′

Nasp1TraGSB 5′-AAAAAAAAGAATTCGAAGCGACCTCATCTTCATTC-3′

Krba1GSF 5′-AAAAAAAAGGTACCGACTCCTCCCCACCCTAGTC-3′

Krba1GSR 5′-AAAAAAAAGAATTCGCCCAGCCATCTTCTACCTT-3′

Tra2aGSF 5′-AAAAAAAAGGTACCTTAATGTTCGTGAAGAAATTGAAGAG-3′

Tra2aGSR 5′-AAAAAAAAGAATTCTCATTAGCCTTCTTTTATCTTGATTTA-3′

Lin28GSF 5′-AAAAAAAAGGTACCCTTGAACTCTCTGATTTTAGGTTCTTC-3′

Lin28GSR 5′-AAAAAAAAGAATTCAACAGACTAACCTGGGGCTGA-3′

CrebγF 5′-AAAAAAGGTACCTCATTGTTCTAGGTGCTATCAAAGG-3′

CrebγR 5′-AAAAAAGAATTCCTGACATATTTTATTTTCTCATAGTAT GTCTCTC-3′

Creb2F 5′-AAAAAAGGTACCGTAACTAAATGACCATGGAATCTGGAGCA-3′

Creb2R 5′-AAAAAAGAATTCCTGGGCTAATGTGGCAATCTGTGG-3′

Supporting Information

Dataset S1 BED file containing the Tra2β CLIP tag sequences and their location in the mouse genome (mm9). This bed file can be saved and added as an optional track on the UCSC mouse genome browser (http://genome.ucsc.edu/). To load this BED file on the UCSC genome browser, use the "manage custom tracks" button under genomes. Alternatively, the bed file can be visualised by up loading the link http://research.ncl.ac.uk/ElliottGroup/UCSC/hub.txt into the My Hubs textbox in the UCSC Track Hubs menu.
(TXT)

Figure S1 Sequence of all the exons analysed using minigenes and some known Tra2β target exons. The Tra2β binding sites predicted from the k-mer analysis are coloured as indicated in Table S1.
(DOC)

Figure S2 Multiple Tra2β binding sites are phylogenetically conserved in *Tra2a* poison exons and Nasp-T exons. (A) Sequence of the *Tra2a* poison exon from mouse. (B) Sequence of Nasp-T exon from mouse. Nucleotides in red are conserved in all vertebrates analysed (mouse, frog, rabbit, human, rat, cow, orang-utan, chimp, macaque, marmoset, guinea pig, dog, horse, elephant, opossum, lizard, zebrafinch, tetraodon, stickleback, medaka, chicken). Nucleotides conserved in all mammals are shown in blue. All other nucleotides are shown in black. The Tra2β binding sites predicted from the k-mer analysis are shaded as indicated in Table S1, and the positions of CLIP tags are underlined (note that some of these underlined regions correspond to multiple overlapping CLIP tags which have been joined in this figure).
(DOC)

Figure S3 Experiment to confirm the specificity of the polyclonal antisera used for CLIP analysis. HEK293 cells were transfected with plasmids expressing the indicated proteins, proteins isolated and analysed by SDS-PAGE and Western blotting. The same blot was probed sequentially with an affinity purified antisera raised against Tra2β [65] and then with a polyclonal specific for GFP to detect expression of the fusion proteins. The affinity purified α-Tra2β antisera detected a single band in HEK293 cells corresponding to endogenous Tra2β protein, and also the Tra2β-GFP fusion protein. No recognition of either Tra2α or Tra2βΔRS1-GFP was observed, indicating that this antisera is highly specific.
(TIF)

Table S1 Properties of the 30 most frequently retrieved 6-mers in the Tra2β CLIP tags. The 6-mers are ordered from the most frequently recovered at the top of the table (AGAAGA) to the 34th most frequently recovered 6-mer at the bottom (GAAGCT). The 6-mers are arranged in colour blocks of 5 according to their frequency of retrieval, and compared and corrected with their frequencies in both the total mouse genome and mouse testis transcriptome. The same colour code of the different 6-mer categories are also used to illustrate the occurrence of these 6-mers within the Tra2β target exons in Figures S1 and S2.
(DOC)

