Investigating the regulation of alternative splicing by Tra2 proteins and RBMX in triple negative breast cancer cells

Gerald Hysenaj

Institute of Genetic Medicine

A thesis submitted for the degree of Doctor of Philosophy

November 2020

Declaration

I, Gerald Hysenaj, 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.

……..

Gerald Hysenaj

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor David Elliott for the continuous support of my PhD study and research, for his never-ending patience, incredible kindness, unmatched mentorship, motivation and scientific knowledge. His wise guidance has been essential to the completion of this work. I could not have imagined having a better advisor and mentor for such an important chapter of my life, and whose step I can only dream of following.

I would also like to thank my second supervisor, Dr Alison Tyson-Capper, for her kindness and support throughout this project. Her care, encouragement and guide have given me hope and motivation to pursue my scientific career.

It has been a great pleasure to work with members of the Elliott lab. Firstly, I would like to thank Dr Andrew Best whose work set the foundations for my project and whose lab lessons are still fresh in my mind. I would like to thank Dr Jennifer Munkley for her continued support, advice and friendship. I would also like to thank Mrs Caroline Dalgliesh and Dr Ingrid Ehrmann for their help and advice. My thanks go to many collaborators whose work helped with this project: Dr Katherine James, Mr Yaobo Xu, Mr Matthew Gazzara, Prof Yoseph Barash, Dr Mahsa Kheirollahi-Kouhestani and Prof Eduardo Eyras. I also gratefully acknowledge the BBSRC for funding this project.

I would like to extend my thanks to three undergraduate students that I had the pleasure to train and work with on this project: Sin Min Chan, Louisa Thomas-Eke and Scott Naylor.

Finally, I would like to thank to my family and friends for their support. I thank my girlfriend Silvia for her loving support during the most difficult part of this PhD. I would also like to thank my younger brother Enea whom I have missed a lot during my time here. Special thanks go to my parents Ramis and Shazije who have shown endless care, love and kindness. Their continual motivation and encouragement have kept me going relentlessly. I love you very much.

Table of Contents

List of Figures and Tables

List of Abbreviations

ATP	Adenosine triphosphate
BER	Base Excision Repair
bp	base pair
CBC	Cap binding complex
cDNA	Complementary DNA
CSK buffer	CytoSKeletal buffer
CDS	Coding DNA sequence
AO	Antisense oligonucleotide
BPS	Branch point sequence
DAPI	4',6-diamidino-2-phenylindole
DE	Differential expression
DEPC	diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DSB	double strand break
EJC	exon junction complex
EMSA	electrophoretic mobility shift assay
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
FL	full length
GFP	Green Fluorescent Protein
GO	Gene Ontology
HEK293T	human embryonic kidney 293 cell line
HeLa	human cervical carcinoma cell line
HITS-CLIP	high-throughput sequencing of RNA isolated by CLIP
hnRNP	heterogenous nuclear ribonucleoprotein
ICL	Interstrand crosslinks
iCLIP	individual-nucleotide resolution Cross-Linking ImmunoPrecipitation
IP	Immunoprecipitation
ISE	Intronic Splicing Enhancer

ISS	Intronic Splicing Silencer
kDa	Kilo Daltons
LB	Luria Broth
LSV	local splicing variation
MAJIQ	Modelling Alternative Junction Inclusion Quantification
MDA-MB-231	MD Anderson human invasive breast cancer cell line
miRNA	micro RNA
MISO	Mixture of isoforms
mRNA	messenger RNA
mTORC	mamalian Target of Rapamycin Complex
NER	Nucleotide Excision Repair
NMD	Nonsense Mediated Decay
ORF	Open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFA	Paraformaldehyde
Pfam	Protein family
PPT	polypyrimidine tract
PSI	percentage splicing inclusion
RBP	RNA Binding Protein
RIPA	Radioimmunoprecipitation assay buffer
RNA	Ribonucleic acid
RNAseq	RNA sequencing
UTR	Untranslated region
Sh	Short
siRNA	small interfering RNA
snRNP	small nuclear ribonucleoprotein
SR-Proteins	Serine/Arginine rich proteins
SROOGLE	Splicing regulation online graphical engine
UCSC	University California Santa Cruz
WST-1	tetrazolium salt used to assay cell proliferation

Abstract

Alternative splicing is a key process that enables the expression of a large number of proteins from a limited number of genes. Regulation of alternative splicing has been directly linked with the hallmarks of cancer. Human Transformer 2-α and -β are two homologous RNA binding proteins involved in the regulation of alternative splicing that are overexpressed in many different cancers. Previous work suggests Tra2β interacts closely with RBMX, encoded by the gene *RBMX,* to regulate the splicing of key pathologically related genes such as *SMN2*. The objective of this study is to investigate the role of alternative splicing regulators in breast cancer in order to further understand their biological function. Firstly, I utilise previously obtained RNAseq data from the double knockdown of Tra2α and Tra2β in the MDA-MB-231 breast cancer cell line to identify an extended number of genes regulated by Tra2 proteins using recently developed bioinformatics pipelines. The analysis revealed a range of alternative events including intron retentions, alternative 3' and 5' splice sites and exon repression events, suggesting different mechanisms of splice site regulation by Tra2 proteins. I experimentally validated the results of the RNAseq analysis by PCR. Subsequent Gene Ontology enrichment analysis of Tra2-regulated targets reveals several genes involved in the cellular response to DNA damage. I employed western immunoblotting to investigate the protein expression changes in these genes and was able to detect short protein isoforms expressed for *MBD4* and *XPA*. Subsequently, nuclear fractionation and immunofluorescence microscopy of Tra2 depleted cells showed cytoplasmic localisation of the short isoforms for both *MBD4* and *XPA*. In order to determine the role of the alternative isoforms regulated by Tra2 proteins, antisense oligonucleotide targeting exon 5 of *BRIP1*, exon 3 of *CHEK1*, and exon 5 of *XPA* were transfected into MDA-MB-231 cells. Expression of the short *XPA* isoform using antisense oligonucleotides results in an increase of γH2AX and reduced cell viability. This data suggests Tra2 proteins sustain the DNA integrity and protect the cells from death by ensuring the expression of correctly spliced DNA repair genes. In order to compare the role of Tra2 proteins with the commonly associated splicing regulator RBMX, I knocked down the expression of RBMX in MDA-MB-231 cells and obtained RNAseq data. Bioinformatics analysis was able to identify many commonly regulated alternative events, but also some novel isoforms regulated by RBMX alone. RNAseq data also unveiled an interesting autoregulatory mechanism for the *RBMX* gene, whereby the expression of *RBMX* is regulated by alternative splicing of the 3' UTR. Subsequent examination of Upf1 knockdown cells showed stabilisation of the alternating isoform of *RBMX* indicating it is targeted by nonsense mediated decay. Future work into the role of RBMX regulated targets similar to the work done for Tra2-regulated genes could reveal a complete and clear picture of the function of RBMX in breast cancer cells.

Chapter 1: Introduction

1.1 The central dogma

The human genome sequencing project completed 16 years ago was a monumental achievement in human genetics that allowed scientists for the first time to navigate through the blueprint of life - the human genome. The central dogma of molecular biology states that information flows from DNA to RNA and finally to protein. However, this principle oversimplifies the true nature of biological processes that occur in order to carry out the instructions of the genome. In prokaryotes the flow of information from DNA to protein is somewhat linear. DNA is transcribed into RNA and almost simultaneously translated into protein. However, in eukaryotes transcription and translation occur separately. The precursor messenger RNA (pre-mRNA) transcribed from the DNA is a complementary molecule that must undergo a series of post-transcriptional modifications prior to translation. Pre-mRNA processing begins with 5' capping and is closely followed by splicing and polyadenylation. The pre-mRNA contains protein coding sequences known as exons separated by non-coding sequences known as introns. One of the first steps of pre-mRNA modifications that occurs simultaneously with transcription is pre-mRNA splicing, whereby the non-coding introns are spliced out, leaving behind a coding sequence that can be translated into protein.

An apparent consequence of post-transcriptional editing of RNA is the increased repertoire of proteins from a limited number of genes. The human genome contains approximately 21,000 protein coding genes (Pertea et al., 2018). However, the number of detectable unique proteins expressed in human cells surpasses 90,000 (Harrison, Kumar, Lang, Snyder, & Gerstein, 2002). This proteomic diversity is a characteristic that becomes more apparent in species of the animal kingdom, especially those further out in the evolutionary tree. Such diversity is a consequence of post-transcriptional modifications such as alternative splicing, alternative polyadenylation and RNA editing (Lander, 2011). More than 95% of the human protein coding genes are alternatively spliced (Q. Pan, Shai, Lee, Frey, & Blencowe, 2008). As a result, a single pre-mRNA can be spliced differently resulting in multiple ‘mature’ mRNA species that code for different protein isoforms. Furthermore, alternative splicing can affect the reading frame of the mRNA creating non-coding mRNA sequences, therefore playing a role in post-transcriptional regulation of gene expression.

Post-transcriptional modifications such as alternative splicing produce protein isoforms that can be structurally and functionally distinct from one another. Post-transcriptional modifications of the RNA coupled with post-translational modifications to proteins, such as phosphorylation and

ubiquitylation, dramatically increase the complexity of the proteome from a limited number of genes (Wilhelm et al., 2014). Understanding the mechanism and role of post-transcriptional (and post-translational) modifications is essential to understanding the full potential of the genome. Evidently the central dogma model of information flow is oversimplified and does not represent the complexity of regulatory pathways involved in genome transcription and translation in human cells.

1.2 RNA transcription and post-transcriptional modifications

RNA transcription is a complex reaction which creates RNA molecules complementary to the antisense DNA strand. There are three steps to RNA transcription: initiation, elongation and termination. The most important multiprotein complex involved in these steps is RNA Pol-II which interacts with a number of processing factors that coordinate transcription (Glover-Cutter, Kim, Espinosa, & Bentley, 2008). RNA Pol-II first recognises and binds to the promoter region of a gene causing the DNA to unwind hence initiation transcription (Mischo & Proudfoot, 2013). During elongation RNA Pol-II sequentially reads the antisense strand of the DNA and adds complementary nucleotides to the nascent RNA being transcribed. The final step of transcription relies on RNA Pol-II recognising the transcription termination signal. RNA Pol-II then recruits the polyadenylation complex via the C-Terminal Domain (CTD), detaches from the DNA and releases the RNA transcript.

1.2.1 5’ capping and 3’ polyadenylation of the mRNA

A general misconception regarding post-transcriptional modifications is the idea that they occur after transcription is complete - probably due to the misleading prefix “post”. However, many studies have shown that most post-transcriptional modifications occur simultaneously with transcription. One of the first mRNA modifications is the addition of a 5’ cap. In humans, this can be performed by two different enzymes: human capping enzyme (HCE) and RNA 7-methyltransferase (MT) (Furuichi & Shatkin, 2000). RNA Pol-II can recruit both HCE and MT via its’ CTD domain when phosphorylated (Shuman, 1997). The 5’ cap is a guanine nucleotide, methylated on the 7th carbon of the guanosine base (m7G), that binds to the 5’ carbon of the first nucleotide of the nascent RNA via its’ own 5’ carbon through a triphosphate bridge. The m7G cap is essential for mRNA stability, but it is the cap-binding complex (CBC), comprised of Cbp20 and Cbp80, which causes importance of the cap in various functional processes of mRNA modifications (Figure 1) (Gonatopoulos-Pournatzis & Cowling, 2014). The CBC can interact with and recruit components of the spliceosome thereby playing a role in pre-mRNA splicing (Konarska, Padgett, & Sharp, 1984; Pabis et al., 2013). Furthermore, the CBC guides the mature mRNA through the nuclear envelope during nuclear export of mRNA (Daneholt, 2001).In the cytoplasm, the cap-bound CBC recruits the eukaryotic translation

initiation factors-4G (eIF4G) and eIF4A to the 5’ end of the mRNA. Depending on the success of the pioneer round of translation the CBC-bound mRNA is targeted by nonsense mediated decay (NMD) or the eIF4E replaces CBC and begins the steady-state round of translation (Maquat, Hwang, Sato, & Tang, 2010). Among other roles, the 5’ cap-bound CBC has also been shown to interact with 3’ processing factors bound to the polyadenylation sites of the mRNA (Ramanathan, Robb, & Chan, 2016). Depletion of CBC in HeLa cells strongly reduced endonucleolytic cleavage and polyadenylation (Flaherty, Fortes, Izaurralde, Mattaj, & Gilmartin, 1997).

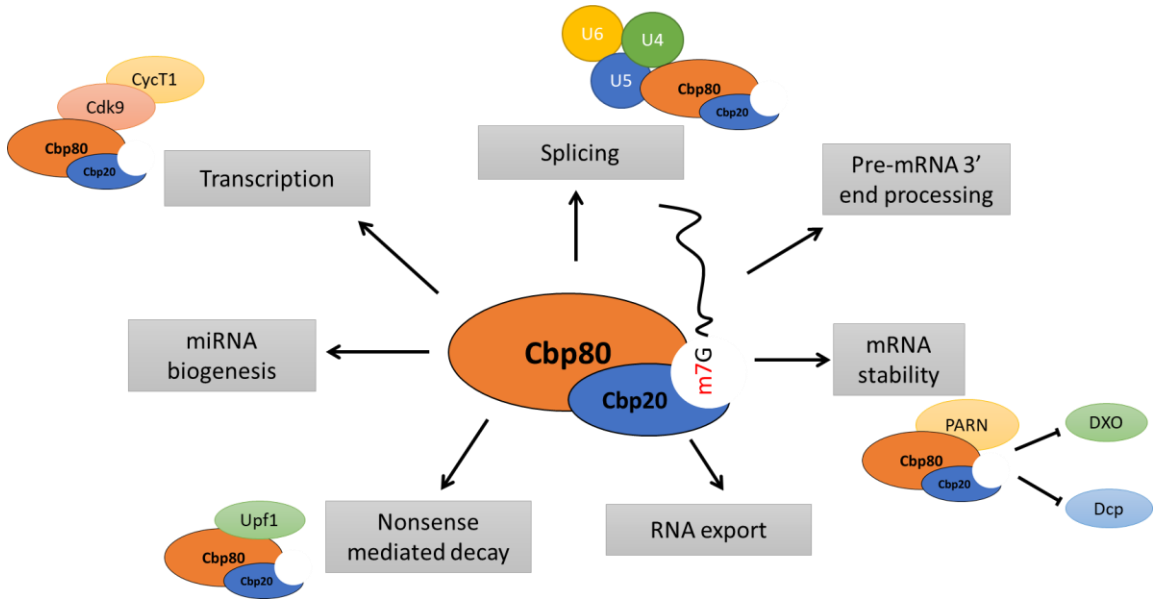

Figure 1 Cap Binding Complex (CBC) role in RNA processing. CBC composed of Cbp80 and Cbp20 subunits binds to m7G cap and facilitates interactions with key components of RNA processing highlighting the importance of the 5’ cap. CycT1, cyclin T1; Cdk9, Cyclin dependent kinase 9; Upf1, RNA helicase and ATPase; U4, U5 and U6, Spliceosome snRNPs; Dcp, decapping mRNA; DXO decapping exoribonuclease. Image adapted from (Gonatopoulos-Pournatzis & Cowling, 2014)

Polyadenylation is in itself another important post-transcriptional modification step that plays a role in mRNA stability and nuclear export. Upon transcription termination several factors are recruited to the 3’ end of the mRNA in order to add the poly(A) tail. A poly(A) tail is a nearly universal feature of mRNAs in eukaryotes. The cleavage and polyadenylation specificity factor (CPSF) is one of the main interacting proteins that recognises the strongly conserved polyadenylation signal AAUAAA and cleaves the mRNA approximately 20 nucleotides downstream (N. J. Proudfoot & Brownlee, 1976). The addition and elongation of the poly(A) chain is then performed by the poly(A) polymerase with the help of the poly(A)-binding protein II (PABII) (Bienroth, Keller, & Wahle, 1993). Interestingly, studies suggest that binding of PABII, even in the absence of a poly(A) tail, is sufficient for mRNA stabilisation (Coller, Gray, & Wickens, 1998), and it is PABII that regulates the length of the poly(A) tail by interacting both with poly(A) polymerase and poly(A) nuclease (Bienroth et al., 1993). PABII

facilitates the transport of the mRNA through the nuclear envelope and recruits translation initiation factors in the cytoplasm (N. K. Gray, Coller, Dickson, & Wickens, 2000; Vinciguerra & Stutz, 2004).

Although well conserved, the polyadenylation signal sequence is relatively short and may occur multiple times within the mRNA. Therefore, it is possible for an mRNA to be polyadenylated at different sites producing multiple isoforms - alternative polyadenylation. An important determinant for polyadenylation site selection is the strength of the signal, which is determined by the distance between the hexanucleotide AAUAAA and the GU-rich element, an auxiliary sequence usually positioned 20-40 nucleotides downstream of the cleavage site (N. Proudfoot, 1991). Another aspect that influences alternative polyadenylation is the concentration of the regulatory factors. As an example, the low concentration of the cleavage stimulation factor (CstF) in B lymphocytes causes the expression of the membrane (μm) bound isoform of the IgM heavy chain (Figure 2) (Colgan & Manley, 1997). While in plasma cells the abundance of CstF means it no longer acts as a limiting factor, therefore allowing for the selection of the weaker μs polyadenylation signal. Additionally, recent studies suggest that splicing of the last intron of an mRNA can be functionally linked with polyadenylation, indicating splicing factors can act as auxiliary elements in poly(A) site selection (Niwa, Rose, & Berget, 1990; Wassarman & Steitz, 1993). The sequence elements and regulatory factors can be classified as *cis*- and *trans*-acting elements, respectively, and the same principle will be discussed again in relation to alternative splicing.

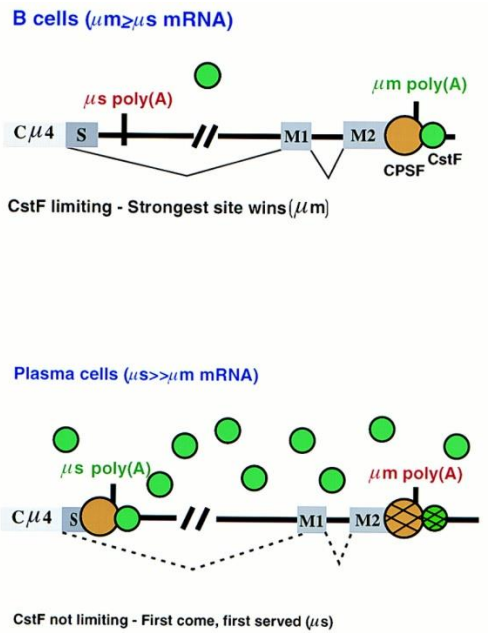

Figure 2 The concentration of Cleavage stimulation factor (CstF) and strength of polyadenylation site are two determining factors in alternative polyadenylation site selection as demonstrated by the example of IgM heavy chain isoforms present in B cells and Plasma cells. Image taken from Colgan and Manley, 1997

1.2.2 Pre-mRNA splicing and alternative splicing in eukaryotes

Pre-mRNA splicing is an important RNA modification that removes non-coding introns and sequentially joins exons together to form a continuous protein coding sequence known as the open reading frame (ORF) ready for translation (Matlin, Clark, & Smith, 2005). However, splicing plays a crucial role in determining the final coding sequence of the mRNA, hence having an impact in both the transcriptome and proteome of the cell. Splicing occurs simultaneously with transcription and begins shortly after transcription initiation and 5' capping. The recognition and splicing of introns is performed by the splicing machinery and is influenced by a range of auxiliary RNA-binding proteins (RBPs) and heterogeneous nuclear ribonucleoproteins (hnRNPs) (Cramer et al., 2001). While some exons are constitutively spliced, meaning they are present in every mRNA isoform produced from a given gene, many are alternatively spliced resulting in alternative splice patterns, a process known as alternative splicing. Previous studies have shown that RNA Pol II can play a role in the determination of splice patterns. Slowing down transcription elongation rate in human and yeast cells through mutations of the RNA Pol II can decrease exon skipping (de la Mata et al., 2003; Ip et al., 2011). Furthermore, the RNA binding protein SRSF2, an important pre-mRNA splicing factor, can bind paused RNA Pol II and activate transcription elongation, indicating that components of the splicing machinery can also influence transcription (Ji et al., 2013).

Alternatively spliced transcripts often differ in their coding capacity, stability, or translational efficiency (Figure 3) (Kelemen et al., 2013). Human protein coding genes produce on average three different mRNA transcripts (Djebali et al., 2012). While the majority of alternatively spliced transcripts introduce diversity to the proteome by affecting the protein coding sequence, an estimated one third of alternatively spliced mRNAs contain alterations to the ORF (Lewis, Green, & Brenner, 2003). A shift in the ORF can potentially introduce premature termination codons (PTCs) causing the mRNA to be targeted by nonsense-mediated decay (NMD), a surveillance mechanism that actively degrades nonsense mRNAs. The close connection between alternative splicing and NMD suggests that alternative splicing can regulate gene expression. Furthermore, alternative splicing of the untranslated regions (UTRs) can also influence gene expression by affecting mRNA stability (Hughes, 2006). The 5' UTR is the region between the 5' cap and the ORF, and is estimated to contain upstream ORFs in 10% of human mRNAs, suggesting alternative splicing can influence gene expression via alternative 5' UTR splicing (Pesole et al., 2001). Furthermore, regulatory elements within the 5' UTR involved in the recruitment of translation initiation factors and ribosomal binding can differ across alternative isoforms of the same gene and therefore influence translation efficiency. The 3' UTR is located between the termination codon and the poly(A) tail, and is on

average five times longer than the 5’ UTR in humans (Pesole et al., 2001). Alternative polyadenylation and alternative splicing can both affect the 3’ UTR sequence. Splice junctions within the 3’ UTR located further than 55 nucleotides downstream of the stop codon can flag the mRNA for degradation via NMD (Nagy & Maquat, 1998; J. Zhang, Sun, Qian, & Maquat, 1998). The importance of accurate splicing is further highlighted by the fact that more than 15,000 variants affect splice sites or regulatory sequences that lead to wide range of cancers (He, Zhou, Zuo, Cheng, & Zhou, 2009). Moreover, there is increasing evidence associating atypically long and complex UTRs containing regulatory elements with oncogenes, tumour suppressors and other genes associated with cell proliferation (Kozak, 1991; Singh, Wolfe, Zhong, Rätsch, & Wendel, 2015; Venkat et al., 2020),

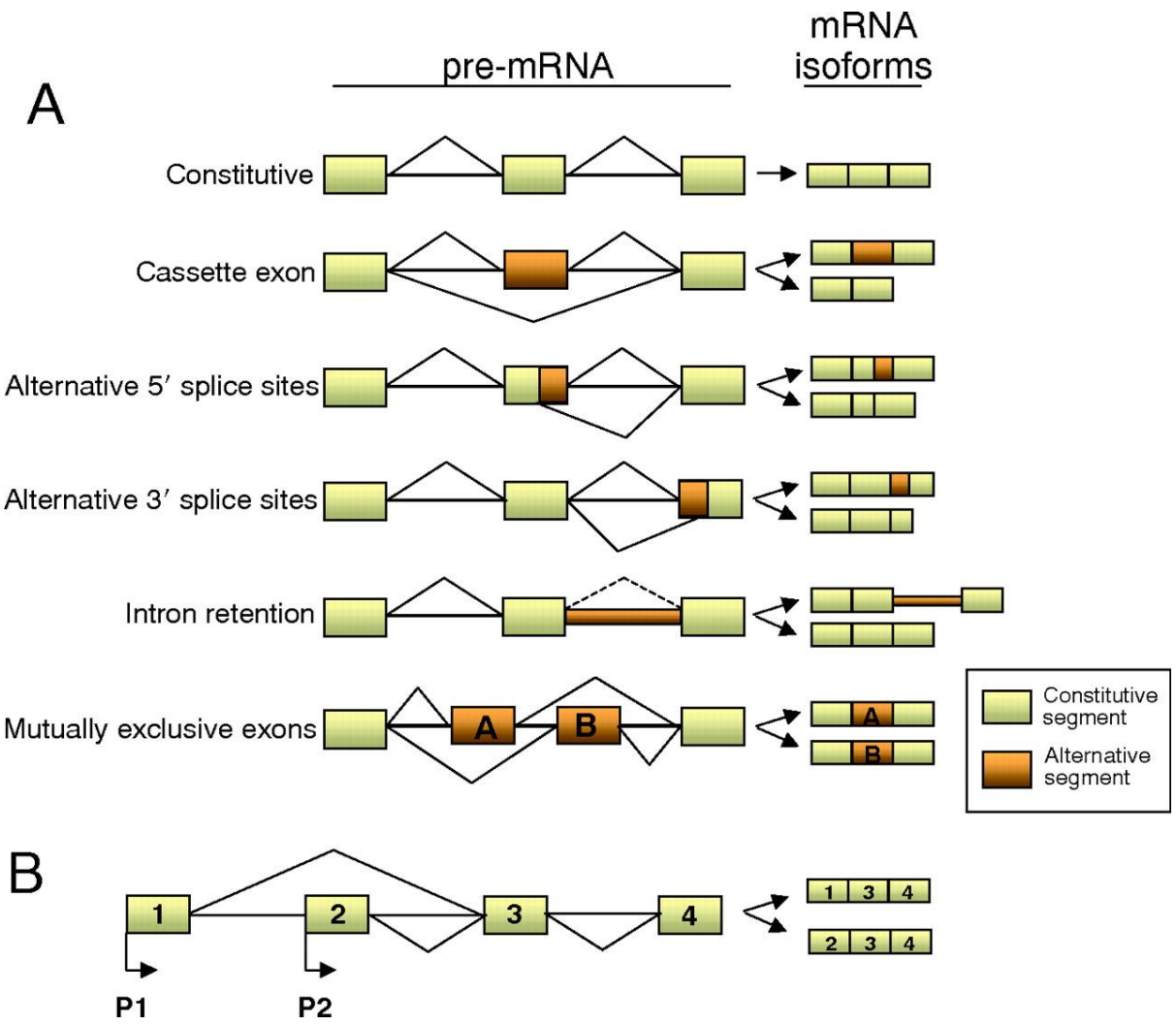

Figure 3 (A) The different modes of alternative splicing, including (B) alternative promoter usage. Image adapted from (Srebrow & Kornblihtt, 2006)

1.2.2.1 Spliceosome machinery and the splicing reaction

The recognition and excision of non-coding intronic sequences from transcribed pre-mRNAs is performed by a multi-megadalton ribonucleoprotein (RNP) complex known as the spliceosome (Papasaikas & Valcárcel, 2016). In humans two spliceosomes exist, the major and the minor spliceosome. They are comprised of over 150 proteins and a set of small nuclear RNPs (snRNPs) (Rappsilber, Ryder, Lamond, & Mann, 2002). The major spliceosome consists of five snRNPs known as U1, U2, U4/U6 and U5 which are involved in the recognition and removal of U2-type introns. Whereas, the minor spliceosome catalyses the removal of the less abundant U12-type introns, utilising the analogous snRNP group of U11/U12 and U4atac/U6atac snRNPs, with the U5 snRNP shared between the machineries (Patel & Steitz, 2003). There is some evidence suggesting the spliceosome is pre-assembled into a megastructure containing four active spliceosomes termed as the supraspliceosome (Azubel, Habib, Sperling, & Sperling, 2006; Patel & Steitz, 2003), however the majority of scientific reports indicate that the spliceosome is assembled anew at each splicing event (Wahl, Will, & Luhrmann, 2009; Will & Lührmann, 2011).

Mechanistically, the spliceosome directs two sequential S_N2-type transesterification reactions (Figure 4) (Wahl et al., 2009). The first step involves the nucleophilic attack on the phosphodiester bond of the 5' splice site (ss) by the branchpoint adenosine nucleotide resulting in cleavage of the 5' exon and generation of the lariat intron. The second transesterification reaction occurs when the 3'OH group of the cleaved 5' exon attacks the 3' splice site (3'ss). Consequently, the exons are ligated together, and the intron is released as a lariat.

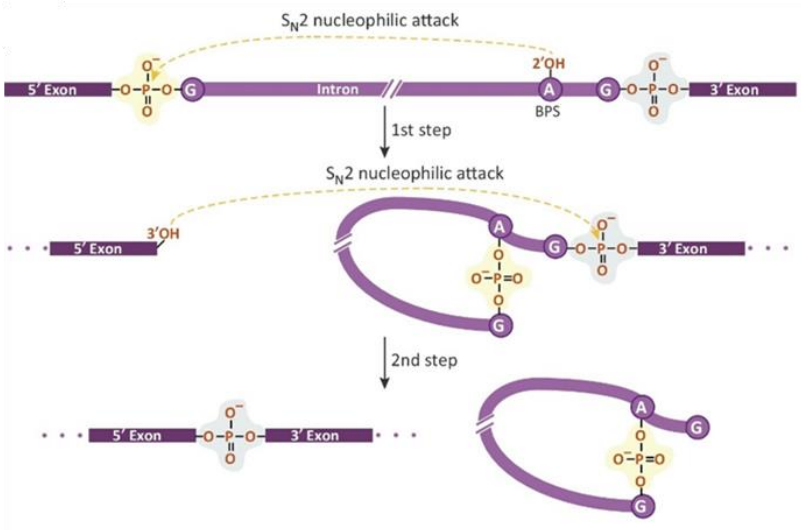

Figure 4 The splicing reaction as represented by the two SN2-type transesterification reactions. These are the two key reactions facilitated by over 200 proteins and RNAs that assemble to form the spliceosome. BPS, Branch point sequence. Image adapted from (Papasaikas & Valcárcel, 2016)

Throughout the process of splicing, the spliceosomal components assemble on the RNA in a defined series of complexes known as the H, A, B and C splicing complexes (Wahl et al., 2009). Complex H is effectively pre-spliceosome assembly and is an essential step that involves the association of several RBPs with the pre-mRNA guiding the spliceosome to the right splice sites. The initiating step of U1 snRNP interaction with the 5' splice site (5'ss) is stabilised by other auxiliary RBPs. The U2 auxiliary factor (U2AF) interacts with the splicing factor 1/branch point bound protein (SF1/BBP) to recognise the branch site upstream the 3' splice site (3'ss) (also known as the E complex). Subsequently, U2 snRNP binds to the branch site in an ATP-dependent manner replacing SF1/BBP, forming complex A. Complex B formation involves the association of a tri-snRNP complex composed of U4/U6 and U5 with the spliceosome which is guided by a large complex of protein known as NineTeen Complex (NTC) (Figure 5) (Hogg, McGrail, & O'Keefe, 2010). During the first transesterification reaction catalysed by the RNA-dependent ATPase Prp2 the spliceosome undergoes remodelling, leading to the dissociation of U1 and U4 forming complex C. In the second transesterification reaction the NTC guides the removal of the intron lariat and subsequent ligation of the upstream 5'ss exon to the 3'ss exon. Upon completion of the splicing reaction, the removed intron lariat is degraded and the snRNPs are recycled in subsequent splicing reactions. Spliceosome formation is reviewed in further detail by (Wahl et al., 2009).

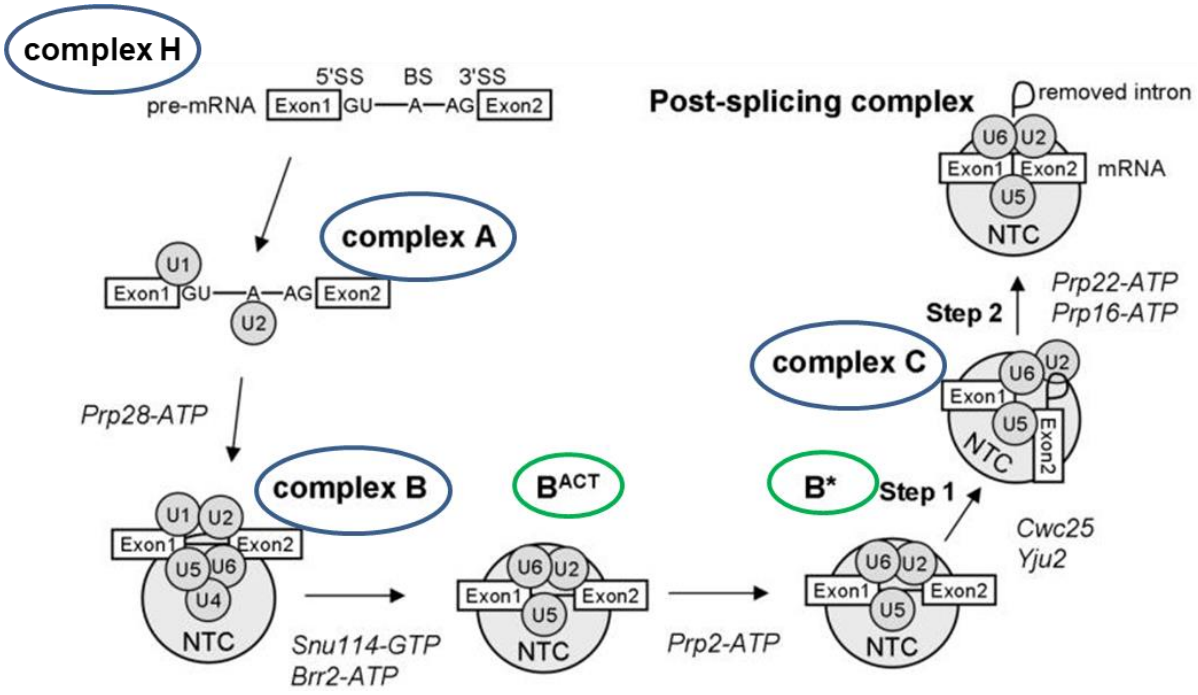

Figure 5 Major spliceosome assembly and the respective complexes formed during intron removal from the pre-mRNA. Image was adapted from (Hogg et al., 2010)

1.2.2.2 Splice site recognition and cis-acting elements

Exons are often classified as constitutive or alternative exons. This classification reflects the strength of the splice site signal recognised by the spliceosome during complexes H and A. The recognition of splice sites by the spliceosome machinery is essential for correct splicing and is dependent on the complementarity between splice site sequences and the snRNPs. Constitutively spliced exons are often flanked by strong splice sites relative to alternative exons. Splice site strength can be predicted using online tools such as Splicing RegulatiOn Online Graphical Engine (SROOGLE) (Schwartz, Hall, & Ast, 2009).

Splice site recognition can be influenced by c*is*-acting elements. These are sequences on the pre-mRNA that enhance (exonic splicing enhancers (ESE) and intronic splicing enhancers (ISE)) or inhibit/silence (exonic splicing silencers (ESS) and intronic splicing silencers (ISS)) splice site selection. Recognition of both constitutive and alternative exons is largely influenced by splicing enhancer and silencer sequences (Matlin et al., 2005). *Cis*-regulatory elements are highly enriched within the regulated exons or close proximity to the splice sites (Sorek & Ast, 2003), but distal regulatory sequences have also been reported (Ule et al., 2006).

1.2.2.3 Trans-acting factors: SR-proteins and hnRNPs

For *cis*-regulatory elements to exhibit their effect on splice site recognition they must be bound by RNA binding proteins. These are *trans*-regulatory proteins which include members of the SR-protein family and the heterogenous nuclear ribonucleoproteins (hnRNPs). Generally, SR-proteins act to enhance splice site recognition by binding to ESEs or proximal ISEs and by recruiting U1 snRNP to the 5'ss during complex A formation (Zhou & Fu, 2013). SR-proteins contain an arginine/serine-rich domain (RS-domain) and one or more RNA recognition motifs (RRMs). The RS-domain plays an important role for protein-protein interactions with other SR-proteins and components of the spliceosome (J. Y. Wu & Maniatis, 1993), whereas the RRM facilitates direct interaction with the RNA through *cis*-elements. Classical SR-proteins (SRSF1 – SRSF6) are defined based on recognition by the mAb104 monoclonal antibody, structural similarities, purification using magnesium chloride and function in both alternative and constitutive splicing (Long & Caceres, 2009). However, structurally similar proteins known as SR-like proteins are also involved in splicing regulation. SR-like proteins contain at least one RS-domain and an RRM allowing them to interact with the *cis*-regulatory elements, other SR-proteins and recruit spliceosomal snRNPs. Examples of SR-like proteins include both subunits of the U2AF heterodimer (U2AF65 and U2AF35), as well as the homologous pair Tra2α and Tra2β (Long & Caceres, 2009). The activity of SR proteins depends on two main factors:

concentration and phosphorylation state. SR-protein kinases such as SRPK1 phosphorylate serine residues on the RS-domains of SR-proteins (Zhong, Ding, Adams, Ghosh, & Fu, 2009). Several models have been previously proposed for the mechanisms of splicing regulation by SR-proteins (Figure 6). The classical model of U2AF recruitment to ESE sequences and stabilisation of the interaction between the spliceosome and the splice sites is the most widely accepted model (Graveley, Hertel, & Maniatis, 2001). SR-proteins may also act by inhibiting the antagonistic function of hnRNPs, which are thought to silence splice sites when bound to exonic *cis*-regulatory elements. This model is concentration dependent as it suggests SR-proteins compete with hnRNPs to bind exonic sequences (Zhu, Mayeda, & Krainer, 2001). Another commonly accepted mechanism of action suggests SR-proteins bind to intronic elements near the splice site and inhibit U2AF recruitment. This model suggests an inhibitory role for SR-proteins that is dependent on the mRNA location of their binding (Erkelenz et al., 2013). Finally, SR-proteins have been suggested to affect the splicing of a target pre-mRNA without directly binding to it, but instead interacting via the RS-domain with the U4/U6 and U5 tri-snRNP to facilitate its recruitment to the spliceosome (Long & Caceres, 2009).

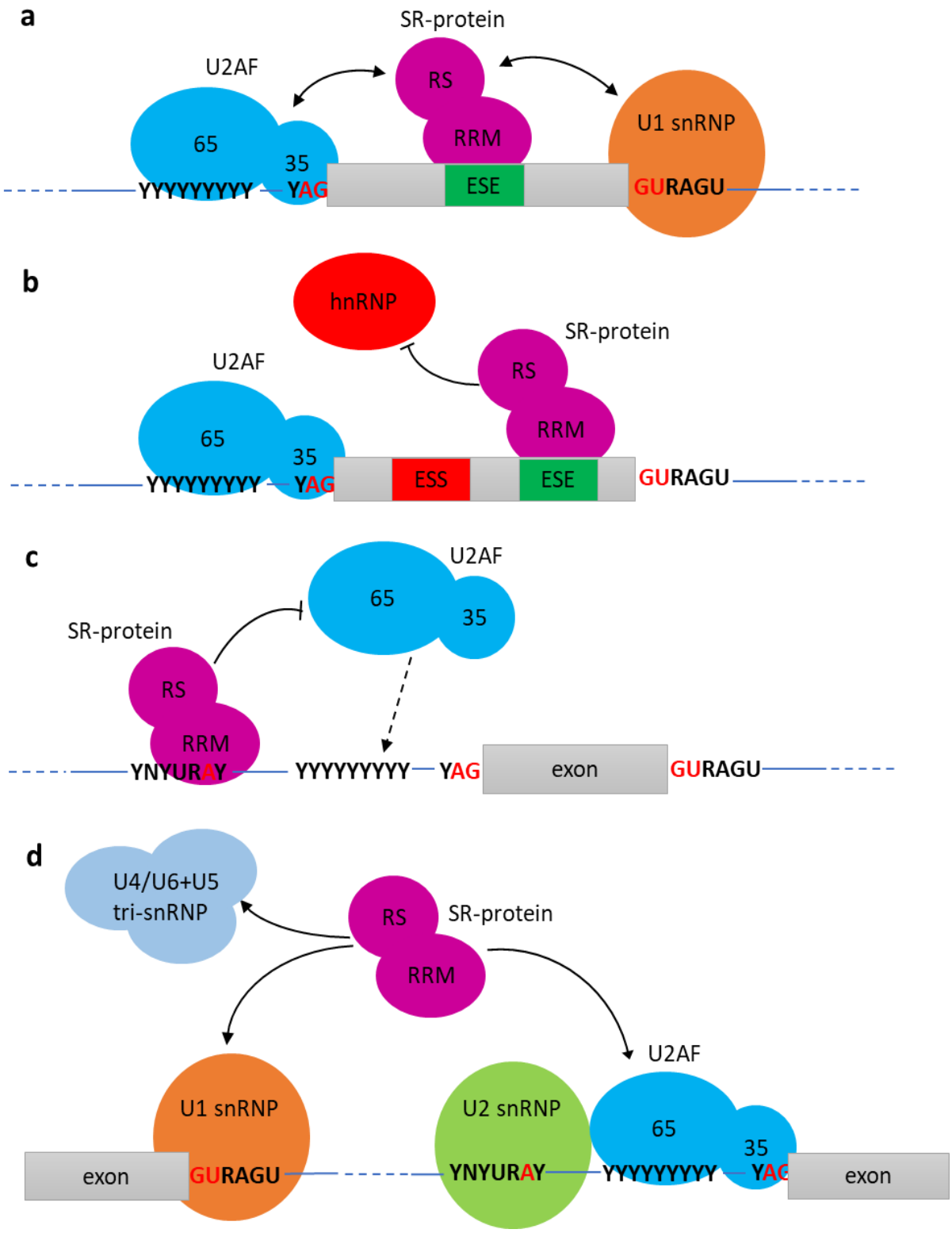

Figure 6 Splicing regulatory models for SR-proteins. A) The classical model of SR-proteins binding to ESEs and stabilising the interaction between U2AF and the 3'ss. B) 'Competitive' model where SR-proteins prevent the antagonistic action of hnRNPs. C) By binding onto intronic sequences SR-proteins can repress splice site recognition. D) SR-proteins may recruit spliceosomal components independently of RNA-binding. Image adapted from Dr Andrew Best thesis, 2014.

Unlike SR-proteins, hnRNPs inhibit splicing when binding to exonic splicing silencer sequences (Erkelenz et al., 2013), but can have either positive or negative effects on splicing when bound to

intronic sites. The activating or repressive activity of hnRNPs is largely dependent on their binding location, relative concentration and affinity to *cis*-elements (Fu & Ares, 2014). hnRNPs can also interact with other RBPs in a cooperative fashion but can also compete for the same binding sites. This was evidenced in the case of polypyrimidine tract-binding protein (PTB, also known as hnRNP I) and Quaking (QK) which were shown to regulate an overlapping network of 172 alternatively spliced cassette exons (Hall et al., 2013). However, some exons were repressed or activated by both (80) and some showed antagonistic regulation by PTB and QK (92).

1.3 The role of Tra2 proteins in the regulation of alternative splicing

Transformer-2 (*Tra2*) is an SR-like protein involved in alternative splicing originally identified in Drosophila melanogaster where it plays a key role in sex differentiation and spermatogenesis (Elliott, Best, Dalgliesh, Ehrmann, & Grellscheid, 2012). In humans, there are two distinct Tra2 genes, *TRA2A* and *TRA2B*, coding for Tra2α and Tra2β respectively (referred to as Tra2 proteins throughout this thesis). In fact, two copies of Tra2 are found in all vertebrates and the sequence is evolutionarily well conserved across species (S. Grellscheid et al., 2011). Tra2α and Tra2β share 63% amino acid sequence homology and have similar structural features (Elliott et al., 2012). They have particularly high sequence conservation across the RRM which accounts for the fact that their target sequences are similar, however most SR-proteins tend to bind GAA-rich targets (Long & Caceres, 2009). High throughput sequencing cross linking immunoprecipitation (HITS-CLIP) and RIP-seq have been used to show Tra2β commonly binds to the AGAAGA hexamer (S. Grellscheid et al., 2011; Uren et al., 2012). An interesting feature of the Tra2 proteins is the interaction between the two. Initially Tra2β was identified to autoregulate its own protein expression by promoting the inclusion of its own alternative exon 2. *TRA2B* exon 2 is a poison exon containing several in-frame stop codons. Inclusion of this exon in the mRNA of *TRA2B* leads to its degradation via nonsense mediated decay (NMD). Grellscheid et al. (2011) discovered that mouse *Tra2a* also contains a poison exon 2. This exon shows high sequence conservation across vertebrates and most interestingly is regulated by Sfrs10 (*TRA2B* in mice). Remarkably, the inclusion of *Tra2a* exon 2 in the mRNA was testis specific which correlates with an increased expression of Sfrs10. Subsequently, human *TRA2A* poison exon 2 was found to be regulated by Tra2β in a human breast cancer cell line (Figure 7)(Andrew Best et al., 2014). The interesting dynamic between the two proteins shows that Tra2β may have a master regulator role out of the two genes and is potentially more important for certain tissue such as the testis.