Table S2 List and properties of all 6-mers recovered by Tra2β CLIP above background levels.
(XLSX)

Table S3 Top functions associated with Tra2β-bound mRNAs determined from Ingenuity Pathway Analysis (IPA).
(DOCX)

Acknowledgments

We acknowledge Lilian Martinez who technically assisted and helped Markus Storbeck and Miriam Jakubik. We thank Jernej UIe (MRC Laboratory for Molecular Biology) for his kind advice, Bill Mattox for the gift of the Tra2α cDNA, Philippa Saunders for advice on immunohistochemistry, and Darren Hoyland for loading the Tra2β BED file so it can be downloaded as a hub from the Newcastle University server.

Author Contributions

Conceived and designed the experiments: DJE SG YL BW MS. Performed the experiments: MS CD YL AB IE KR. Analyzed the data: DJE SG CD MS AB YL TC CFB JS DG MSJ BW. Contributed reagents/materials/analysis tools: MJ YM. Wrote the paper: SG DJE.

References

1. Kan Z, Garrett-Engele PW, Johnson JM, Castle JC (2005) Evolutionarily conserved and diverged alternative splicing events show different expression and functional profiles. Nucleic Acids Res 33: 5659–5666.
2. Lareau LF, Green RE, Bhatnagar RS, Brenner SE (2004) The evolving roles of alternative splicing. Curr Opin Struct Biol 14: 273–282.
3. Stamm S (2002) Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. Hum Mol Genet 11: 2409–2416.
4. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446: 926–929.

5. Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, et al. (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev 21: 708–718.
6. McGlincy NJ, Smith CW (2008) Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? Trends Biochem Sci 33: 385–393.
7. Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463: 457–463.
8. Grellscheid SN, Smith CW (2006) An apparent pseudo-exon acts both as an alternative exon that leads to nonsense-mediated decay and as a zero-length exon. Mol Cell Biol 26: 2237–2246.
9. Licatalosi DD, Darnell RB (2010) RNA processing and its regulation: global insights into biological networks. Nat Rev Genet 11: 75–87.
10. Wang Z, Burge CB (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA 14: 802–813.
11. Smith CW, Valcarcel J (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem Sci 25: 381–388.
12. Tacke R, Tohyama M, Ogawa S, Manley JL (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. Cell 93: 139–148.
13. Beil B, Screaton G, 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: 679–690.
14. Dauwalder B, Amaya-Manzanares F, Mattox W (1996) A human homologue of the Drosophila sex determination factor transformer-2 has conserved splicing regulatory functions. Proc Natl Acad Sci U S A 93: 9004–9009.
15. Shepard PJ, Hertel KJ (2009) The SR protein family. Genome Biol 10: 242.
16. Belote JM, Baker BS (1983) The dual functions of a sex determination gene in Drosophila melanogaster. Dev Biol 95: 512–517.
17. Clery A, Jayne S, Benderska N, Dominguez C, Stamm S, et al. (2011) Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-beta1. Nat Struct Mol Biol.
18. Tsuda K, Someya T, Kuwasako K, Takahashi M, He F, et al. (2011) Structural basis for the dual RNA-recognition modes of human Tra2-{beta} RRM. Nucleic Acids Res.
19. Mende Y, Jakubik M, Riessland M, Schoenen F, Rossbach K, et al. (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: 2154–2167.
20. Daoud R, Da Penha Berzaghi M, Siedler F, Hubener M, Stamm S (1999) Activity-dependent regulation of alternative splicing patterns in the rat brain. Eur J Neurosci 11: 788–802.
21. Nayler O, Cap C, Stamm S (1998) Human transformer-2-beta gene (SFRS10): complete nucleotide sequence, chromosomal localization, and generation of a tissue-specific isoform. Genomics 53: 191–202.
22. Stoilov P, Daoud R, Nayler O, Stamm S (2004) Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. Hum Mol Genet 13: 509–524.
23. Mattox W, Palmer MJ, Baker BS (1990) Alternative splicing of the sex determination gene transformer-2 is sex-specific in the germ line but not in the soma. Genes Dev 4: 789–805.
24. Hazelrigg T, Tu C (1994) Sex-specific processing of the Drosophila exuperantia transcript is regulated in male germ cells by the tra-2 gene. Proc Natl Acad Sci U S A 91: 10752–10756.
25. Madigan SJ, Edeen P, Esnayra J, McKeown M (1996) att, a target for regulation by tra2 in the testes of Drosophila melanogaster, encodes alternative RNAs and alternative proteins. Mol Cell Biol 16: 4222–4230.
26. Elliott DJ, Millar MR, Oghene K, Ross A, Kiesewetter F, et al. (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: 3848–3853.
27. Venables JP, Elliott DJ, Makarova OV, Makarov EM, Cooke HJ, et al. (2000) RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing. Hum Mol Genet 9: 685–694.
28. Grellscheid SN, Dalgliesh C, Rozanska A, Grellscheid D, Bourgeois CF, et al. (2011) Molecular design of a splicing switch responsive to the RNA binding protein Tra2β. Nucleic Acids Res 39: 8092–104.
29. Venables JP, Bourgeois CF, Dalgliesh C, Kister L, Stevenin J, et al. (2005) Up-regulation of the ubiquitous alternative splicing factor Tra2beta causes inclusion of a germ cell-specific exon. Hum Mol Genet 14: 2289–2303.
30. Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, et al. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456: 464–469.
31. Sanford JR, Wang X, Mort M, Vanduyn N, Cooper DN, et al. (2009) Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. Genome Res 19: 381–394.
32. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520–562.
33. Xue Y, Zhou Y, Wu T, Zhu T, Ji X, et al. (2009) Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. Mol Cell 36: 996–1006.
34. Mattick JS, Makunin IV (2006) Non-coding RNA. Hum Mol Genet 15 Spec No 1: R17–29.
35. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141: 129–141.
36. Darnell RB (2006) Developing global insight into RNA regulation. Cold Spring Harb Symp Quant Biol 71: 321–327.
37. Keren H, Lev-Maor G, Ast G (2010) Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 11: 345–355.
38. Ruppert S, Cole TJ, Boshart M, Schmid E, Schutz G (1992) Multiple mRNA isoforms of the transcription activator protein CREB: generation by alternative splicing and specific expression in primary spermatocytes. EMBO J 11: 1503–1512.