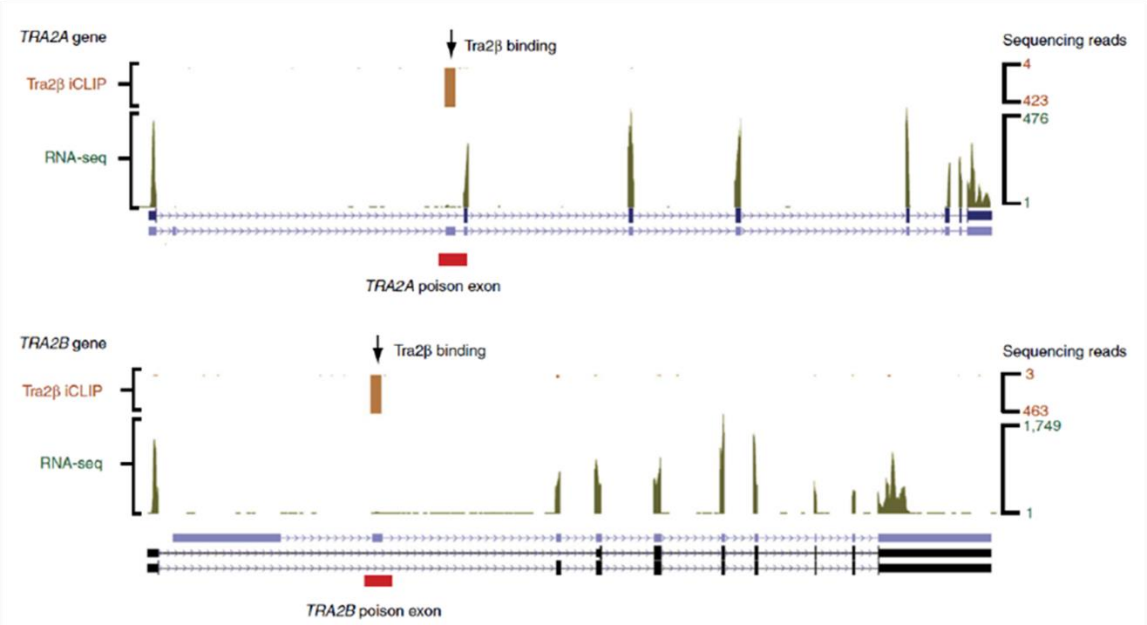

Figure 7 RNAseq and iCLIP data showing *TRA2A* and *TRA2B* poison exons. Vertical arrows indicate Tra2β iCLIP tags present on the poison exons for both *TRA2A* and *TRA2B*. However, RNAseq reads (green peaks) mapped on the human genome from the UCSC genome browser do not detect these poison exons due to nonsense mediated decay. RNAseq and iCLIP data derived from MDA-MB-231 cells. Image adapted from Best et al. (2014)

The TRA2B gene produces five mRNA isoforms including the isoform containing the poison exon described above. Only two of the other mRNAs produce protein isoforms: the full-length isoform and Tra2β-3 which misses RS-domain 1 (Tra2βΔRS1) (Figure 8) (Nayler, Cap, & Stamm, 1998). Although Tra2β may regulate its own expression levels via inclusion of the poison exon, it has not been observed to influence the expression of the Tra2β-3 isoform. The role of Tra2β-3 is not known but the expression of this isoform is conserved in vertebrates and is tissue specific, so it is likely to carry out some function (Cléry et al., 2011). The full-length isoform has also been described to regulate several other important exons.

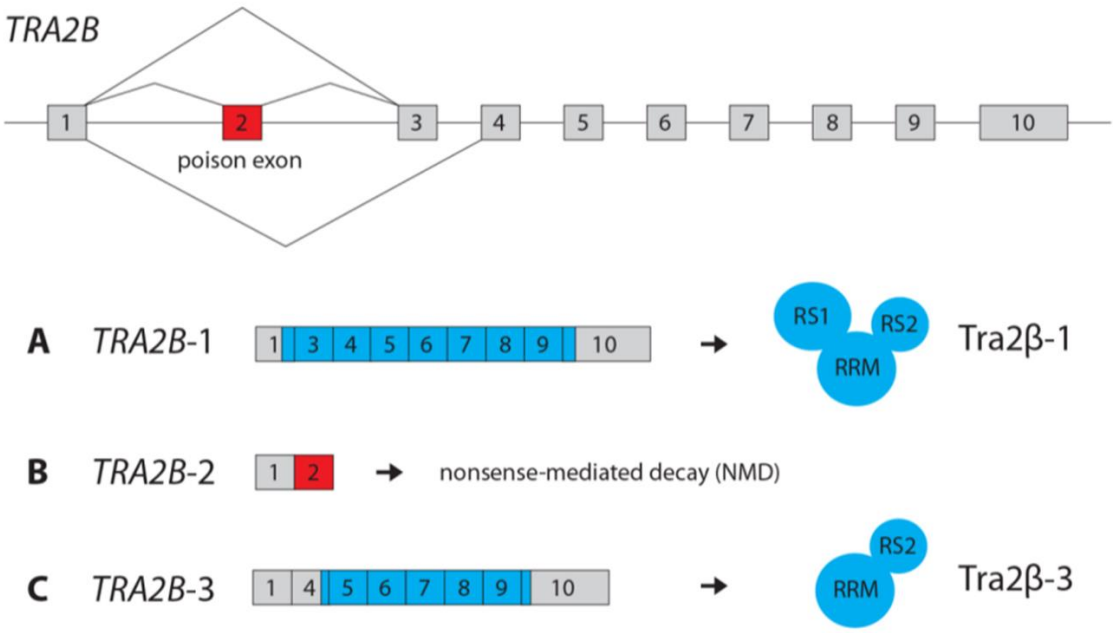

Figure 8 Alternative splicing events with the *TRA2B* gene. Highlighted in blue are open reading frames which produce two protein isoforms (A) and (B). The poison exon highlighted in red expresses an mRNA isoform targeted by NMD. Image adapted from Soilov *et al.*, 2004.

Tra2β has been shown to directly interact with RBMX (also known as hnRNP G), coded by the gene *RBMX*, to promote the inclusion of exon 7 in the *SMN2* gene (Hofmann & Wirth, 2002). This gene is particularly important for patients with spinal muscular atrophy who have non-functional *SMN1* alleles because normally *SMN2* expression can rescue neuronal function. However, exon 7 of *SMN2* is normally skipped and cannot produce a functional SMN protein. Tra2β can act to promote exon 7 inclusion and potentially restore in vivo function (Hofmann, Lorson, Stamm, Androphy, & Wirth, 2000). Alternative splicing of Homeodomain interacting protein kinase 3 (HIPK3) is also regulated by Tra2β (Julian P. Venables et al., 2005). The *HIPK3* exon T is a poison exon and is exclusively spliced in the testes suggesting that Tra2β is directly regulating *HIPK3* expression in the testis via alternative splicing of exon T (S. N. Grellscheid et al., 2011). Another testis specific exon regulated by Tra2β is the *Nasp*-T exon. Unlike *HIPK3*-T, *Nasp*-T is a coding exon that contains 975 nucleotides with approximately 37 Tra2β-binding sites (Elliott et al., 2012). The *Nasp* gene is essential for mouse development (Richardson et al., 2006), and the isoform including exon-T is associated with meiotic chromosomes and monitors double-strand break repair (Richardson et al., 2000). There is an overall tendency for Tra2β to regulate exon inclusion in a tissue-specific manner, particularly in the testis. An interesting find is the discovery that Tra2β, an alternative splicing regulator, is vital to the development of embryos in mice. Mende *et al*. (2010) found that double-Tra2β deletion is embryonic lethal (Mende et al., 2010). The exact mechanism behind this fundamental role in physiology is not well understood but it raises many questions about the function of Tra2β in

development. Interestingly, development of male germ cells continues into the adult which suggests that the reason Tra2β is important in development might be associated with regulation of the testis specific exons. The fact that Tra2β strictly controls alternative splicing in the testis and is essential in embryo development may suggest a role in cells undergoing rapid cell division.

1.4 Role of Tra2α/β and RBMX in cancer

Rapid cellular division is a hallmark of cancer. Previous reports have identified many mutations disrupting normal alternative splicing in cancer (H.-X. Liu, Cartegni, Zhang, & Krainer, 2001; Lukas et al., 2001). Likewise, recent investigations into the mechanisms that drive tumorigenesis have suggested that RNA binding proteins play a significant role in cancer by regulating post-transcriptional mRNA expression (Galante et al., 2009; M.-Y. Kim, Hur, & Jeong, 2009). Since alternative splicing applies to more than 95% of protein coding genes it is possible that disruptions of this mechanism can lead to the generation of an array of protein isoforms, some of which may contribute to tumorigenesis (Pajares et al., 2007). Involvement of Tra2β in lung squamous cell carcinoma (LUSC) and adenocarcinoma (LUAD), among other regulators of alternative splicing (Coomer, Black, Greystoke, Munkley, & Elliott, 2019), highlights the importance of alternative splicing in aggressive cancers. There are already several reports suggesting that Tra2 proteins are overexpressed in cancers (Fischer et al., 2004a; Gabriel et al., 2009; Watermann et al., 2006). Tra2β overexpression has been associated with poor prognosis in cervical cancer and LUSC patients (Sebestyén et al., 2016). A study of invasive breast cancer patients showed significant overexpression of Tra2β proteins compared to other SR-proteins (Figure 9). This suggests that Tra2β may play a significant role in breast cancer development through alternative splicing regulation.

Similarly, overexpression of RBMX has been correlated with expression of the pro-apoptotic Bax gene in breast cancer (Martínez-Arribas et al., 2006). Identification of lung cancer patients with truncating mutations in the *RBMX* gene suggests RBMX may act as a tumour suppressor (D. Zhang, Qu, Zhou, Wang, & Zhou, 2018). However, the general consensus is that RBMX, a hnRNP, functions as an RNA binding protein to regulate splicing. This has been previously demonstrated in conjunction with Tra2β in the co-regulation of *SMN2* exon 7 (discussed previously) and antagonistic regulation of *TPM3* (Nasim, Chernova, Chowdhury, Yue, & Eperon, 2003). The role of RBMX in disease is reviewed by (D. J. Elliott, C. Dalgliesh, G. Hysenaj, & I. Ehrmann, 2019).

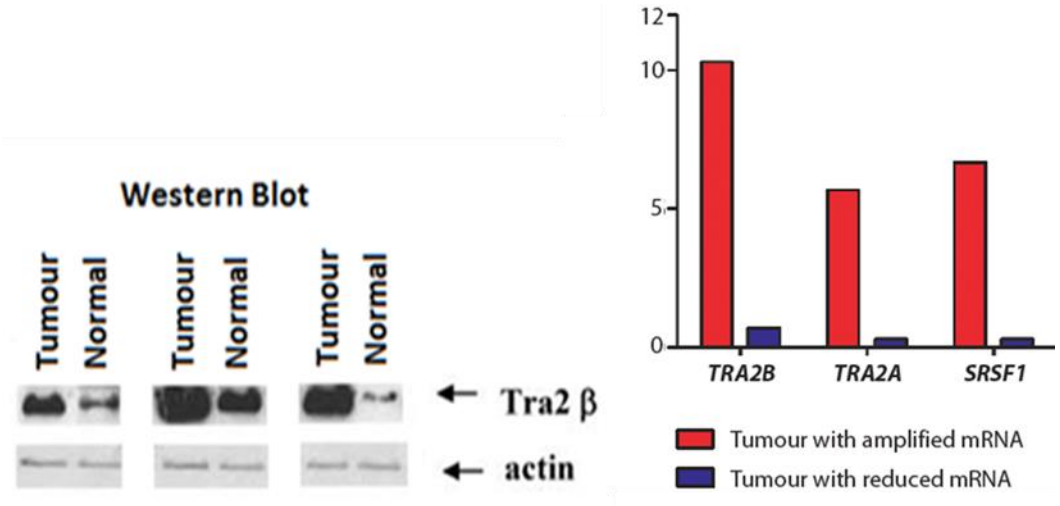

Figure 9 Expression of *TRA2A* and *TRA2B* in breast cancer. Both mRNA and protein expression of the SR-like proteins are significantly increased in breast cancer tissue. Image adapted from Watermann *et al.* (2006)

1.5 Background to this thesis

The work on this thesis in relation to Tra2 proteins is a continuation of a project initially undertaken by Dr Andrew Best. Dr Best's work was mainly focused on investigating Tra2β regulated exons in the MDA-MB-231 cell line, a model of invasive breast cancer. To identify such targets, he performed iCLIP (individual-nucleotide resolution Cross-Linking and ImmunoPrecipitation) analysis to identify RNA targets bound by Tra2β in MDA-MB-231 cells. Subsequently, Dr Best performed RNA sequencing of the MDA-MB-231 cells treated with negative control siRNA and siRNA targeting *TRA2A* and *TRA2B.* As shown on his Nature Communications publication, in order to observe a significant effect on the expression of different isoforms regulated by Tra2β, a knockdown of both Tra2 proteins was required (Andrew Best et al., 2014). The double knockdown was necessary as the absence of one homologue could be overshadowed by the overexpression of the other as a consequence of the negative feedback pathway between the two genes. Due to the similarities in the structure and amino acid homology of the two Tra2 genes, targets regulated by Tra2β could also be regulated by Tra2α. The RNAseq reads from control and Tra2 double knockdown samples were mapped to the human reference genome hg19 using Tophat 2. The results of the investigation revealed a set of 54 potential Tra2 targets, one of which was the *CHEK1* gene. This gene was shown to be directly regulated by Tra2β, which promotes exon 3 inclusion and leads to the production of full-length functioning *CHEK1* in cancer cells. The CHK1 protein is known to be an important checkpoint regulator that can control cell cycle progression and the Tra2 overexpression in breast cancer is pushing the cell through the cell cycle checkpoints by increasing CHK1 protein levels. The evidence so far suggests that *CHEK1* is one of the strongest cancer-driving targets being regulated by

Tra2 in breast cancer. However, the full profile of alternatively spliced targets regulated by Tra2 is potentially much larger. Depletion of Tra2 proteins in MDA-MB-231 cells lead to reduced cell viability (Andrew Best, James, Hysenaj, Tyson-Capper, & Elliott, 2016). Nuclear morphology of cells treated with siRNAs targeting both *TRA2A* and *TRA2B* showed distinct features of abnormal of chromosomal organisation. This indicates the role of Tra2 proteins extends well beyond the regulation of alternative splicing of *CHEK1.* In this project I utilise the data from the RNAseq and iCLIP experiments to investigate more potential targets of Tra2α/β in the MDA-MB-231 breast cancer cell line.

1.6 Research aims and objectives

The key objective of this thesis is to investigate the role of the RNA-binding proteins Tra2α, Tra2β and RBMX in the regulation of alternative splicing of essential genes.

The aim of this project was to more efficiently analyse the RNAseq data obtained from triple negative breast cancer MDA-MB-231 cells by a previous PhD student, Dr Andrew Best. In his work, Dr Best identified several important genes under the regulation of Tra2 proteins, particularly *CHEK1*. However, *CHEK1* overexpression did not rescue cells from death after Tra2α/β double knockdown (Andrew Best et al., 2014). In this project, I was prompted to robustly analyse the RNAseq data using multiple improved bioinformatics pipelines in order to identify novel alternatively spliced events regulated by Tra2α/β. I also investigated enriched Gene Ontology pathways that elude to the potential functional role of Tra2 proteins in the breast cancer cells.

Alternative splice changes regulated by Tra2 proteins induce the expression of alternative isoforms that can potentially alter protein structure and function. Therefore, we hypothesised that protein changes corresponding to these splice changes must be detected in order to confidently determine if these isoform influence breast cancer cell function and survivability. Following identification of alternative events that affect protein isoform expression of Tra2-regulated genes, I used modified antisense oligonucleotides to imitate the individual splice changes. Subsequently, I assessed the functional impact of the induced isoforms in breast cancer cell function and survival.

Regulation of alternative splicing is a complex process that involves the interactions of many RNA binding proteins. Tra2β has been known to interact with RBMX in the regulation of *SMN2* splicing, a clinically relevant gene (Cléry et al., 2011; Moursy, Allain, & Cléry, 2014). Therefore, I aimed to investigate the regulatory overlap between Tra2 proteins and RBMX using transcriptome-wide analysis of RNAseq data obtained from RBMX knockdown MDA-MB-231 cells. In this thesis, I analyse the first global analysis of RBMX-regulated alternative events with the aim of identifying mechanistic patterns or splice regulation by RBMX and affected genes potentially important to the biology of breast cancer cells.

Chapter 2: System wide screening of alternative isoforms regulated by Tra2α and Tra2β in breast cancer

2.1 Introduction

Tra2 proteins are ubiquitously expressed in the human body. However, increased expression of both Tra2α and Tra2β observed in several cancers raises a question to the role of these genes in the development and maintenance of cancer cells. The knockdown of both Tra2 proteins in the MDA-MB-231 breast cancer cell line had severe effects on the cell survivability and clearly affected the splicing of several key genes (Andrew Best et al., 2014). Particularly, the alternatively spliced exon 3 of Checkpoint Kinase 1 (*CHEK1*) was shown to be directly bound and regulated by Tra2β. In turn, skipping of exon 3 led to loss of CHK1 protein, and a parallel increase in DNA damage monitored by γH2AX accumulation. Direct siRNA depletion of *CHEK1* also caused an accumulation of γH2AX levels. However, cell rescue experiments using tetracycline induced *CHEK1* overexpression did not restore the viability of cells treated with siRNA against *TRA2A* and *TRA2B*. This was a clear indication that Tra2α/β are likely to be regulating other genes essential for cell survival.

The full scope of Tra2-regulated alternative events might not have been exposed by the original analysis of the RNAseq. A potential limiting factor could be the analysis power provided by the bioinformatics tool DexSeq. While DexSeq is a package specifically developed to detect alternative splicing events, there are many evident drawbacks. DexSeq is a bioinformatics package focused on finding differential exon usage by using RNA-seq exon counts. The ratio of the reads overlaying an exon and the reads mapped to the flanking exons of the same gene is considered when calculating exon usage. This difference is defined as differential exon usage (similar to Percentage Splicing Inclusion, PSI, which is often used to calculated splice site usage in RT-PCR experiments) and considers the total read counts over each exon in the calculations. While the method takes into consideration differences from biological replicates, there are several drawbacks that may have prevented our initial investigations from identifying the full range of target genes regulated by Tra2 proteins. Firstly, the DexSeq algorithm relies on read counts mapped over predefined exons using a binning method. Reads that map onto intronic regions or cryptic exons that have not been defined in the reference genome will not be used in the analysis. This limits the potential of DexSeq to identify novel cryptic exons or intron retention changes. Secondly, the PSI calculations rely purely on the number of reads overlapping with an exon with respect to all reads falling on the whole gene. Therefore, changes in gene expression may present false changes in alternative exon usage. Finally,

the exon-exon junction reads (reads that map on two exons while skipping the spliced sequence in between) are simply used to add to the total count of reads for each respective exon they map to. Such simplistic sorting of junction reads devalues the information they provide.

Having such drawbacks in mind we thought to re-analyse our RNAseq data using other bioinformatics pipelines that have sought to tackle these issues. Mixture of ISOforms (MISO) and Modelling Alternative Junction Inclusion Quantification (MAJIQ) are two bioinformatics pipelines that implement exon-exon junctions in the analysis of each splice junction. Moreover, both MISO and MAJIQ will consider reads that map to non-exonic regions and so are able to identify splicing patterns involving cryptic exons or intron retentions (Figure 10). While the packages have similarities and are improvements on DexSeq, MAJIQ is much more recently developed pipeline. The developers, Prof Yoseph Barash's bioinformatics group that collaborated on this project, are continuously improving the pipeline to increase sensitivity and reduce false positive calls. In addition, they have developed Voila, a visualisation package that creates interactive summary files with gene splice graphs, local splicing variations (LSV) and their quantification (Vaquero-Garcia, Barrera, Gazzara, González-Vallinas, et al., 2016).

The benefit of analysing our data using multiple additional bioinformatics pipelines is two-fold. Firstly, we can comprehensively identify alternatively spliced essential genes regulated by Tra2 proteins. This is would help identify other genes/isoforms responsible for the dramatic effect of Tra2 double knockdown. Secondly, we can identify the full landscape of Tra2-regulated alternative isoforms. This is an important outcome of the analysis that can hint to the role of Tra2 proteins in the long term. While investigating the effects of strong changes in essential genes is important to evaluate the immediate effect of Tra2-protein knockdown in the cells, exploring the full range of target genes regulated by Tra2 can direct us to their normal role in the cell and potential function in both normal biology and carcinogenesis.

Gene Ontology enrichment analysis will be a useful tool to highlight important biological and cellular pathways regulated by Tra2 proteins. In this chapter, we make use of GO terms associated with Tra2-regulated genes identified by all three bioinformatics pipelines. We also investigate the effect of some of the splice changes at the protein level to determine a role of Tra2 proteins in GO-highlighted cellular pathways.

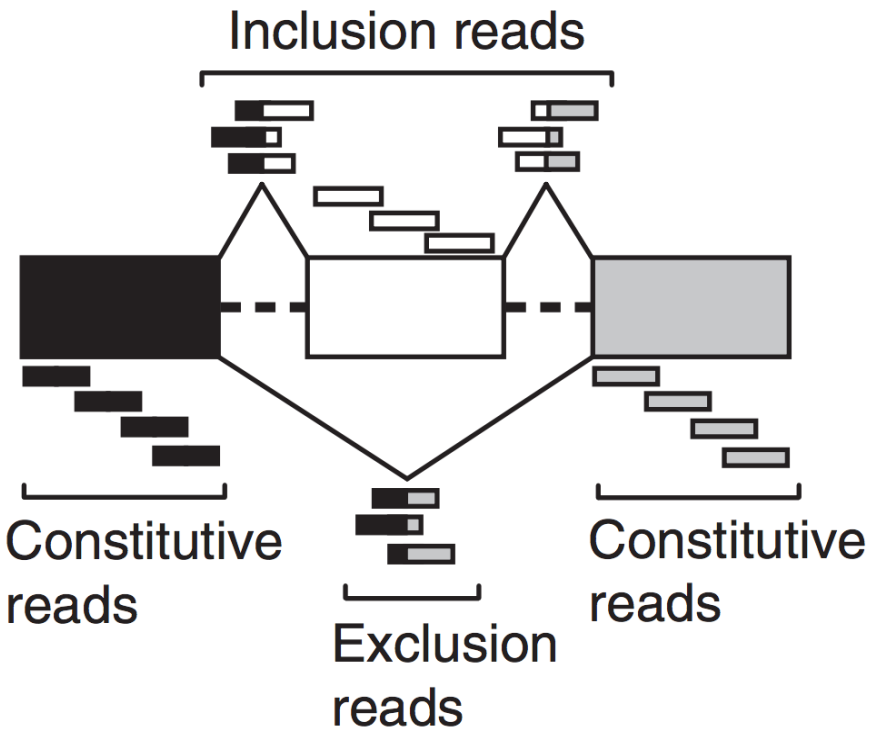

Figure 10 RNAseq reads overlaping the splice junctions are utilised by both MISO and MAJIQ in determining the isoform ratios. Image adapted from MISO, Software documentation (https://miso.readthedocs.io/en/fastmiso/)

2.2 Chapter aims and objectives

The main objective of this chapter is to understand the role of Tra2α/β proteins in the cell, and how they might contribute to the development of breast cancer. Here we aim to:

Interrogate our previous RNAseq data using new bioinformatics programmes to determine if there are additional Tra2β-dependent target exons in the MDA-MB-231 cell line.

Investigate the composition and predicted effects of alternative splicing events regulated by Tra2 proteins in MDA-MB-231 cells.

Investigate the Gene Ontology enrichment data to highlight cellular/biological pathways under the potential regulation of Tra2 proteins.

Identify candidate genes and test their protein expression changes upon Tra2 double depletion.

Investigate the effect of Tra2 splicing control on protein expression from Tra2-regulated genes.

2.3 Materials and methods

2.3.1 Bioinformatics analysis of RNAseq data

The RNAseq data published originally in the nature communications paper by Best *et al.* 2014 was analysed using DexSeq and MISO (this analysis was kindly performed by Mr Yaobo Xu at the Institute of Genetic Medicine). In collaboration with the bioinformatics team based at the University of Pennsylvania lead by Prof. Yoseph Barash we reanalysed the data with a more comprehensive bioinformatics tool named MAJIQ, which utilises a similar approach to MISO to identify alternative events within the RNAseq data (Vaquero-Garcia, Barrera, Gazzara, González-Vallinas, et al., 2016). RNAseq data analysis using MAJIQ was performed by Matthew Gazzara at the University of Pennsylvania. MAJIQ provides in depth detail regarding each local splicing variation (LSV) and utilises improved algorithms to identify alternative events based on a larger database of predefined exons. Both MISO and MAJIQ analysis methods use a 20% change in percentage splicing inclusion (PSI) as a minimum cut-off.

To identify biological pathways that are potentially regulated by the Tra2 proteins we used Gene Ontology (GO) analysis on the list of targets that showed to be alternatively spliced upon Tra2 double depletion. The GO enrichment analysis was performed by Dr. Katherine James using all three datasets (DexSeq, MISO and MAJIQ) separately and combined.

2.3.2 Cell culture

The breast cancer cell line MDA-MB-231 was maintained in Dulbecco's Modified Eagle Medium with 10% fetal bovine serum and 1% Penicillin and Streptomycin antibiotics. The cells were incubated at 37°C with 5% CO_2 and stored in 75 cm^2 tissue flasks. Cells were split regularly at 4-5 day intervals, at a 1 in 5 ratio. The cell line was provided to me by Dr Andrew Best but was originally purchased from the American Type Culture Collection.

2.3.3 siRNA knockdown

Tra2α and Tra2β transient double knockdown was established using siRNAs targeting *TRA2A* (Ambion IDs: s26664) and *TRA2B* (Ambion IDs: s12749) transfected with Lipofectamine RNAiMax transfection protocol (ThermoFisher). Negative control cells were transfected with scramble siRNA (Ambion Cat#: 4390843). Cells were transfected with siRNA transfection mixtures diluted in 150µl of Opti-MEM using 12µl of 10µM negative control siRNA or 6µl of 10µM siRNA targeting TRA2A and 6µl of 10µM siRNA targeting TRA2B simultaneously. Each transfection mixture was then combined with 6µl of Lipofectamine RNAiMAX reagent (ThermoFisher) which was diluted in 150µl of Opti-MEM. The combined mixture was incubated for 5 minutes at room temperature before transfection. MDA-MB-

231 cells were seeded on 6-well plates 24 hours prior to transfection at a confluence of approximately 1x10^6, measured using a haemocytometer. Upon siRNA transfection cells were for the incubated for 72 hours at 37°C, after which they were harvested using trypsin. Samples were split to provide RNA (cells treated with Trizol) and protein (cells treated with 2x Protein Loading buffer).

2.3.4 cDNA and PCR quantification

The RNA from cells treated with siRNA against *TRA2A* and *TRA2B* was extracted using standard Trizol RNA extraction (Life Technologies) following manufacturer's instructions. In each case cDNA was synthesized from 500ng total RNA in 10µl reactions using Superscript VILO cDNA synthesis kit (Invitrogen) following manufacturer's instructions. To analyse the splicing profiles of the alternative events primers were designed using Primer 3 Plus and the predicted PCR products were confirmed using the UCSC *In-Silico* PCR tool. Primers were designed to have a melting temperature of 60°C (+/- 1°C) often including an internal screening primer (Full list of primers for all chapters in the Appendix). All PCR reactions were performed using GoTaq G2 DNA polymerase kit from Promega. All PCR products were examined using the Qiaxcel capillary electrophoresis system (Qiagen) and PSI was calculated using concentration from three biological replicates of control and Tra2 double knockdown cels. The following formula was used to derive PSI:

$$\frac{\text{Conc of 'exon included' PCR band (ng/µl)}}{\text{Conc of 'exon included'+'exon exluded' PCR bands (ng/µl)}} x100 = PSI\ (\%)$$

End point PCR reaction mix

5x Reaction buffer	4µl
Forward Primer (10µM)	1µl
Reverse Primer (10µM)	1µl
dNTPs (10µM)	0.4µl
GoTaq Polymerase	0.1µl
dH_2O	12.5µl
cDNA	1µl
TOTAL	20µl

Thermocycler programme

1 Heat Activation	95°C	5 min
2 Denaturation	95°C	20 sec
3 Annealing	58°C	30 sec
4 Extension	72°C	60 sec (Cycle to step 2 x29)
5 Final Extension	72°C	10 min
6 Cooling	4°C	∞

2.3.5 Western Immunoblotting

Cells dissolved in 2x Protein Loading buffer (4% SDS, 20% glycerol, 200mM DTT, 0.01% bromphenol blue and 0.1 M Tris HCl, pH 6.8) were used to obtain protein samples from each siRNA knockdown case. Samples were briefly sonicated and then denatured at 100°C for 5 minutes. Protein was separated using 10% SDS-polyacrylamide gel and then transferred onto PVDF membrane (Hybond-P, GE Healthcare) using semi-dry transfer methods. Membranes were blocked for 1 hour with 5% non-fat milk, and in some cases 1% horse serum, diluted in Tris-Buffered Saline with Tween 20 (TBS-T). Often the membrane was incubated overnight at 4°C (or at room temperature for 1 hour) with the primary antibody diluted in blocking solution at ratios suggested by the manufacturer. Membranes were then washed with TBST prior to the addition of the species-specific secondary antibody conjugated to horseradish peroxidise. Enhanced chemiluminescent (ECL) Prime Western Immunoblotting Detection kit (Amersham) was used to activate the horseradish peroxidise for film development. Quantification of western blot bands was done using ImageJ (Fiji software). A list of antibodies used is shown in the table below.

Primary Antibodies	Cat. number	Dilution
CENPQ	Rockland, 200-401-B48S	1/1000
SUV39H2	Proteintech, 11338-1-AP	1/2000
ZCCHC11	Proteintech, 18980-1-AP	1/1000
Sin1 (MAPKAP1)	Cell Signalling, D7G1A	1/2000
Tra2β	Abcam, ab31353	1/2000
β-Actin	Sigma-Aldrich, A5441	1/2000
α-Tubulin	Sigma-Aldrich, T5168	1/2000
H3K4me2	Millipore, 07-030	1/2000
H3K4me3	Diaogenode, MAb-003-050	1/1000
H3K27me3	Diagenode, pAb-069-050	1/1000

2.3.6 Histone H3 methylation state

Histone H3 methylation profiling was performed by Western blotting by Dr. Alejandro Villarreal at Freiburg University, Germany (Prof. Tanja Vogel laboratory). For this experiment MDA-MB-231 cells were treated with siRNA using same transfection protocol as above but were incubated for 72 hours and 108 hours. Protein extraction was performed using RIPA buffer containing protease inhibitor cocktail.

2.4 Results

2.4.1 MISO and MAJIQ identify over 500 additional changes upon Tra2 double depletion

2.4.1.1 Mixture of ISOforms, MISO

The computational analysis using MISO was performed by Yaobo Xu (Newcastle University) and predicted 84 alternative events that are potentially regulated by Tra2 proteins, 6 of which were previously identified by DexSeq. The analysis used a 20% dPSI minimum cutoff. 69 out of 84 (82%) alternative events identified by MISO were cassette exons (Figure 11, and Appendix Table 1). These include 10 cases of mutually exclusive exons. An example of the mutually exclusive exon splicing is in the *COX10* gene. COX10 encodes an important protein member of the Cytochrome C Oxidase assembly. The mutually exclusive splice event regulated by the Tra2 proteins involves the exclusive use of exons 3 and 4 of COX10. The UCSC genome database predicts that each of these alternative isoforms can produce translatable mRNA that encode for proteins of 251 amino acids (exon 3 included) and 226 amino acids (exon 4 included). To determine if there are any evident changes in the structures of these two isoforms, I ran the amino acid sequences through the Pfam protein analysis online tool found at https://pfam.xfam.org/ (El-Gebali et al., 2019). Both sequences contain the UBiA prenyltransferase protein subunit which is predominantly unaffected in either isoform. Since there is little change in the predicted structure of the protein it is difficult to anticipate the effect of the isoform switch in the assembly of Cytochrome C Oxidase. See Appendix, Table 2 for a list of MISO identified targets and the anticipated Pfam domains affected.

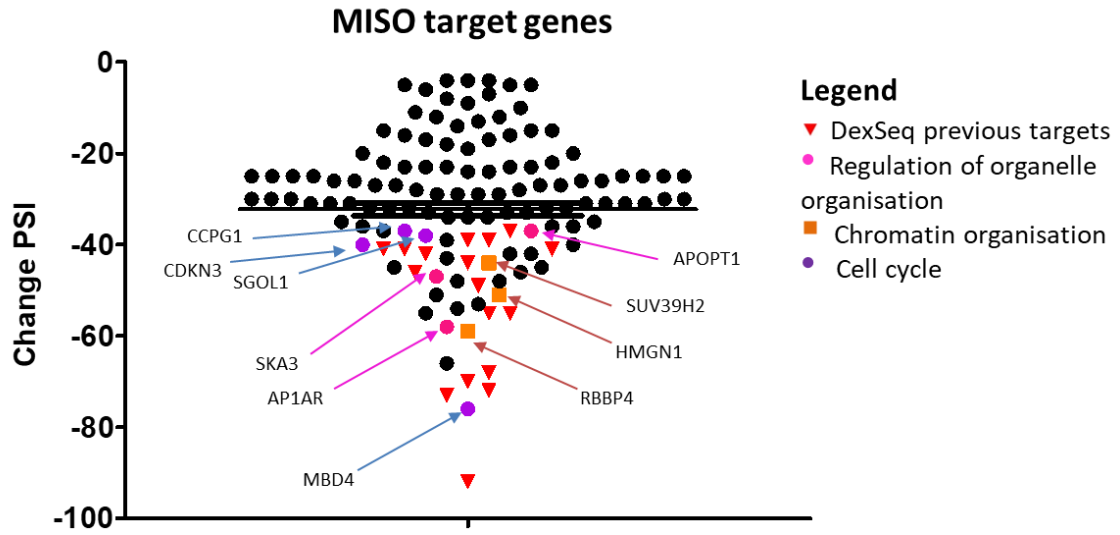

Figure 11 Tra2-regulated alternatively spliced genes identified by MISO. Percentage Splicing Inclusion of alternative events identified by MISO showing several novel targets from key cellular pathways.

The use of exon-exon junction reads enabled MISO to clearly identify alternative use of 5’ and 3’ splice sites in 15 different genes (4 and 11 respectively). The identification of alternatively selected splice sites was novel compared to the results from DexSeq. One example of the use of an alternative 3’ splice site is found in the *Retinoblastoma binding protein 4* (*RBBP4*) gene (Figure 12). *RBBP4* encodes a ubiquitously expressed nuclear protein that is involved in histone acetylation and chromatin assembly. The alternative 3’ splice site used in the Tra2 double knockdown cells is located in the 3’UTR region of the last exon. The iCLIP data shows a large cluster of Tra2β binding sites near the control-expressed 3’ splice site. This is strong evidence suggesting Tra2 proteins are directly regulating the use of this alternative splice site. The alternative last exon introduces a different amino acid sequence at the C-terminus of the protein that leads to a shorter protein. However, Pfam domain predictions show the shorter protein still contains all five WD40 domains, a domain often found in transcription factors and signal transduction proteins.

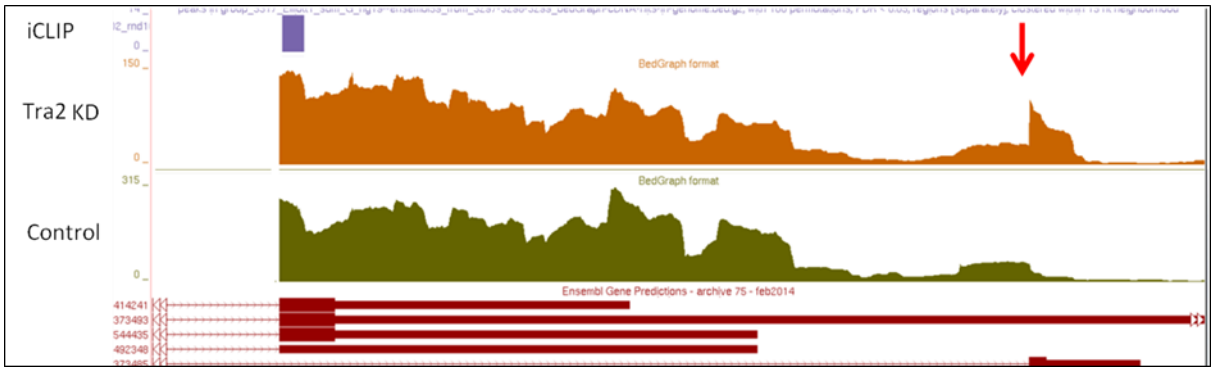

Figure 12 Alternative splicing of RBBP4. RNAseq data from the Tra2 knockdown (top) and the control (bottom) samples, as viewed on the UCSC browser, show a clear switch in the 3' splice site usage (indicated by red arrow) corresponding to predicted isoforms by Ensembl

To verify the results from MISO we compared the splicing profiles between MDA-MB-231 cells transfected with a negative control siRNA or *TRA2A* and *TRA2B* siRNAs. Total RNAs were extracted and analysed by reverse transcriptase polymerase chain reaction (rt-PCR) using primer pairs or other primer combinations able to detect the expected isoforms. In total, we designed primers for and tested 34 novel alternative events identified by MISO. In 31 (91%) of these we were able to confirm the alternative changes predicted by RNAseq (Appendix, Table 1). While validation by PCR is important, particularly in cases where there are no iCLIP tags in the vicinity of the alternative events, we are confident that the calls made by MISO are pointing to real changes in the RNA splice isoform ratios and experimental confirmation of the remaining targets would not prove anything new.

2.4.1.2 Modelling Alternative Junction Inclusion Quantification, MAJIQ

The bioinformatics team of Prof Yoseph Barash at the University of Pennsylvania analysed our RNAseq data using their analysis package, MAJIQ. A total of 679 local splicing variations (LSV) were found in 474 genes - note that one alternative event may have more than one LSV, each indicating exon splicing changes observed from the point of view of the flanking exons. 15 of these genes were found to have 2 alternative events being regulated by Tra2 proteins. Much like MISO, most targets identified by MAJIQ were alternative cassette exons (502 out of 679 LSV, 74%). MAJIQ was the only pipeline able to confidently identify 35 repressed exons (all previously identified splice changes have negative PSIs in response to Tra2α/β depletion, while these 35 exons had positive PSI changes) (Figure 13). MAJIQ was also capable of identifying changes in intron retention for 19 genes, a role that Tra2 proteins have not been previously known for.

The extensive use of RNAseq reads overlapping exon junctions allowed MAJIQ to identify a range of complex alternatively spliced events. 334 out of 679 LSVs found by MAJIQ are considered as complex due to the involvement of three or more splice junctions. Complex LSV can pose a problem when trying to determine the correct splicing event. As an example, consider an alternatively spliced exon that also contains an alternative 3' splice site within its coding sequence. In this example LSV a dPSI change of -60% could be calculated based on a -50% change due to the exon being fully skipped and -10% due to the use of the alternative 3' splice site. Figure 14a shows the alternative splice change predictions by MAJIQ for the *MBD4* gene. The multiple exon-exon junctions in play between exon 2, 3 and 4 of the gene can create a range of possible alternative isoforms. Indeed, RT-PCR shows several faint bands in the Tra2α/β depleted cells (Figure 14b). An analysis of the cumulative distribution function (CDF) of the dPSIs for the third junction in the 334 complex LSVs shows that in approximately 10% of the cases the 3rd junction changes by 10% (the complex LSV analysis was performed by Matthew Gazzara, University of Pennsylvania). The visualisation programme Voila helps us understand better the complex splice events and directs us in designing the right primers to test such cases. Interestingly, some complex alternative events occur over multiple exons. We were able to identify and validate by PCR seven such events in the following genes: *GINS1, INO80C, OARD1, SLC35A1, UBAC2, UIMC1* and *XPA*. We will investigate a few of these cases in more detail later on in the chapter.

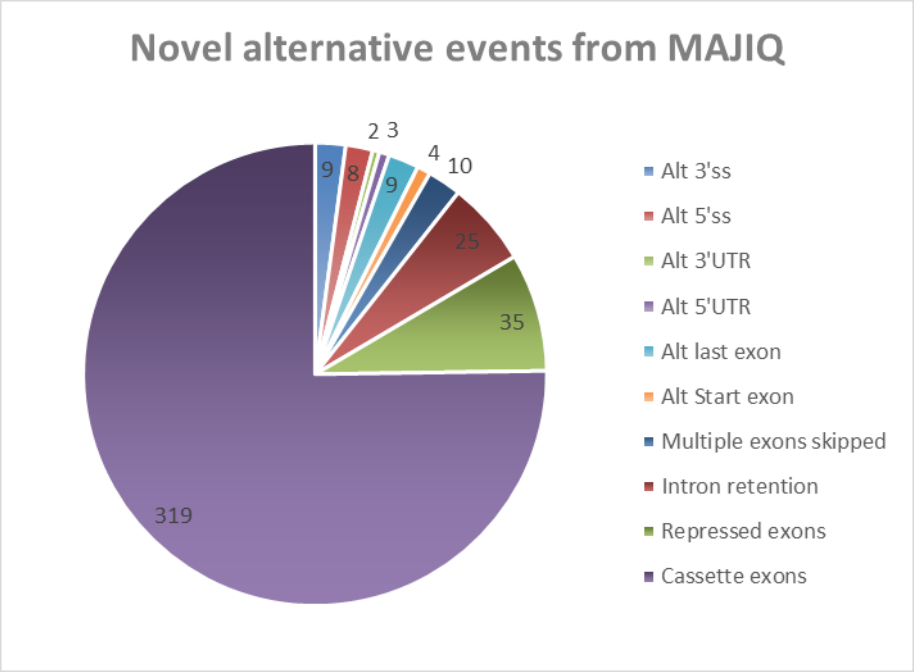

Figure 13 Alternative events identified by MAJIQ include a variety of complex events, however the majority of Tra2α/β targets are cassette exons

As with MISO, we set out to validate a proportion of the alternative splicing events identified by MAJIQ using RT-PCR and capillary gel electrophoresis. In total we tested the splicing profiles of 94 genes encompassing a range of alternative events. 83 (88%) of the events tested by PCR confirmed the results predicted by MAJIQ (Appendix table 2). However, the dPSI values obtained by our PCR results (both for MISO and MAJIQ) were not always the same, often smaller than those predicted by the pipelines. The discordance in the dPSI value measured could be due to a range of confounding variables and differences between processes used to obtain the results. One obvious reason is the different experimental methods; MISO/MAJIQ use data from RNAseq whereas our validation is based on the PCR band intensity measured by capillary gel electrophoresis. Primer affinity and amplification bias of shorter amplicons can heavily affect the dPSI results from PCR. Furthermore, the knockdown efficiency in the samples sent for sequencing may be slightly different from that obtained in the samples used to perform the validations. All things considered however we were able to validate approximately 90% of the MISO and MAJIQ targets tested, all showing changes in the expected direction which gives us confidence that the bioinformatics analysis of the RNAseq data is identifying real alternatively spliced isoforms regulated by Tra2 proteins.

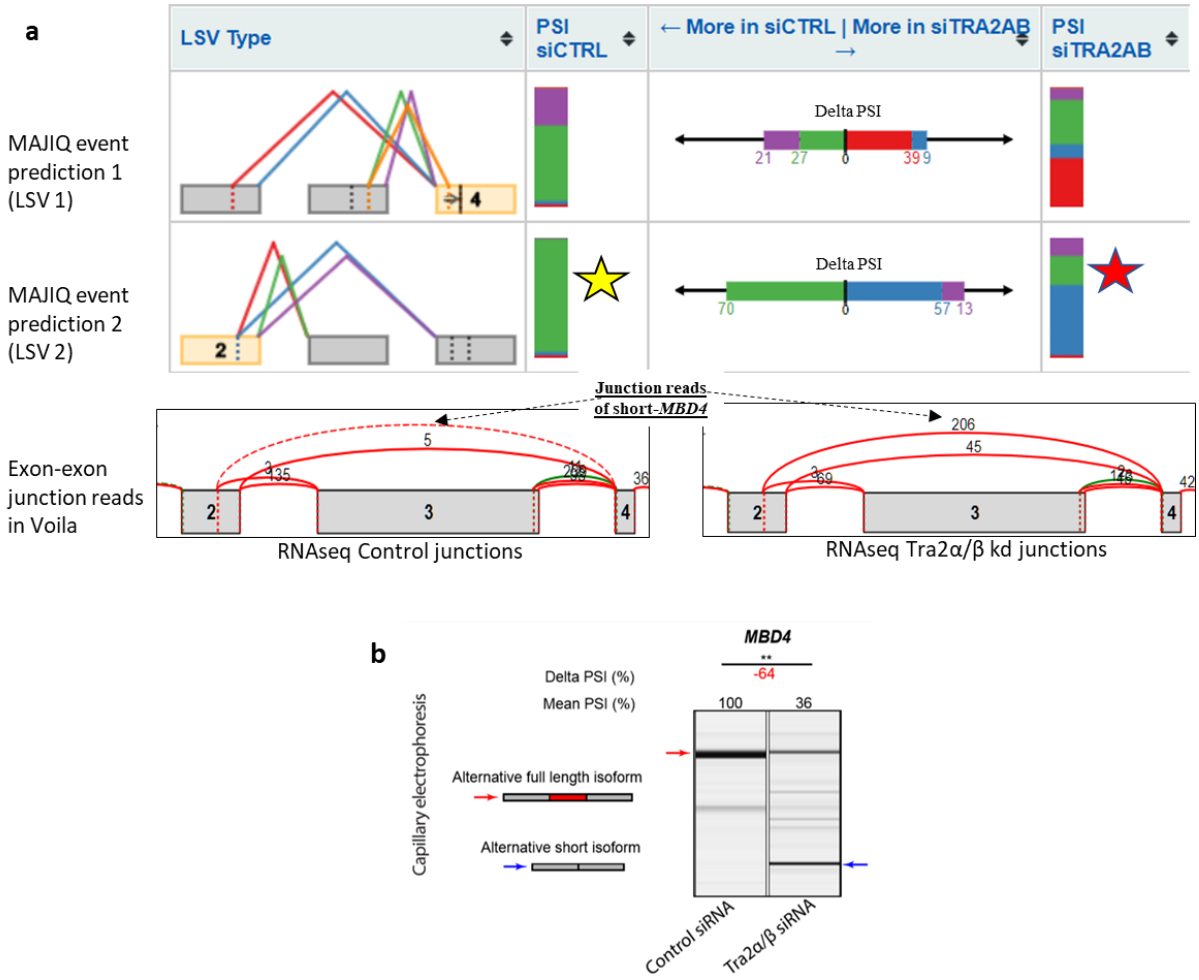

Figure 14 Complex LSV example in the *MBD4* gene. Two Local Splicing Variants (LSVs) identified by MAJIQ shown in part a) are in fact describing the same exon-exon junction reads but calculate the PSI values differently. In LSV2 of this example the proposed main isoform uses the downstream splice site of exon two which connects to exon 3 (green line, green bar). The green bar (next to yellow star) represents the use of this junction, which is then drastically reduced (-70%) in the knockdown cells where alternative splice junctions are used (bars next to red star). Part b) of the figure shows a capillary gel electrophoresis of the RT-PCR for *MBD4*. Several smaller bands are visible in the Tra2α/β knockdown cells suggesting a multitude of alternative isoforms may be present in these conditions. ** represents $p<0.01$, calculated by Student's *t*-test from three biological replicates.