39. Zhang LY, Zeng M, Chen P, Sun HQ, Tao DC, et al. (2009) Identification of messenger RNA substrates for mouse T-STAR. Biochemistry (Mosc) 74: 1270–1277.
40. Ray D, Kazan H, Chan ET, Pena Castillo L, Chaudhry S, et al. (2009) Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. Nat Biotechnol 27: 667–670.
41. Sorek R, Shamir R, Ast G (2004) How prevalent is functional alternative splicing in the human genome? Trends Genet 20: 68–71.
42. Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB (2005) Identification and analysis of alternative splicing events conserved in human and mouse. Proc Natl Acad Sci U S A 102: 2850–2855.
43. Richardson RT, Alekseev OM, Grossman G, Widgren EE, Thresher R, et al. (2006) Nuclear autoantigenic sperm protein (NASP), a linker histone chaperone that is required for cell proliferation. J Biol Chem 281: 21526–21534.
44. Coles JL, Hallegger M, Smith CW (2009) A nonsense exon in the Tpm1 gene is silenced by hnRNP H and F. RNA 15: 33–43.
45. Zhang XH, Chasin LA (2004) Computational definition of sequence motifs governing constitutive exon splicing. Genes Dev 18: 1241–1250.
46. Zhang XH, Kangsamaksin T, Chao MS, Banerjee JK, Chasin LA (2005) Exon inclusion is dependent on predictable exonic splicing enhancers. Mol Cell Biol 25: 7323–7332.
47. Tronche F, Kellendonk C, Kretz O, Gass P, Anlag K, et al. (1999) Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety. Nat Genet 23: 99–103.
48. Alekseev OM, Richardson RT, O'Rand MG (2009) Linker histones stimulate HSPA2 ATPase activity through NASP binding and inhibit CDC2/Cyclin B1 complex formation during meiosis in the mouse. Biol Reprod 81: 739–748.
49. Welch JE, O'Rand MG (1990) Characterization of a sperm-specific nuclear autoantigenic protein. II. Expression and localization in the testis. Biol Reprod 43: 569–578.
50. Osakabe A, Tachiwana H, Matsunaga T, Shiga T, Nozawa RS, et al. (2010) Nucleosome formation activity of human somatic nuclear autoantigenic sperm protein (sNASP). J Biol Chem 285: 11913–11921.
51. McGuffin ME, Chandler D, Somaiya D, Dauwalder B, Mattox W (1998) Autoregulation of transformer-2 alternative splicing is necessary for normal male fertility in Drosophila. Genetics 149: 1477–1486.
52. Gabut M, Chaudhry S, Blencowe BJ (2008) SnapShot: The splicing regulatory machinery. Cell 133: 192 e191.
53. Hofmann Y, Lorson CL, Stamm S, Androphy EJ, 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: 9618–9623.
54. Jiang Z, Tang H, Havlioglu N, Zhang X, Stamm S, et al. (2003) Mutations in tau gene exon 10 associated with FTDP-17 alter the activity of an exonic splicing enhancer to interact with Tra2 beta. J Biol Chem 278: 18997–19007.
55. Stamm S, Casper D, Hanson V, Helfman DM (1999) Regulation of the neuron-specific exon of clathrin light chain B. Brain Res Mol Brain Res 64: 108–118.
56. Yamada T, Goto I, Sakaki Y (1993) Neuron-specific splicing of the Alzheimer amyloid precursor protein gene in a mini-gene system. Biochem Biophys Res Commun 195: 442–448.
57. Disset A, Bourgeois CF, Benmalek N, Claustres M, Stevenin J, et al. (2006) An exon skipping-associated nonsense mutation in the dystrophin gene uncovers a complex interplay between multiple antagonistic splicing elements. Hum Mol Genet 15: 999–1013.
58. Young PJ, DiDonato CJ, Hu D, Kothary R, Androphy EJ, et al. (2002) SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 beta 1. Hum Mol Genet 11: 577–587.
59. Cartegni L, Krainer AR (2002) Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. Nat Genet 30: 377–384.
60. Kashima T, Manley JL (2003) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. Nat Genet 34: 460–463.
61. Elliott DJ, Grellscheid SN (2006) Alternative RNA splicing regulation in the testis. Reproduction 132: 811–819.
62. Venables JP, Klinck R, Koh C, Gervais-Bird J, Bramard A, et al. (2009) Cancer-associated regulation of alternative splicing. Nat Struct Mol Biol 16: 670–676.
63. Wang Z, Kayikci M, Briese M, Zarnack K, Luscombe NM, et al. (2010) iCLIP predicts the dual splicing effects of TIA-RNA interactions. PLoS Biol 8: e1000530. doi:10.1371/journal.pbio.1000530.
64. Hofmann Y, Wirth B (2002) hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1. Hum Mol Genet 11: 2037–2049.

65. Sergeant KA, Bourgeois CF, Dalgliesh C, Venables JP, Stevenin J, et al. (2007) Alternative RNA splicing complexes containing the scaffold attachment factor SAFB2. J Cell Sci 120: 309–319.
66. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.
67. Mignone JL, Kukekov V, Chiang AS, Steindler D, Enikolopov G (2004) Neural stem and progenitor cells in nestin-GFP transgenic mice. J Comp Neurol 469: 311–324.
68. Bourgeois CF, Popielarz M, Hildwein G, Stevenin J (1999) Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5′ or 3′ splice site activation. Mol Cell Biol 19: 7347–7356.
69. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, et al. (2011) The UCSC Genome Browser database: update 2011. Nucleic Acids Res 39: D876–882.