2.4.1.3 Tra2 proteins are predominantly involved in the regulation of alternative splicing

The Tra2 proteins are SR-like proteins and were classically known to be involved in alternative splicing alone (Tacke, Tohyama, Ogawa, & Manley, 1998). However, the cassette exons identified by DexSeq, and reported by Dr Best (Andrew Best et al., 2014), included a large number of constitutive exons being spliced out of the mRNA in the absence of Tra2 proteins - 26 out of 54 exons have a control PSI of 95% or higher. Among the target cassette exons identified by MISO only 19% are constitutive exons, less than half the rate of constitutive exons identified by DexSeq. SImilarly, the MAJIQ-identified differentially spliced constitutive cassette exons only comprise 17% of the total alternative events (Figure 15). These findings confirm that Tra2 proteins may be involved in the

regulation of constitutive exons but confirms their role as being predominantly involved in regulating alternative splice.

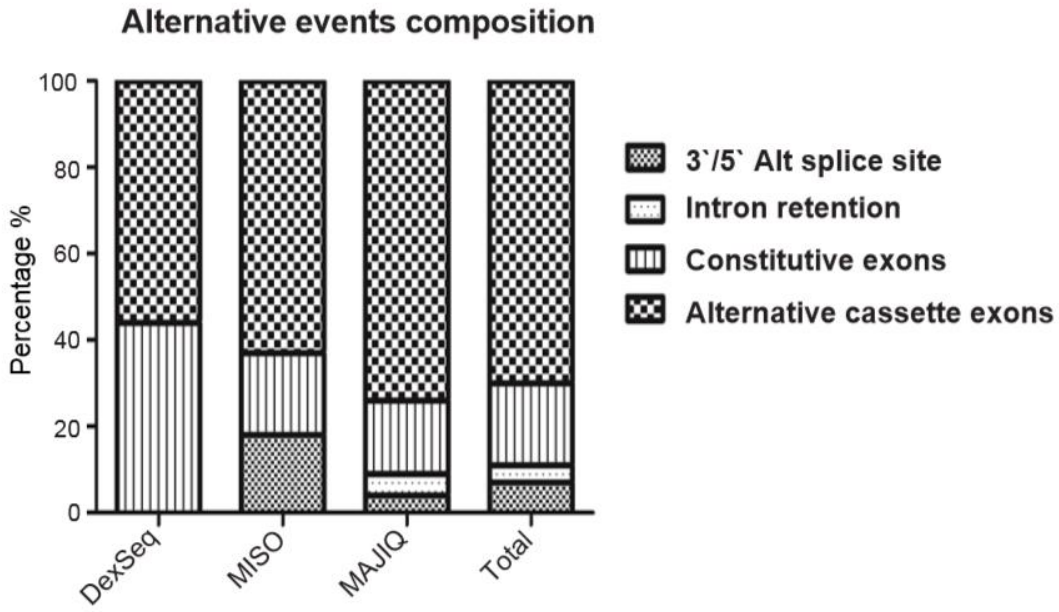

Figure 15 Data combined from all three bioinformatics pipelines confirm that Tra2 proteins predominantly regulate splicing of alternative cassette exons, but also highlight other mechanistically interesting events

2.4.1.4 Combined data from all three bioinformatics analysis shows that only a small proportion of alternative events are commonly identified by all three pipelines

Rerunning the RNAseq data through MISO and MAJIQ helped identify a total of 502 novel differentially spliced targets. Figure 16 shows the overlap of target genes identified by each bioinformatics tool. Only 49 novel alternative events were commonly identified by MISO and MAJIQ (Figure 16). Although this is only 10% of the total number of novel events, it accounts for 58% of all the exons identified by MISO. Despite the differences this overlap suggests that MAJIQ is capable of identifying the majority of MISO targets, but the opposite is less likely. To our surprise I found that only 5 events were identified by all three bioinformatics tools (*CHEK1* ex3 included). This low level of concordance between the three packages is a clear indicator of the differences in the algorithms used to investigate differential splicing. However, MAJIQ is clearly a more powerful analysis tool over the others as it was able to detect almost 10 times more targets, with a high validation rate, including a wide range of complex alternative events that reveal new mechanisms by which Tra2 proteins may be regulating splicing. Appendix, Table 4 shows all previously identified DexSeq alternative events that are also identified by at least one of the new bioinformatic tools..

Despite the clear computational advantages of MISO and MAJIQ in analysing the RNAseq data, 37 out of the 54 exons identified by DexSeq were not detected. One reason for this could be the thorough use of the iCLIP data in conjunction with the results from DexSeq. This allowed the identification of several exons that were showing lower than 20% PSI changes, which is the minimum cutoff used by both MISO and MAJIQ. Although some targets identified by MISO and MAJIQ do not have iCLIP tags, the RNAseq data clearly shows they are alternatively spliced in Tra2 double knockdown cells. These changes in alternative splicing can still be attributed to the direct effect of Tra2 proteins (although with lower confidence) since the iCLIP tags are only showing binding sites for Tra2β.

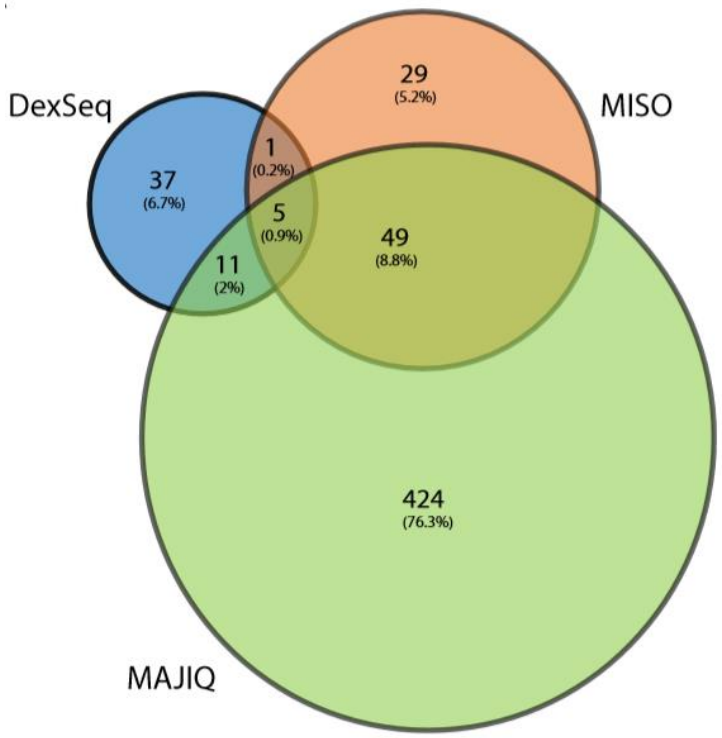

Figure 16 Tra2 regulated targets identified by bioinformatics. The total number of alternative events identified by all 3 bioinformatics packages, indicating only 5 events were commonly detected. Percentages represent the portion of targets relative to the total number of alternative events identified by all 3 bioinformatics packages.

2.4.2 MISO and MAJIQ reveal an array of mechanistically complex splicing events that include alternative 3' and 5' splice sites, intron retentions and multiple exon skipping events

2.4.2.1 Tra2α and Tra2β repress the splicing of 35 cassette exons

As SR-like proteins Tra2-α and Tra2-β can bind exonic sequences known as exonic splicing enhancer signals. As such Tra2 proteins have been previously described to enhance the splicing of the target

exons they bind to. We were able to identify target cassette exons that are being repressed by Tra2α and Tra2β. MISO identified a few targets showing positive dPSI but they were low confidence hits with dPSI changes below the 20% cutoff. MAJIQ identified 35 target exons that were being regulated by Tra2 proteins and showed increased PSI in the Tra2 knockdown cells.

In order to understand the mechanism behind the regulation of these repressed exons the binding patterns of Tra2β were analysed based on the iCLIP data. We utilised the previously reported and validated non-changing control exons (Best et al. 2014) as our baseline control cassette exons, and compared the iCLIP patterns over constitutive, activated and repressed cassette exons identified by MAJIQ. Figure 17 shows the total iCLIP tags counted over the upstream, target and downstream exons for each of the categories. Notice that not all the exons identified in each category have iCLIP tags associated with them, hence the number of exons used for this analysis is different to the total number of exons identified in each group. In order to test the distribution of Tra2β binding sites we used chi-square test for each category under the null hypothesis of no correlation between iCLIP tags and target exons. The results show significant abundance of Tra2β binding sites within target exons of activated (p-value 0.009) and constitutive (p-value 0.007) exons. Moreover, repressed exons are shown to have significantly fewer iCLIP tags (p-value 0.03) in the target exons relative to the flanking upstream/downstream exons, whereas control non-changing target exons show no significant correlation between number of iCLIP tags and target exons. Our data suggests that Tra2β may be repressing these exons by binding onto downstream/upstream exons. A study by Ghigna et al. 2005, previously reported their findings on the regulation of exon 11 of Ron, an important tyrosine kinase receptor (Ghigna et al., 2005). Their data showed that the strength of the enhancer signal in the downstream exon 12 bound by the splicing factor SRSF1 (SF2/ASF) would parallel the skipping of exon 11. It is possible therefore that Tra2 proteins could repress upstream exons in a similar manner, by binding to strong enhancer signals.

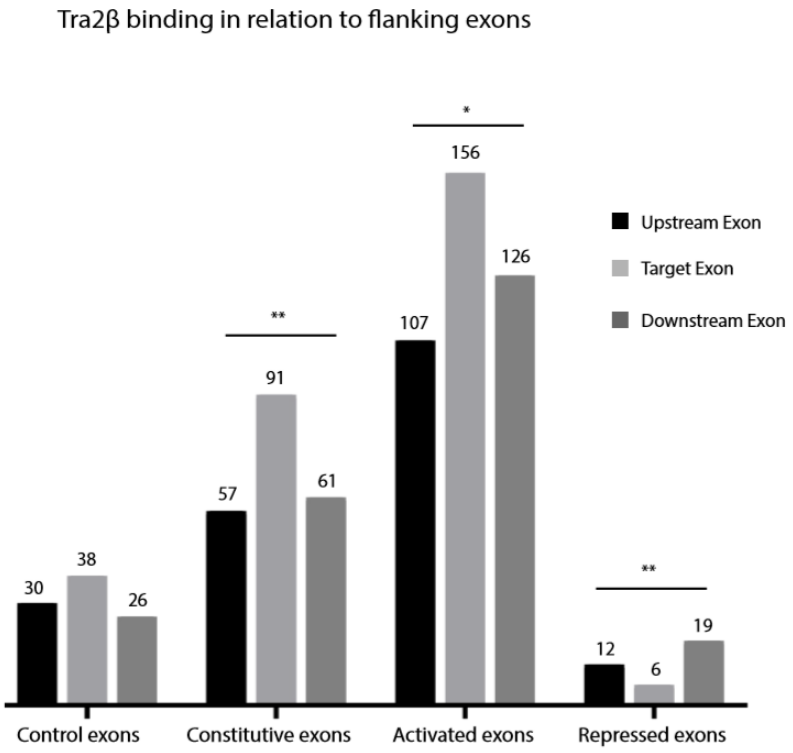

Figure 17 Bar chart indicating the relative presence of Tra2β iCLIP binding tags for constitutive, activated and repressed exons. Chi square test p-values *p<0.05 and **p<0.01

2.4.2.2 Tra2 depleted cells show changes in the inclusion of several introns/bleeding exons

Intron inclusion is not a common occurrence in the transcriptome, and alternative splicing of introns is often overlooked. In order to detect changes in intronic regions splicing analysis packages must consider reads that map onto intronic regions and junction reads that overlap junctions between exons and introns. In our data, MAJIQ was the only pipeline capable of detecting changes in intron retentions. The analysis identified 19 target genes with delta percentage intron retention (dPIR) over 20%. Only 3 target genes, *NT5C3A*, *KIAA0141* and *PTPN2*, were found to contain introns that are repressed by Tra2α and Tra2β (positive dPIR). 16 introns show negative dPIR including introns in *CHEK1*, *AKT1S1* and *BUB3*, genes predicted to be regulated by Tra2 proteins in the MISO analysis too.

Although the data from MAJIQ is very promising, it is important to bear in mind that these are only predictions and a manual inspection of each case is ultimately necessary to fully understand what is behind each event. Upon closer inspection there are a few intron retention cases that may need to be excluded. Interestingly, these are all the three cases also identified by MISO to be alternatively spliced. For instance, MISO classifies the *AKT1S1* alternative event as an alternative 5' splice site that matches with the 5' splice site of the "intron" defined by MAJIQ. The decrease in read coverage in this region is mistakenly taken as a change in intron inclusion by MAJIQ because of the definition of

the coding areas in the reference genome MAJIQ uses, the Ensembl Gene Predictions database (2014). Similarly, the alternative events in *BUB3* and *CHEK1* involve alternative splice sites located in the final exons of the genes (Figure 18). Because of the predicted, but not validated, reading frames in the reference genome these alternative events are listed as intron retentions by MAJIQ. In order to confidently classify these events, we designed primers to test the splice junctions by RT-PCR. Primers were designed to bind the flanking exons and where necessary an internal primer was added. The PCR results show a clear change in the alternative splicing of these events but no evidence of intron inclusion. Although the classification of these events may be incorrect, MAJIQ was still able to flag up the splice changes.

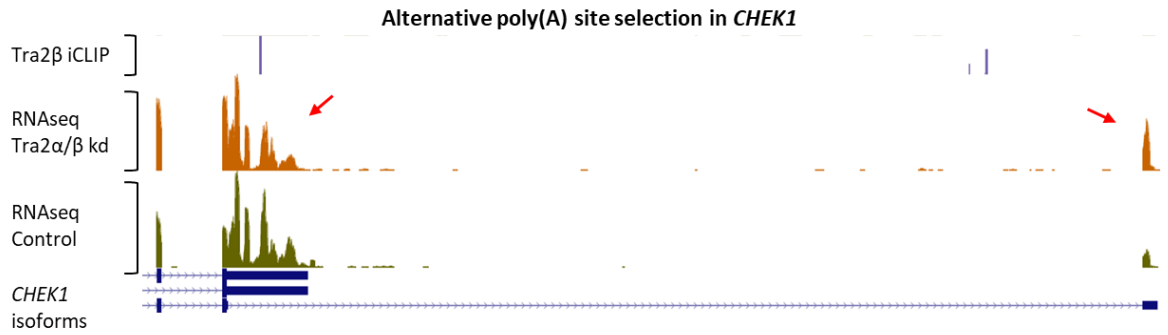

Figure 18 A second alternative event identified in *CHEK1* shows an alternative poly(A) site selected further downstream from the normally expressed isoform. This event was incorrectly classified as an intron retention event by MAJIQ. UCSC genome browser view of RNAseq data, red arrows highlighting the changing RNAseq read coverage peaks accordingly.

In order to confirm that Tra2 proteins are indeed causing changes in intron inclusion we also tested four other intron retention cases; *NT5C3A*, *POLR3C*, *ANTXR2* and *NCK1* (Figure 19). The dPIR changes measured by PCR were considerably lower than those calculated by MAJIQ for the remaining three targets. This could be due to PCR amplification bias or potential differences in the knockdown efficiency. However, the changes measured were in the expected direction indicating that Tra2 proteins can regulate intron retention. Interestingly, the event identified in *POLR3C* resembles an exitron. The affected exon is 256 nucleotides long and the region being spliced out is only predicted to occur by the Ensembl Gene Prediction database. However, the prediction indicates this is a non-coding transcript. Indeed, the exitron is 113 nucleotides long and would cause a frameshift in the mRNA.

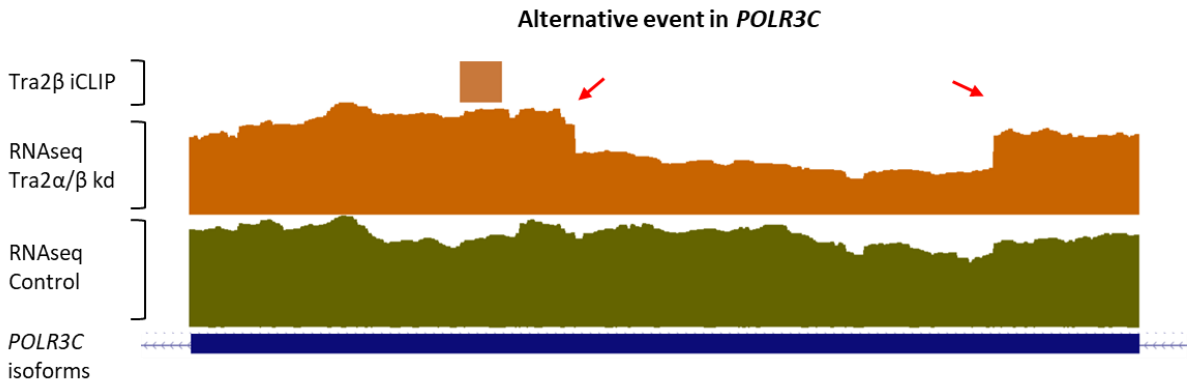

Figure 19 Alternative splicing in *POLR3C* shows an exitron is spliced out in Tra2 double knockdown cells. UCSC view of RNAseq data. Red arrows highlight the boundaries of the exitron.

2.4.3 Alternative events regulated by Tra2 proteins directly impact the open reading frame or crucial protein domains

Alternative splicing can have a range of effects on the final product of a gene. In our approach of analysing the RNAseq data I set out to identify and rank the alternative splicing events based on their impact on the quantity of alternative isoforms expressed. While this was helpful to highlight the target genes whose splicing patterns are strongly dependent on the presence of Tra2 proteins, it does not readily bring any light to the effect on the proteins encoded by these genes. To gain an insight to the potential effect on the proteome caused by the changes in alternative splicing we manually inspected each novel alternative event identified by MISO and MAJIQ to assess the effects anticipated for the proteins expressed by each gene. All 556 alternatively spliced targets were analysed using the Pfam domain prediction tool, which identified a potential change in important protein domains in approximately 37% of the Tra2 targets (Figure 20). Although these predictions are simply based on the anticipated effect on the protein sequence, it is likely that the newly expressed protein isoforms upon Tra2 double depletion may not be the same as their control-counterparts. In addition, we observed that approximately 12% of the total alternative events introduce a premature stop codon in the mRNA sequence. In this context a premature stop codon is not only a stop codon that is earlier in the sequence of the mRNA, but specifically it is located prior to at least one exon-exon junction. This way we are more confident to classify such alternative isoforms as targets of nonsense mediated decay. Hence, overall our results show that in the absence of Tra2 proteins a large proportion (49%) of the alternative isoforms are likely to produce a structurally different protein or no protein at all. A strong indicator supporting this scenario comes from the previously described conclusion that Tra2 proteins are necessary for the splicing of some constitutive exons. In principle, constitutive exons are often very well conserved exons and can contain crucial sequences without which the protein structure may not function properly. It is clear

that the involvement of Tra2 proteins in the regulation of constitutive exons and already established alternative isoforms has a strong impact on the proteome of the cell.

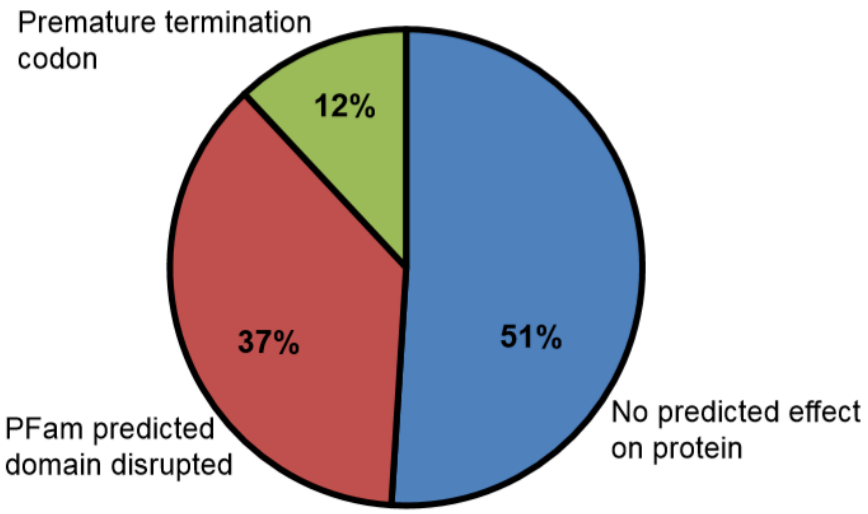

Figure 20 Percentages of alternative events regulated by Tra2 proteins that affect reading frame or disrupt predicted protein domains. All 556 target genes identified by DexSeq, MISO and MAJIQ were analysed and are accounted for in this chart

2.4.4 Tra2β binding is associated with alternatively spliced exons

The Tra2β iCLIP data from the MDA-MB-231 cells has been essential to help create a full picture of the landscape of genes bound by Tra2β. Extensive analysis of the RNAseq data has been useful at identifying the full range of differential splicing events potentially regulated by Tra2 proteins. Although there is a clear overlap of Tra2β bound and differentially spliced exons, there are a significant number of differentially spliced exons with no Tra2β iCLIP tags nearby, and several exons bound by Tra2β that do not show differential splicing. Such occurrences can be due to several possibilities but importantly they challenge our hypothesis that 'Tra2 protein binding is essential for alternative splicing'. To test this hypothesis, we sought to interrogate the RNAseq from MAJIQ and the Tra2β iCLIP data. Matthew Gazzara at the University of Pennsylvania compared the total number of differentially spliced targets that contain Tra2β iCLIP clusters in at least one exon of the regulated LSV to a set of high-confidence, non-changing LSVs. Figure 21 shows a cumulative fraction plot and a significant association of Tra2β iCLIP clusters and differentially spliced LSVs ($p<6x10^{-10}$, two-tailed Fisher's Exact test).

In order to determine if Tra2β binding to the target changing exons is related to the delta PSI of the bound exon we extended the analysis to include all novel cassette exon skipping events found irrespective of the observed dPSI upon knockdown of Tra2 proteins (2068 examples). The results

show that Tra2β-bound changing-exons (Bound A, Figure 21) are differentially spliced at a significantly higher rate ($p<1.5x10^{-20}$, Kolmogornov-Smirnov test). This indicates that Tra2β binding can strongly influence the total change in splicing inclusion of the bound exon.

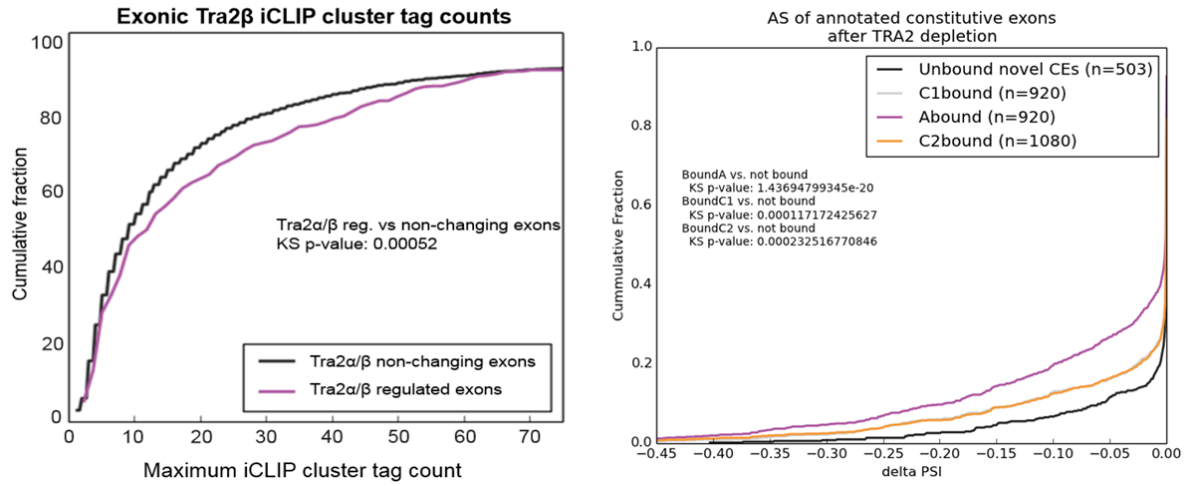

Figure 21 Analysis of Tra2α/β bound exons. *Left*: Cumulative fraction analysis of Tra2β bound exons that show alternative splicing vs non-changing exons. Kolmogornov-Smirnov test, p-value: 0.00052. *Right*: Cumulative fraction analysis of Tra2β iCLIP tags shows stronger dPSI changes associated with target exons (Abound) relative to flanking exons (upstream "C1bound" and downstream "C2bound"). Note, C1bound and C2bound curve overlap.

2.4.5 Tra2 proteins regulate the splice patterns of other splice factors

While our results from the analysis of Tra2-bound changing vs non-changing exons show strong evidence that Tra2-binding can significantly increase the dPSI response, we also observe several alternatively spliced genes that have no Tra2β iCLIP tags in the nearby sequences. Such splice changes could be due to other RNA-binding proteins regulated by Tra2α/β that in turn are also regulators of alternative splicing. Two very good examples of other splicing regulators that could be controlled by Tra2α and Tra2β are the alternatively spliced genes Quaking (*QKI*) and DEAD-box Helicase 5 (*DDX5*), two RNA-binding proteins known to regulate pre-mRNA splicing (Galarneau & Richard, 2005; Kar et al., 2011). The changing exons in both examples are associated with large clusters of Tra2β iCLIP tags indicating direct binding of Tra2 proteins. Interestingly, the cassette exons affected in both genes are constitutive exons, but their lengths are multiples of 3 (3n nucleotides), hence the spliced isoforms should remain in-frame and express shorter proteins. Pfam domain prediction indicates that the changes in protein sequence would disrupt the KH-domain for *QKI* and the Helicase conserved C-terminal domain in the case of *DDX5*. We were able to confirm these targets by PCR but have not tested the changes at the protein level. Despite this, it would be

difficult to determine exactly which Tra2-unbound alternative events are due to the effects of *QKI* or *DDX5* (or other genes) considering that the conditions here involve a multitude of other gene changes (Figure 22). However, the presence of these alternative splicing regulators being clearly affected by the Tra2-double knockdown may provide an explanation for the lack of Tra2β iCLIP tags over some strongly changing alternative events in our data.

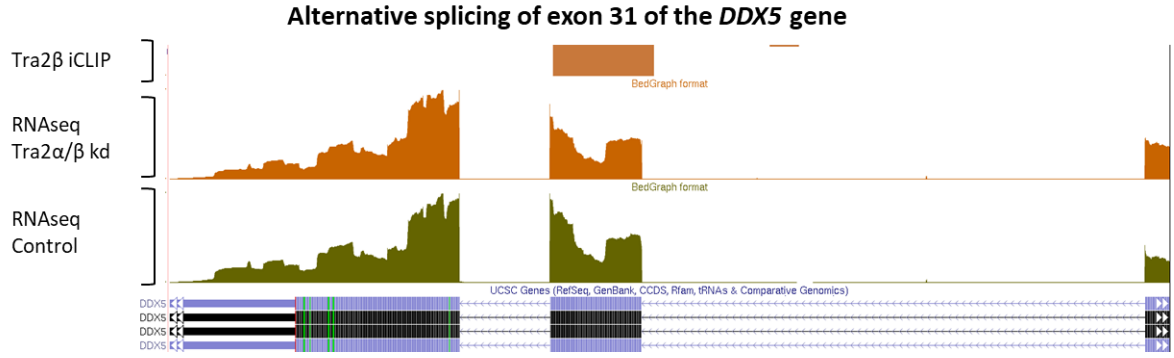

Figure 22 Alternatively spliced exon 31 of *DDX5* is also associated with many Tra2β iCLIP binding tags. Figure shows the RNAseq read coverage over the target and flanking exons as viewed on the UCSC browser.

2.4.6 Gene Ontology enrichment analysis identifies only one enriched term: "Chromatin Organisation"

Gene Ontology (GO) enrichment analysis can be a useful tool to investigate enrichment of biological or cellular pathways in the list of target genes regulated by Tra2 proteins. In the course of my PhD we have performed this analysis several times using several lists of target genes and the results have varied significantly, especially due to the changes in the database of GO terms used. During the early stages of our work the GO enrichment analysis performed by Katherine James (Newcastle University) using the total target lists, including genes from DexSeq, MISO and MAJIQ, we identified 7 enriched GO terms. However, recent changes in the Bioconductor database of Gene Ontology terms the results from the GO enrichment analysis show only 1 term to have statistically significant enrichment in our list of genes (*chromatin organisation*). Much of the analysis that directed our lab work was based on the original results of the GO enrichment analysis.

2.4.6.1 Genes involved in chromosome biology show strong changes in RNA splicing

Our initial Gene Ontology enrichment analysis identified 3 enriched terms strongly implicated in chromosome regulation (*'sister chromatid segregation'*, *'chromosome segregation'* and *'chromatin remodelling'*), and a further two terms that are associated with chromosome modification (*'histone H3-K9 modification'* and *'regulation of gene expression, epigenetic'*) (Figure 23). These results are somewhat similar to those of the GO enrichment analysis performed on the DexSeq gene list (Best et

al. 2014), indicating a clear importance of the Tra2 role in these pathways. The results from Dr. Best's work indicated that the effect of Tra2 proteins on chromosome biology could be the cause of the abnormal nuclear morphology reported in Best et al. 2014. Identifying similar biological pathways enriched in a much larger dataset of Tra2-regulated genes is a strong indicator of the role of Tra2 proteins in chromosome regulation. In order to be confident however we must confirm that indeed the Tra2 knockdown cells are having an effect in the proteins expressed by the target genes involved in 'chromosome biology'. We set out to test the expression of genes from each GO enriched term using western immunoblotting. Target genes were selected based on certain criteria. Firstly, it is important to carefully investigate each alternative event in order to determine if the splice switch is predicted to either produce a shorter/larger protein or none at all. Some alternative protein isoforms made by splice changes such as mutually exclusive exons (that are similar in length) or alternative 5' and 3' splice sites that differ by few amino acids may be impossible to test by western blotting. Secondly, we ranked the genes within the enriched groups by dPSI. We reasoned that lower dPSI events may not make much change to protein isoform levels. Finally, we selected targets based on the extent of available literature on their individual roles and function. Genes that are well known to have an effect in chromosome biology and are alternatively spliced upon Tra2 double knockdown could hold the key behind the effects of the knockdown on the nuclear morphology.

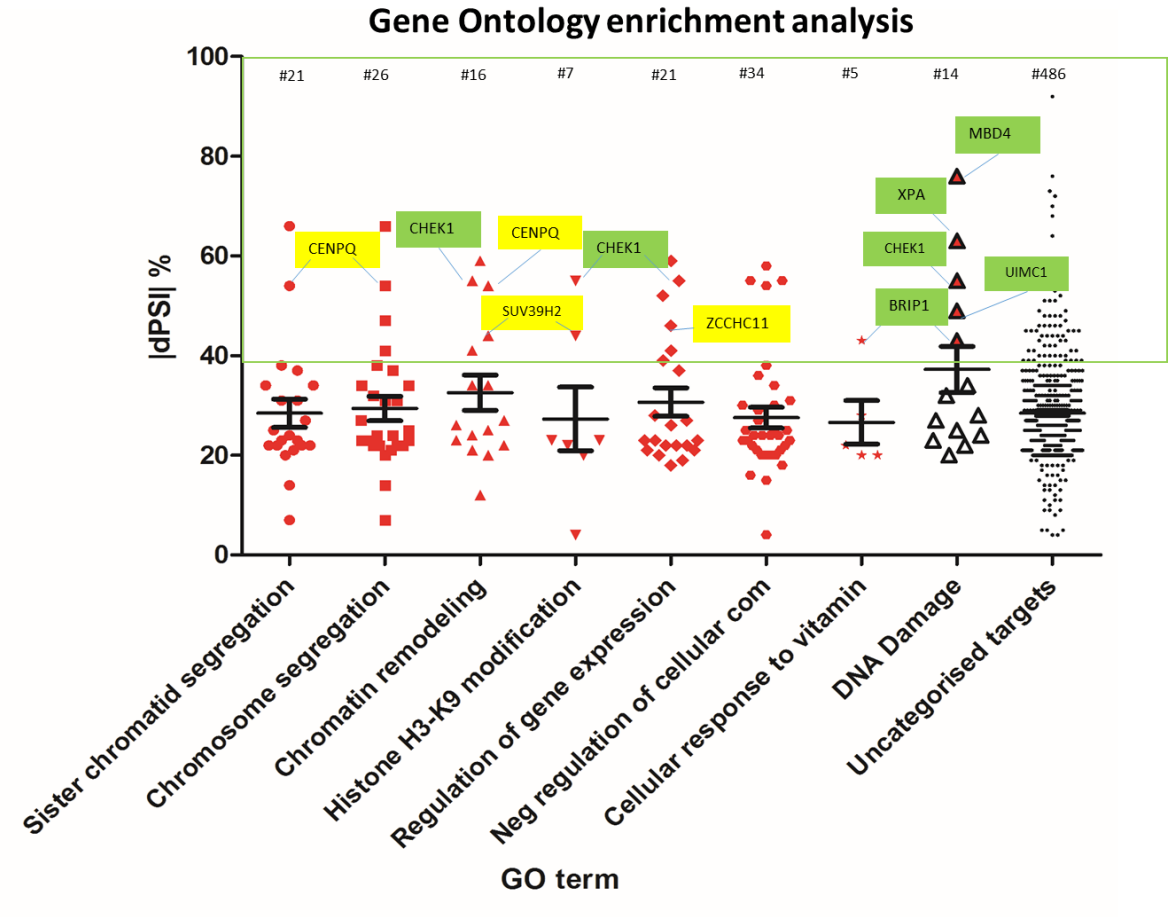

Figure 23 Initial GO enrichment analysis highlighted several GO enriched terms associated with chromosome biology. Note that "DNA Damage" genes were manually highlighted, some of which are investigated further in the following chapter. #Number atop of each column represent the total number of genes in each GO term/group. Genes in highlighted in green boxes are DNA damage associated genes, and in yellow boxes are genes involved in chromosome biology.

2.4.6.2 Tra2-regulated DNA binding proteins involved in chromosome biology do not show changes at the protein level

The substantial number of Tra2α/β target associated with chromosome biology raises the question of the potential role alternative splicing can have in these genes. To investigate a potential role of the alternative changes in chromosome biology we must first determine these changes are affecting the protein expression pattern of the target genes.

My attention was directed at the *CENPQ* gene (Figure 24). This gene encodes for the Centromere Protein Q and is a component of the CENPA-CAD complex that is recruited to the centromeres and is involved in mitotic progression and chromosome segregation. Interaction of *CENPQ* and *PIBP1* was shown to be essential for their localisation to the centromere, and therefore the localisation of the CENPA-CAD (Kang et al. 2011). Our bioinformatics search has also identified alternative events occurring in two other related genes: *CENPK* and *CENPT*. However, the dPSI effect measured in

CENPQ upon Tra2 double knockdown was significantly stronger (-54%) than those measured in *CENPK* (-34%) and *CENPT* (-20%). Moreover, the alternative event in *CENPQ* is anticipated to create a shorter isoform of the protein which should be detectable by western immunoblotting. However, the results were disappointing as no short *CENPQ* isoform was detected and the primary *CENPQ* protein isoform does not significantly change expression even when we tried testing the knockdown after 96 hours of transfection to account for protein turnover (Figure 25).

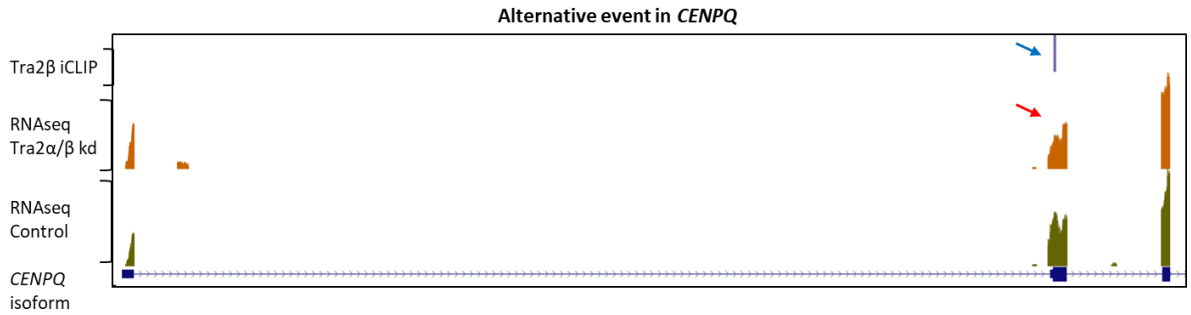

Figure 24 Alternative splicing in *CENPQ*. UCSC genome browser view of the RNAseq tracks shows a reduction in read coverage of the first coding exon (red arrow) associated by a cluster of Tra2β iCLIP tags.

Similar to the CENP-family of proteins regulated by Tra2α/β we also noticed that a group of zinc finger DNA-binding proteins undergo alternative splice changes at high dPSIs. Namely the CCHC-domain containing genes *ZCCHC7*, *ZCCHC9* and *ZCCHC11* show dPSIs ranging from -46% to -72%. The *ZCCHC11* gene, also known as *TUT4*, encodes for an RNA uridyltransferase that uses UTP to add uridines to the 3' end of substrate RNA molecules. Although *ZCCHC11* does not show the highest changing dPSI among the three candidate zinc finger proteins, the alternative event involves the skipping of exon 2 that in turn is predicted to encode a much shorter protein isoform (108kDa). The shorter isoform lacks 5 predicted domains present in the full-length protein (185kDa). We were able to validate the isoform switch consistently at the RNA level. However, the antibody used against *ZCCHC11* was not able to detect the expression of the shorter isoform we predicted or any other potential isoform that could indicate the alternative splice switch is affecting the protein level of *ZCCHC11*.

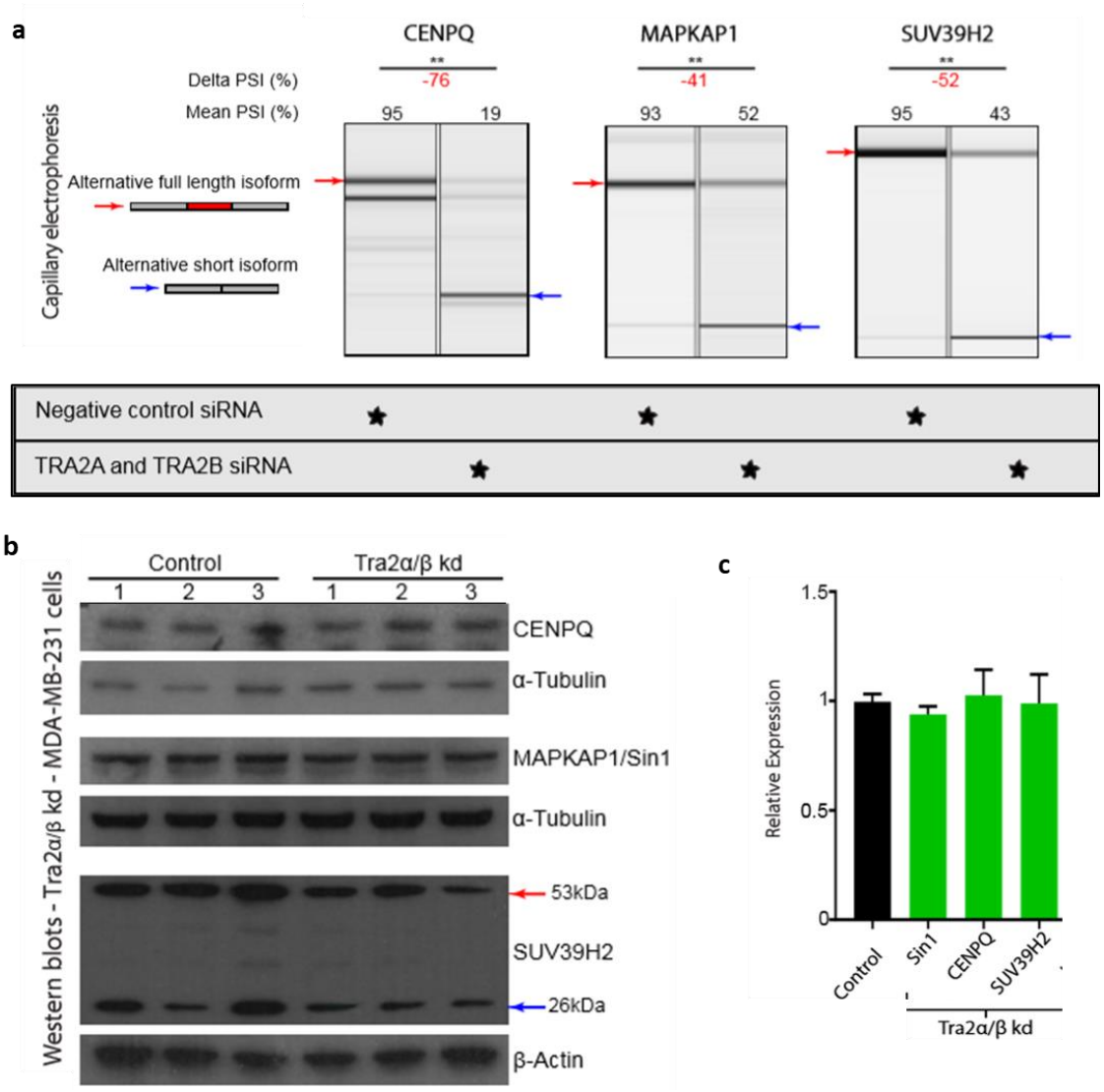

Figure 25 Reverse transcription PCR validation of *CENPQ, MAPKAP1* and *SUV39H2* shows strong and statistically significant dPSI changes in Tra2α/β depleted cells. b) Western Immunoblotting of corresponding proteins does not show significant reductions (c) at the protein level relative to control samples. Quantification of protein expression was done using ImageJ for samples with three biological replicates. Significance measured using Student's *t*-test. ** represents $p<0.01$

The *MAPKAP1* gene encodes a protein called SIN1 (Stress-activated map kinase interacting protein 1) which is a component of the mTOR complex 2. This complex is a key regulator of cell growth and survival in response to endogenous growth factors. There are six known protein coding isoforms for this gene listed on UniProt KB, also defined by the RefSeq genes on the UCSC genome database. The alternative event being affected by the Tra2 double knockdown indicates the use of an alternative cassette exon. The RNAseq changes over this exon are also associated with a cluster of Tra2β iCLIP tags. The estimate dPSI change by the bioinformatics tools is -36%. We were able to confirm this change by RT-PCR of the cDNA from control vs Tra2 double knockdown samples. Alternative splicing of exon 2 of *MAPKAP1* leads to the expression of a well-defined mRNA isoform that uses a different

start codon in the downstream exon 5. The remainder of the protein sequence is in frame. The shorter isoform is predicted to express a 37kDa protein which is missing the SIN1 motif as illustrated by the Pfam protein domain prediction tool. The SIN1 motif has been shown to interact with Ras and act as a Ras-inhibitor (Schroder 2007). The results from our western immunoblotting however were unable to detect the shorter isoform at all (Figure 25). This could be due to the specificity of the antibody to the N-terminus. However, we did not see a change in the protein level of the full length SIN1 isoform even when we repeated the knockdown for 96 hours to allow for protein turnover.

Although we only investigated a relatively small number of proteins from the group of genes involved in chromosome biology, our data from western immunoblotting have consistently failed to identify the predicted short protein isoforms or decreased expression of the control-expressed isoforms. Taken as a whole, analyses so far indicated that it was important to examine isoform switches at the protein level as well as the RNA level, in order to understand why depletion of Tra2α and Tra2β caused cell death.

2.4.6.3 Tra2 deficient cells show a switch in the methylation pattern of histone 3

The GO enrichment analysis pointed our attention towards the regulation of chromosomes. A noticeable number of target genes are directly involved in histone methylation/demethylation such as the *KDM1A*, *KDM3A* and *SUV39H2*. Therefore, we sought to investigate if Tra2 double knockdown has an effect on the methylation pattern of Histone 3 in collaboration with Prof. Tanja Vogel's lab (Freiburg University). We knocked down the expression of Tra2α and Tra2β in MDA-MB-231 cells over 72 and 108 hours as previously described and collected the protein in RIPA buffer. The methylation of Histone 3 at lysine residue K4 and K27 was investigated with respect to the total presence of Histone 3 expression. Interestingly, levels of H3K4me2 were significantly reduced after the 72 hour knockdown but no significant change was measured after 108 hours. H3K4me2 has been previously reported to be positively associated with transcription factor binding regions (Ying Wang, Li, & Hu, 2014). H3K4me3 however remained unchanged across both experiments. The trimethylation of lysine 27 (H3K27me3) was observed to increase in the 72 hour and 108 hour Tra2 siRNA transfected cells, however the change was only statistically significant in the 108 hour treated cells (Figure 26). H3K27 trimethylation plays a role in maintaining gene transcriptional repression and is deposited preferentially at CpG-dense promoters (Ferrari et al., 2014). Tra2 double knockdown cells clearly show alterations in the methylation patterns associated with transcriptional repression. Pasini et al. 2004 showed that SET-domain-containing methyltransferases are directly involved in the methylation of histone H3K9/27 (Pasini, Bracken, Jensen, Denchi, & Helin, 2004). A clear candidate gene from our Tra2-regulated targets containing a SET domain is *SUV39H2*.

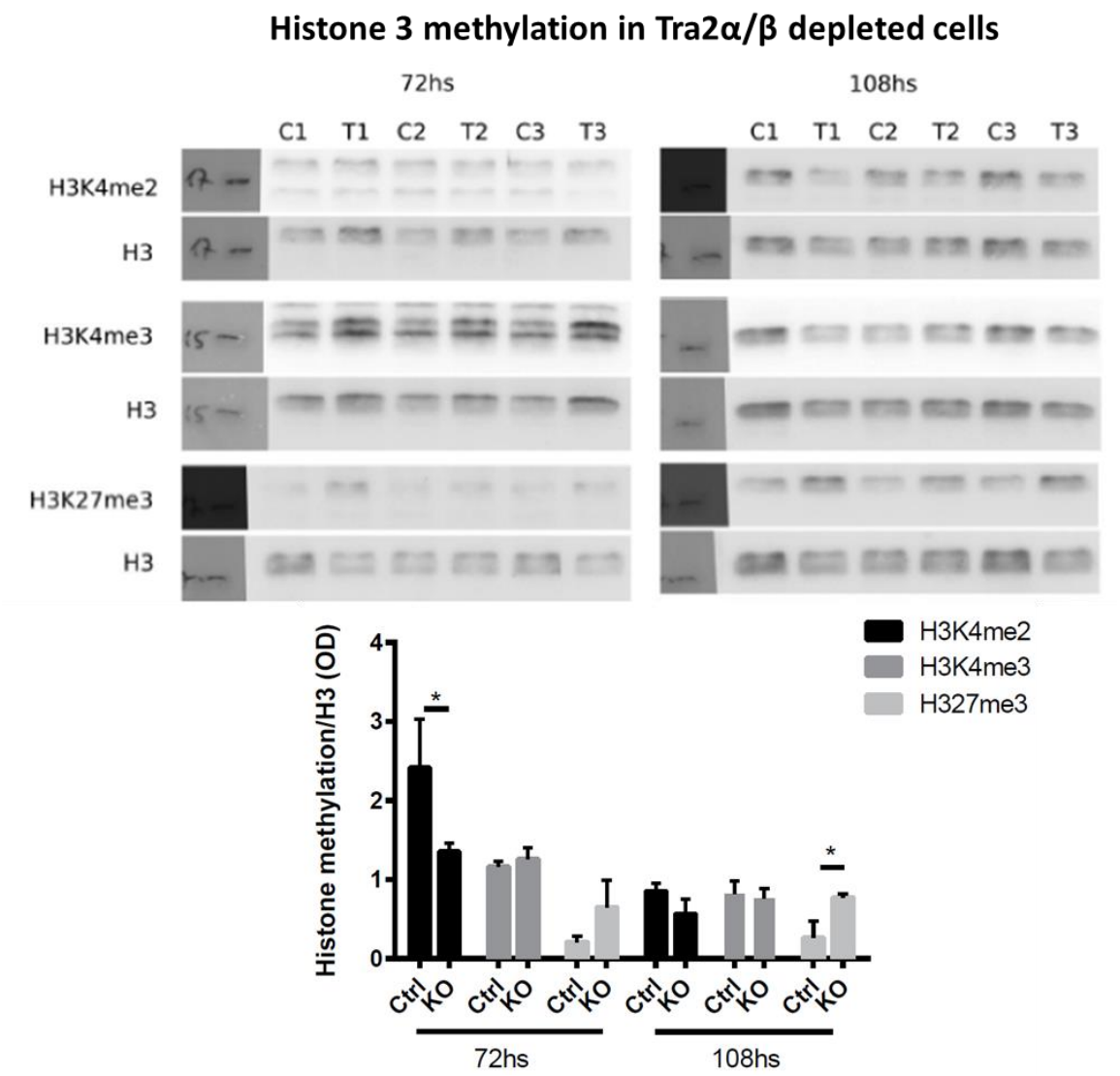

Figure 26 Western immunoblotting of different methylation conformations for Histone 3 show a significant switch from dimethylated of H3K4 to trimethylated H3K27. This work was done by Tanja Vogel's laboratory, Frieburg University. *p<0.05

The protein encoded by *SUV39H2* is a histone methyltransferase that specifically trimethylates Lys-9 of histone 3 near the centromeric region (Dillon 2005 and Schuhmacher 2015). MDA-MB-231 cells depleted of Tra2 proteins express a shorter mRNA isoform of *SUV39H2* that makes use of an alternative 5' splice site within exon 3. A cluster of Tra2β iCLIP tags over exon 3 suggests the involvement of Tra2 proteins in the regulation of this alternative event. The short alternative isoform is well defined by the UCSC genome database and is predicted to encode a 26kDa protein isoform. To evaluate the potential impact of the alternative splice switch we ran the sequences of the two isoforms through the Pfam protein domain prediction tool. The shorter isoform is clearly missing the "pre-SET" motif, a zinc binding protein motif functioning as a 3-dimensional scaffold to accommodate the SET domain which is present in both isoforms. Western immunoblotting shows

that although the shorter protein isoform of *SUV39H2* was detected in both control and knockdown cells, there is no statistically significant change in either isoform (Figure 26). Although a change in *SUV39H2* protein isoform levels could not be detected, the methylation pattern changes measured in Tra2 knockdown cells may well be due to the isoform switch of other methyltransferases.

2.5 Discussion

2.5.1 Multiple bioinformatics prediction tools are required for a comprehensive analysis of RNAseq data

To further understand the role of Tra2 proteins in breast cancer cells, a key priority was to identify the full scope of genes whose alternative splicing was directly regulated by Tra2 proteins. To do so we used MISO and MAJIQ, two improved bioinformatics pipelines compared to DexSeq, to analyse the original RNAseq data obtained from the Tra2α/β double knockdown in MDA-MB-231 triple negative breast cancer cells. The increased capacity to detect alternative events by MISO and MAJIQ expanded the pool of alternatively spliced genes tenfold (from 54 to 556). Moreover, both MISO and MAJIQ were restricted to search for splice changes with a minimum of 20% dPSI, something that was not implemented in the DexSeq analysis. The large number of identified splice changes by MISO and MAJIQ suggest that the DexSeq pipeline has major limitations at identifying the full set of genes affected by the knockdown experiment. Although the newly expanded list of potentially Tra2-regulated genes is substantially larger, a noticeable drawback is the small number of overlapping exons identified by all three pipelines. Despite this, the experimental validation success rate of the newly identified targets indicates that the predicted splice changes are indeed occurring upon Tra2 double knockdown. A crucial conclusion to be drawn here is the fact that each bioinformatics package has its weaknesses and therefore must be used in conjunction with other packages in order to robustly identify the full range of alternative events. This was also highlighted by work published recently with collaborators in the recent publication of the newly developed SUPPA2 bioinformatics package used to analyse alternative splicing results (J. L. Trincado et al., 2018). Although SUPPA2 was able to achieve higher accuracy compared to other methods, especially at low sequencing depth and short read length, the investigation of Tra2-regulated targets identified several novel hits, suggesting MAJIQ, MISO and DexSeq had still not identified all alternative events from our RNAseq data.

The presence of iCLIP tags on alternatively spliced targets indicates direct binding and regulation by Tra2 protein. The alternative splice changes identified by MISO and MAJIQ do not directly depend on the presence of Tra2β iCLIP tags. The analysis of the MAJIQ data (by Matthew Gazzara at the University of Pennsylvania) indeed shows that Tra2β binding is significantly associated with a change in alternative splicing. However, there were a number of alternative events identified by both MISO and MAJIQ that do not contain any Tra2β iCLIP tags associated with them. A potential explanation for this could be the effect of Tra2 proteins on the regulation of splicing of other splice factors such as *QKI* and *DDX5.* Another factor contributing to the indirect regulation of alternative events could

be due to stress induced changes in the cells by the absence of Tra2 proteins. Moreover, absence of Tra2β iCLIP binding tags is not direct evidence that Tra2α does not bind to these exons. Although previous work clearly indicates that Tra2 proteins share similarities in structure and target binding, it is possible that some of these targets are preferentially bound and regulated by Tra2α. Therefore, it would be unwise to exclude from the analysis any target genes showing strong splice changes based purely on the absence of Tra2β iCLIP tags.

2.5.2 Tra2 proteins regulate a range of mechanistically different alternative events

A strong advantage of using MISO and MAJIQ in the analysis of Tra2-regulated alternative events is their capability to detect complex alternative splice changes (as opposed to just cassette exons by DexSeq). Amongst the 502 novel alternative events, 334 alternative events were classified as complex. These include alternative 3' and 5' splice sites, intron retention events and mutually exclusive exons. However, overall analysis shows that 70% of Tra2-regulated events are alternative cassette exons, and a further 19% are constitutive cassette exons. This finding supports the concept that Tra2 proteins act by binding to and promoting the inclusion of exons via direct interactions through the exonic splicing enhancer sequences (Desmet et al., 2009; Tacke et al., 1998). However, the bioinformatics analysis also identified 35 repressed cassette exons. The analysis of iCLIP tag distribution shows a significant correlation between the flanking exons bound by Tra2β and repressed target exons, suggesting a potential mechanism of action for Tra2 proteins similar to that previously reported in the regulation of exon 11 of Ron by SRSF1 (Ghigna et al., 2005). Moreover, previous reports in the fruitfly *Drosophila melanogaster* have shown the tra2 protein can directly repress the splicing of its own M1 intron, which ultimately limits its expression (Chandler, Qi, & Mattox, 2003; Qi, Su, & Mattox, 2007). However, to the best of my knowledge there is no evidence of human Tra2 proteins acting to repress the splicing of exons. An interesting example showcasing exon repression by Tra2 proteins is clear in the *DROSHA* gene (Figure 27). Exon 6 of *DROSHA* has no associated iCLIP tags, but both flanking exons do. Furthermore, the largest cluster of Tra2β iCLIP tags is present in the downstream exon 7 of *DROSHA*, further backing up the notion of Tra2 proteins repressing the splicing of upstream exons.

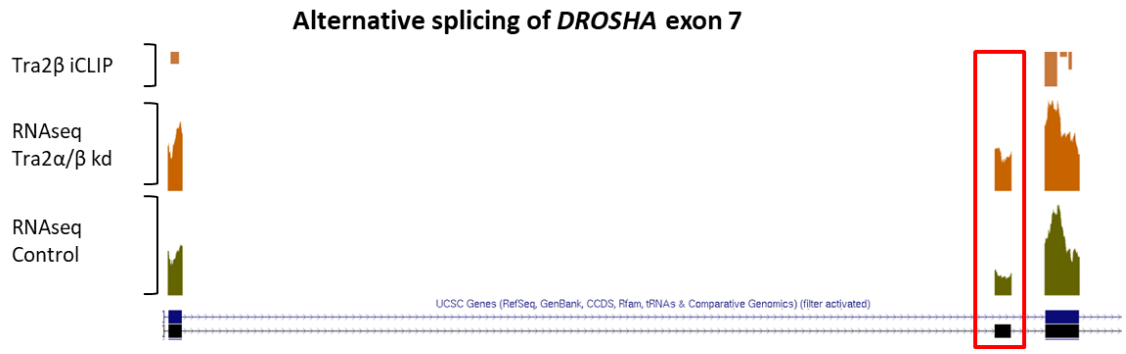

Figure 27 Alternative splicing of *DROSHA* exon 7. Tra2β iCLIP tags cluster on the downstream exon 8 but Tra2α/β knockdown shows exon 7 skipping indicating a potential mechanism where Tra2 proteins repress splicing of an upstream exon

2.5.3 Tra2 proteins indirectly regulate expression of the *TTC14* gene

Previous reports have clearly demonstrated the capability of Tra2β regulating the expression of itself and that of Tra2α via the alternative splicing of poison cassette exons (Andrew Best et al., 2014; Stoilov, Daoud, Nayler, & Stamm, 2004). Our data highlights an interesting example where Tra2 proteins can indirectly influence the expression of a gene via splicing of a key isoform. The alternative cassette exon 11 of *TTC14* detected from DexSeq as alternatively spliced in cells treated with siRNA against *TRA2A* and *TRA2B* shows a clear relative drop in inclusion (Figure 28). The calculated PSI reduction is over 90% and this result was also confirmed by PCR in Dr. Best's thesis. The Tra2β iCLIP data in MDA-MB-231 cells indicates a bundle of iCLIP tags (more than 100) on the alternative cassette exon 11 of *TTC14* showing strong evidence of binding by Tra2β. Interestingly, the expression of the *TTC14* gene was markedly increased in cells treated with *TRA2A/B* siRNAs. Dr. Best confirmed the increased gene expression by quantitative PCR during his PhD together with a range of other genes that also showed differential gene expression. Since there are no known associations of Tra2 proteins with regulation of transcription, the changes in expression measured upon Tra2 proteins knockdown were generally accredited to downstream effects such as cell growth or the cell cycle. However, in the case of *TTC14*, the increase in expression may be directly due to aberrant splicing in the absence of Tra2 proteins. On closer inspection of the RNAseq data the relative abundance of the isoform including exon 11 of *TTC14* is clearly lower in the knockdown cells. However, the total expression of the isoform containing exon 11, as indicated by the RNAseq read peaks, remains the same. While this effect may be an attempt from the cell to maintain the expression of this specific isoform of *TTC14*, it is clearly the absence of Tra2 proteins and the subsequent missplicing of *TTC14* that causes such an increase in the transcription of the gene. There is very little known about the function of *TTC14*. To date there is only one study that eludes to the importance of *TTC14* in lung cancer patients. Wang and colleagues (2016) investigated the presence

of plasma autoantibodies (AAb) in patients with lung cancer vs smokers/patients with benign pulmonary nodules (J. Wang et al., 2016). Their results showed that two panels of 5 autoantibodies each could be used to differentiate between patients with lung adenocarcinoma vs smoker control participants and patients with lung adenocarcinoma vs patients with benign pulmonary nodules. Interestingly, *TTC14* and the well-known proto-oncogene BRAF were in both five-AAb panels. The paper also highlights the results from The Cancer Genome Atlas database on lung adenocarcinomas (LUAD), showing a significant increase in the protein expression of *TTC14* in LUAD patients. However, there is no indication as to which isoform of *TTC14* the study refers to. Future investigations into the specific function of each isoform could highlight the importance of Tra2 proteins in the regulation of splicing for *TTC14.*

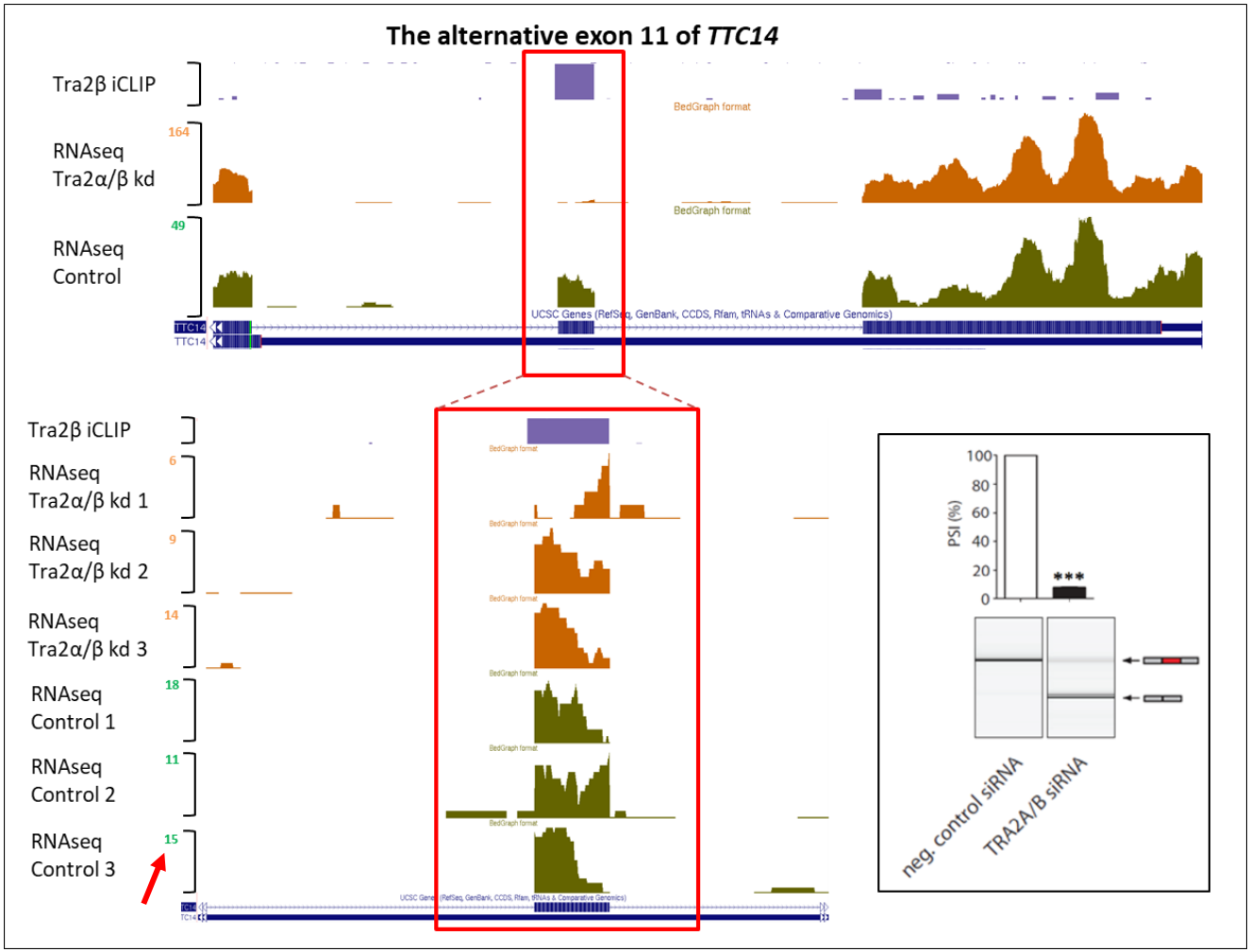

Figure 28 Exon 11 of *TTC14* is alternatively spliced upon Tra2 double knockdown. However, total expression of the long isoform is not substantially different as shown by max peak reads (red arrow) across all biological replicates in the UCSC browser.

2.5.4 Histone methylation pattern changes are too minor to explain the severe structural changes in the nucleus of Tra2-knockdown cells

The Gene Ontology enrichment analysis highlighted a panel of GO terms associated with chromosome biology. Furthermore, previous investigations into the effect of Tra2 double knockdown in MDA-MB-231 cells show abnormal nuclear morphology (Andrew Best et al., 2014), suggesting absence of Tra2 proteins leads to aberrant splicing of key proteins involved in nuclear conformation and chromatin packaging. However, our results indicate that the RNA splice changes incurred by the Tra2 double knockdown are not reflected at the protein level for some of the tested genes (*CENPQ, MAPKAP1, SUV39H2* and *ZCCHC11*). While we cannot measure a direct influence on the protein isoforms of the genes in enriched GO terms it does not mean that these changes are not representative of Tra2-regulated splice patterns in the cells. Tra2 proteins are ubiquitously expressed but the level of expression, particularly of Tra2β, can differ depending on cell type and developmental stage (A. Best et al., 2013). Therefore, the detrimental effects that Tra2 double knockdown has on cell survival may not be a good representation of the long-term effects of minor splice changes induced by Tra2 overexpression in cancer cells.

Furthermore, GO enrichment analysis relies on a well annotated and updated database. On several occasions during the course of this PhD we have analysed the same gene set and obtained different GO enrichment results, suggesting that GO enrichment analysis should not be used alone as a guiding tool to highlight important cellular pathways. As a concrete example, the GO term “cellular response to vitamins” was identified as a significantly enriched term although only five genes were highlighted in this pathway. This is of course due to the small number of genes annotated in the “cellular response to vitamins” GO term database (only seven). Nonetheless, this GO term does not represent a large enough group of genes (out of the 556) to confidently conclude that one of the main roles of Tra2 proteins is to regulate cellular response to vitamins.

The analysis of Histone H3 methylation patterns in Tra2 double knockdown cells indicates changes in alternative splicing are having some effect on histone methylation. However, it is difficult to explain such changes in methylation if none of the tested genes involved in chromosome biology show altered protein expression. One possible explanation could be the interaction between chromatin readers and modifiers is affected by the alternative splice changes induced by the Tra2 knockdown (Torres & Fujimori, 2015). This suggests that the methyltransferases, such as *SUV39H2*, could have their function impaired if the protein domains required for protein-protein interaction is spliced out, regardless of the expression of the gene. Future experiments interrogating the proteome of the cells

after Tra2 double knockdown would potentially bring to light affected chromatin remodelling proteins that could explain the cellular phenotypes observed.

2.6 Chapter summary

In this chapter, I have utilised a range of recently developed bioinformatics tools to identify a large number of alternatively spliced targets in MDA-MB-231 cells regulated by the homologous splicing regulators Tra2α and Tra2β. Tra2 proteins were found to regulate predominantly alternative cassette exons, but intriguingly also repress cassette exons and regulate intron retention. I validated a substantial number of alternative events, some of which did not contain Tra2β iCLIP binding tags, suggesting some of these genes could be regulated indirectly by Tra2 proteins. Gene Ontology enrichment analysis directed us to test the methylation patterns of Histone H3. Western immunoblotting of proteins encoded by candidate genes that were alternatively spliced did not show any significant changes, suggesting that other cellular pathways and target genes may be responsible for the cell viability and nuclear conformation changes measured upon Tra2 double knockdown.

Chapter 3: Induction of specific splice isoforms in breast cancer cells

3.1 Introduction

In chapter 2, MISO and MAJIQ helped identify many alternatively spliced target genes regulated by Tra2 proteins in MDA-MB-231 cells. The expanded list of target genes reveals a more realistic picture of the true scale of influence of Tra2 proteins in alternative splicing. Gene Ontology enrichment analysis is often a useful tool to highlight biological and cellular pathways from a large pool of genes. However, the data from chapter 2 suggests that enriched pathways of genes alternatively spliced upon Tra2 double knockdown are not necessarily affected at the protein level. This means that we cannot confidently suggest that these genes are affecting the cells' function and viability in the 72 hours of Tra2 double knockdown. While the role of these alternative splice changes in the long term may be important, in the short-term other genes could be playing a more essential role in cell survival. In order to identify such essential target genes a different approach would have to be employed. Firstly, GO enrichment analysis of differential gene expression previously reported in Dr Best's thesis can be used to indicate the effect of the Tra2 double depletion in gene expression and in turn highlight pathways activated as a consequence of alternative splicing. Secondly, investigating splice targets with the strongest dPSI changes that have a clear and detrimental effect on the predicted protein could be a better way of identifying essential target genes. Finally, Dr Best's previous investigation into DNA damage suggests Tra2 double knockdown has a clear and marked effect on DNA damage accumulation. The key role of *CHEK1* in DNA damage response has been reported previously (Chen & Sanchez, 2004; Patil, Pabla, & Dong, 2013) and the drastic decrease of CHK1 protein levels upon Tra2 double knockdown suggest that DNA damage is accumulating. Therefore, it is possible that DNA repair genes other than *CHEK1* are also affected by the Tra2 depletion, which could explain why *CHEK1* overexpression alone did not rescue the cells in the previous experiments (Andrew Best et al., 2014).

In this chapter we investigate the impact of individual splice changes in MDA-MB-231 cells caused by Tra2α and Tra2β protein depletion. The effects of alternative events that introduce a premature stop codon in the reading frame, such as in the case of *CHEK1*, can be easily tested by knocking down the expression of the gene. However, alternative events that affect protein sequence and therefore induce the expression of a different protein isoform may be difficult to investigate by siRNA experiments. In order to recapitulate the alternative splice changes that occur upon Tra2 double depletion for individual genes we employ the use of modified antisense oligonucleotides. Various backbone modifications to oligonucleotides such as phosphorodiamidate morpholino oligomer

(PMO) and 2'O methyl phosphorothioate (2'OMePS) have been previously used to induce exon skipping (Shimizu-Motohashi, Murakami, Kimura, Komaki, & Watanabe, 2018; van Putten, Tanganyika-de Winter, Bosgra, & Aartsma-Rus, 2019). Morpholino oligomers introduce a neutrally charged backbone and are often used in clinical trials due to their reduced reactivity and adverse side effects. However, negatively charged oligonucleotides such as phosphorothioate modified oligos are more readily taken up by the cells making them ideal for transfection *in vitro* (Pitout, Flynn, Wilton, & Fletcher, 2019). The last part of this chapter will focus on the potential of 2'OMePS antisense oligonucleotides (AO) as a method of inducing gene specific alternative splice changes (Figure 29). Experimental evidence shifted the focus of the investigation towards genes involved in DNA repair, a largely regulated cellular pathway by Tra2 proteins that was not immediately highlighted by the Gene Ontology enrichment analysis. Phosphorylation of Histone H2AX (known as γH2AX) is a primary cellular response to DNA double strand breaks (DSB) which helps direct DNA repair mechanisms at damage foci (Pilch et al., 2003). γH2AX is a well-established marker of DNA damage accumulation and as such was used as the primary test to interrogate the function of individual isoforms.

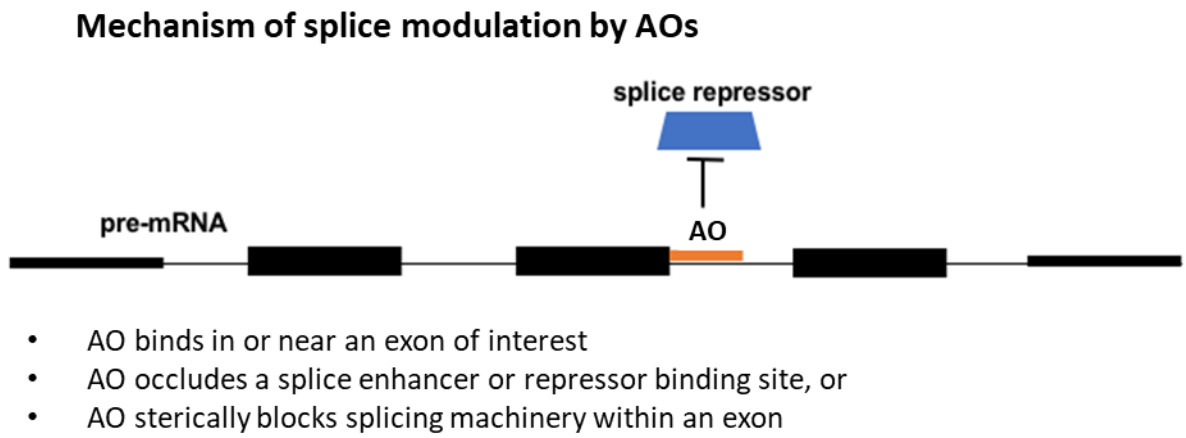

Figure 29 Mechanism of splice modulation by Antisense Oligonucleotides

3.2 Chapter aims and objectives

The main objective of this chapter is to investigate the function of alternative isoforms regulated by Tra2 proteins in breast cancer cells. The specific aims are:

To identify a cellular pathway regulated by Tra2 proteins that is indicative of their role in breast cancer cells

To investigate the effect of alternative splicing on the structure and function of DNA repair genes regulated by Tra2 proteins

To induce the expression of individual isoforms regulated by Tra2 proteins

To interrogate the specific role of such isoforms in cell viability and DNA damage accumulation

3.3 Materials and methods

3.3.1 Cell culture

MDA-MB-231 cells were maintained in DMEM (no phenol red) plus 10% fetal bovine serum and 1% Penicillin Streptomycin. Human embryonic kidney fibroblast (HEK293T) cells were maintained in DMEM with phenol red plus 10% fetal bovine serum and 1% Penicillin Streptomycin. Human bone osteosarcoma (Saos2) cells were maintained in RPMI media plus 10% fetal bovine serum and 1% Penicillin Streptomycin. All cell lines were incubated at 37°C with 5% CO_2 and stored in 75cm^2 and 25cm^2 (Saos2) tissue flasks. These cell lines were provided to me by Dr Andrew Best but were originally purchased from the American Type Culture Collection and LGC Standards, Europe.

3.3.2 siRNA and 2'O Methyl Phosphorothioate transfections

siRNA transfections were performed as described in chapter 2. Antisense oligonucleotides were designed to target the appropriate 3' splice site sequence of the target exons of *BRIP1, CHEK1* and *XPA* (listed in the table below). Scramble sequence derived from IDT own negative control siRNA sequence was used to design negative control antisense oligonucleotide with modified nucleotides. MDA-MB-231 cells were seeded on 6-well plates for 24 hours prior to transfection at a confluence of 1x10^6 in 1.5mL DMEM. The transfection mixture was made similarly to that described for siRNA: 300µl Opti-MEM, 9µl RNAiMAX (ThermoFisher) and 4µl of 100µM *BRIP1* AO or 10µl of 100µl *XPA* or *CHEK1* AO to reach final concentrations of 200nM and 500nM, respectively. The combined mixture was incubated for 5 minutes at room temperature before transfection.

Target gene	2'OMePS sequence	Exon
BRIP1	mU*mC*mA*mG*mG*mG*mG*mA*mG*mU*mC*mU*mU*mA*mU*mA*mU*mA*mA*mG	5
CHEK1	mC*mA*mA*mG*mU*mU*mG*mA*mA*mC*mU*mC*mU*mA*mA*mA*mA*mA*mG*mG	3
XPA	mU*mC*mU*mU*mC*mA*mC*mA*mA*mU*mC*mU*mA*mC*mA*mA*mC*mA*mC*mA	5
Neg control	mC*mG*mU*mU*mA*mA*mU*mC*mG*mC*mG*mU*mA*mU*mA*mA*mU*mA*mC*mG	N/A

3.3.3 Western immunoblotting

Standard Western blotting was performed as described previously in chapter 2. Antibodies used in this chapter (including fluorescent antibodies used for immunohistochemistry) are as follows:

Antibody target	Cat. number	Dilution
BRIP1	#4578 (Cell signalling)	1/1000
MBD4	Abcam, ab12187	1/2000
UIMC1 (RAP80)	Bethyl Labs, A300-763A	1/2000
XPA	Proteintech, 16462-1-AP	1/2000
RBMX	Sigma-Aldrich, HPA057707	1/100
GFP	Clontech, 632381	1/4000
FLAG	Sigma-Aldrich, F3040	1/2000
β-Actin	Sigma-Aldrich, A5441	1/2000
α-Tubulin	Sigma-Aldrich, T5168	1/2000
GFP	Abcam, ab1218	1/2000
H2AX	Santa Cruz Biotech, sc54-606	1/500
γH2AX	Santa Cruz Biotech, sc517348	1/1000
2° Fluorescent Antibody		
Anti-mouse Alexa594	Abcam, ab150120	1/200
Anti-rabbit Alexa488	Abcam, ab150077	1/200

3.3.4 Nuclear fractionation

MDA-MB-231 cells harvested after 72 hours of transfection with anti-Tra2 siRNAs were fractionated to separate the nucleus from the cytoplasmic components. The cells were trypsinised and transferred into Eppendorf tubes. Following PBS washing, cells were re-suspended in 1ml cytoskeletal (CSK) buffer (composed of 100 mM NaCl, 300 mM sucrose, 3 mM MgCl2 and 10 mM PIPES (pH 6.8)). The buffer was removed after centrifugation at 3000 RCF for 2minutes. Cells were then re-suspended in 50µl of CSK buffer and 2µl of 5% Triton X-100 and incubated in ice for 3 minutes. Samples were then spun at 3000 RCF at 4°C for 2 minutes. The resulting supernatant containing the cytoplasmic proteins is transferred into a fresh Eppendorf containing 50µl of 2x Protein Loading buffer. The pellet that remains is washed with 50µl of CSK buffer three times before being suspended in 50µl of 2x Protein Loading buffer.

3.3.5 Immunocytochemistry

Cells were fixed onto coverslips in 24-well plates overnight before treatment with siRNA for 72 hours as described above. At the end of treatment cells were fixed with 4% Paraformaldehyde (PFA dissolved in Phosphate Buffered Saline) for 15 minutes at room temperature. The fixed cells were then permeabilised with 0.1% Triton X-100 in PBS and incubated at room temperature for 15 minutes. Coverslips were then blocked with 10% horse serum in PBS for 15 minutes at room temperature. Primary antibody was diluted in blocking solution at 1:200 ratio and was then incubated with the cover-slips for 1hour. Coverslips were washed three times with PBS and then were incubated for 1 hour in fluorescently labelled anti-rabbit secondary antibody diluted at 1:400 in

PBS while minimising light exposure. Coverslips were again washed three times with PBS and then were mounted onto slides using Vectashield Mounting Medium with DAPI (Vector Labs, H-1200) for nuclear detection. Images were collected using a ZEISS Axioplan2 fluorescent microscope with Axiovision software.

3.3.6 *MBD4* Plasmid Cloning for MBD4

3.3.6.1 Cloning

Transformation of ligated plasmids was done in α-Select Chemically Competent Cells (Bioline) using the manufacturer's heat shock protocol. Ampicillin-resistant transformed cells were plated onto LB agar plates containing ampicillin and incubated at 37°C for 16 hours. Colonies were screened by PCR using primers respective to the vectors. All PCR reactions during the cloning steps are performed using the Phusion High-Fidelity DNA polymerase kit (Thermo Scientific). Following screening, plasmids were sent for sequencing at Source BioScience. Purified plasmid DNA samples were diluted to 100ng/µl and analysed using the splice junction primers for the *MBD4* alternative event or the pXJRTF forward primer and pXJB1 reverse primers (10µM). Resulting sequences were manually mapped onto the genome (UCSC genome browser) and visually inspected.

Phusion PCR reaction

5X HF Reaction Buffer	2µl
Template (insert, vector, 20ng/µl	0.5µl
Forward Primer (10µM)	1µl
Reverse Primer (10µM)	1µl
dNTPs (10µM)	0.2µl
dH_2O	5.2µl
Phusion DNA Polymerase	0.1µl
TOTAL	10µl

Thermocycler programme

1 Heat Activation	95°C	5 min
2 Denaturation	95°C	20 sec
3 Annealing	58°C	30 sec
4 Extension	72°C	60 sec (Cycle to step 2 x29)
5 Final Extension	72°C	10 min
6 Cooling	4°C	∞

3.3.6.2 GFP- and Flag-tagged MBD4

Short (Sh) and full length (FL) MBD4 inserts were prepared using cDNA extracted from knockdown cells. PCR amplified products were separated by gel electrophoresis and subsequently extracted using a QIAquick Gel Extraction kit (Qiagen). MBD4-GFP inserts were prepared using primers with

added *EcoR*I and *Xho*I restriction sites, whereas MBD4-FLAG inserts contained added *Not*I and *Xba*I flanking restriction sites. Utilising these restriction sites, the inserts were ligated with pGFP3 and p3xFLAG vectors (N-Terminal) using T4 DNA ligase (New England Biolabs) following manufacturer's procedure.

3.3.6.3 MBD4 Minigene

Minigene inserts were designed to cover the genomic DNA region that encompasses exons 2 and 3 of MBD4, parts of the flanking introns and intron 2 (chr3:129155088-129156987; hg19, UCSC genome browser). The inserts were synthesised by IDT UK as two g-block fragments (IDT UK) overlapping exons 2 and 3 of *MBD4*, and flanking intronic sequences that may potentially affect Tra2β binding (Figure 30). These fragments are then joined with the common downstream fragment to form the minigene inserts. The *wt* minigene insert was digested with *EcoR*I restriction enzyme. The empty pXJ41 vector was digested with *Mfe*I restriction enzyme and the linearised vector was treated with Antarctic Phosphatase (New England Biolabs) following manufacturer's procedure. The *wt* minigene insert and pXJ41 vector were ligated with T4 DNA ligase (NEB).

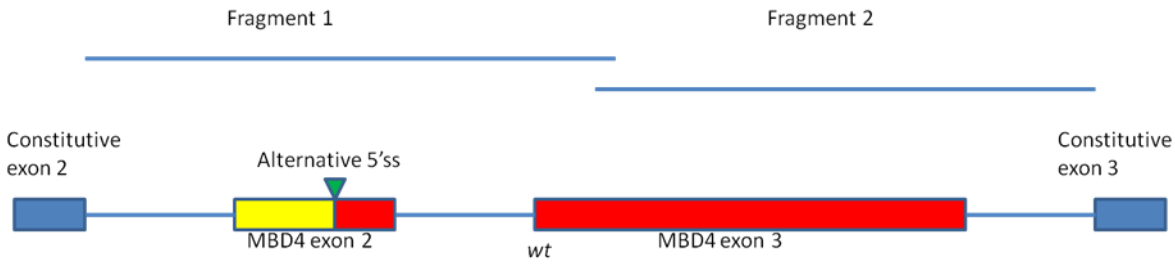

Figure 30 G-block fragments designed to reconstitute the alternatively spliced site in exon 2 of *MBD4* in a pXJ41 plasmid.

3.3.6.4 Plasmid transfection – for immunocytochemistry

HEK293T and Saos2 cells were used for MBD4-GFP and MBD4-FLAG transfection (Dr Best performed transfection of Saos2 cells with MBD4-FLAG). HEK293T cells were seeded onto 24-well plates with cover-slips and incubated for 24 hours at 37°C. Transfection mixes containing 500ng of plasmid, 100µl of Opti-MEM and 3µl of GeneJammer (incubated at room temperature for 30-40 min) were prepared for each of the following plasmids: FL MBD4-GFP, Sh MBD4-GFP, empty pGFP3, FL MBD4-FLAG, Sh MBD4-FLAG and empty p3xFLAG. Each mix was split into 4 wells. Cells were incubated for a further 24-hour following transfection after which they were fixed using 4% PFA.

3.3.6.5 Plasmid transfection – for fractionation

The experimental procedure is like plasmid transfection for immunocytochemistry with the exception that the cells are plated into 6-well plates (4x volume of cells) with no cover-slips. Two

biological replicates were prepared for each plasmid. Cells harvested were fractionated as described above.

3.3.6.6 Minigene transfection

HEK-293 cells were seeded into 6-well plates (no cover-slips) for 24 hours at 37°C. Three biological replicates in each case were prepared with 200ng of the MBD4 minigene and 500ng of Tra2β-GFP plasmid (or pGFP3 plasmid as control). The plasmids were mixed with 100µl of Opti-MEM and 3µl of GeneJammer and incubated at room temperature for 30-40 min. Following transfection cells were incubated for 24 hours and harvested for RNA extraction.

3.3.6.7 One-step RT-PCR

HEK-293 cells transfected with minigene plasmids were harvested and the RNA was extracted using standard Trizol RNA extraction (Life Technologies) following manufacturer's instructions. Reverse transcription and PCR were done in a single reaction, using One-step RT-PCR kit (Qiagen). The reaction mix contained pXJRTF and pXJB1, which are primers targeting the two flanking primer sites in the pXJ41 vector, and an internal screening primer mapping on exon 3 of MBD4. The standard 5µl reaction is outlined below:

One-step RT-PCR

5x reaction buffer	1µl
Q-solution	1µl
pXJRTF primer (10µM)	0.3µl
pXJB1 primer (10µM)	0.3µl
Internal primer (10µM)	0.3µl
dNTPs (10µM)	0.2µl
Enzyme mix	0.2µl
RNA (50ng/µl)	1.7µl
TOTAL	5µl

Thermocycler programme

1 Reverse Transcription	50°C	30 min
2 Initial PCR step	95°C	15 min
3 Denaturation	94°C	30 sec
4 Annealing	58°C	30 sec
5 Extension	72°C	60 sec (Cycle to step 2 x29)
6 Final Extension	72°C	10 min
7 Cooling	4°C	∞

2.3.7 Microsatellite instability assay

Microsatellite instability assay using the Promega MSI analysis kit was performed by Dr Ghanim Alhilal, Newcastle University. For this experiment cells were transfected with negative control and

Tra2α/β siRNAs for 72 hours as per protocol. Genomic DNA was extracted using standard Phenol-chloroform extraction protocol and diluted to 1 ng/ul for the MSI test. Samples were run on a 3130 Genetic Analyzer.

3.3.4 WST-1 cell survivability assay

To quantify the cytotoxicity of 2'OMePS antisense oligonucleotides targeting *BRIP1*, *CHEK1* and *XPA*, a WST-1 colorimetric assay was used. Cells were seeded onto a 96-well plate for 24 hours prior to transfection at $4x10^3$ cells/well. Transfection mixture volume was adjusted accordingly maintaining the same ratios as in section 3.3.2 of this chapter. Every 24 hours after transfection 10µl WST-1 was added to each well containing cells and: a) negative control AO, b) targeting AO, c) DMEM media and d) DMEM media only (no cells). The absorbance of each well was measured at 450nm wavelength using spectrophotometer Varioskan (ThermoFisher).

3.4 Results

3.4.1 Differential gene expression changes and enrichment of DAVID terms indicate a role of Tra2 proteins in the regulation of DNA damage response genes

Differential gene expression analysis using DeSeq identified a large number of genes showing up- or down-regulation. The function of these genes was investigated using GO enrichment analysis (Figure 31 - data from Dr Best's thesis). Because Tra2 proteins are mainly implicated in controlling splicing rather than regulation of transcription, the gene expression changes were attributed to secondary effects of the Tra2-double knockdown. However, looking at the results of GO enrichment analysis of the differentially expressed genes we identified a large subset of genes involved in two major cellular processes: Cell cycle regulation and DNA damage response/repair. These data suggest the effect of the knockdown on the cells elicits a DNA damage response and also interferes with the cell cycle. Such response may be a consequence of the splice changes in genes that are involved in DNA maintenance and repair. *CHEK1* is one example of a DNA repair gene being directly regulated by Tra2-proteins that has a clear role in managing DNA damage accumulation. Therefore, we searched the new and expanded Tra2-dependent differentially spliced gene list for genes involved in DNA damage repair/response. Interestingly, the most recent GO enrichment analysis highlighted nearly 10% (47/556) of the genes are involved in '*Cellular response to DNA damage stimulus',* but after the Benjamini-Hochberg adjustment the term is no longer significantly enriched. This does not however change the fact that a relatively large number of alternatively spliced genes are involved in DNA damage response - including *CHEK1*.

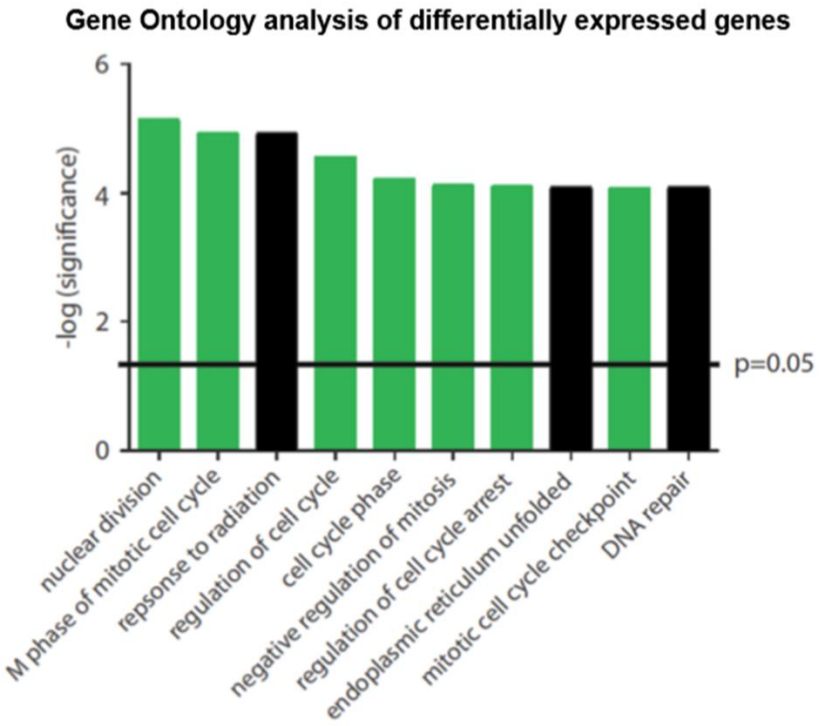

Figure 31 Differentially expressed genes upon Tra2 double knockdown in MDA-MB-231 cells. Data from Dr. Best's thesis

3.4.2 Important components of different DNA repair pathways are alternatively spliced in Tra2α/β knockdown MDA-MB-231 cells

DNA damage can occur at an increased rate in fast replicating cancer cells. Mutations that occur in non-cancer cells can give rise to oncogenes or alter the expression/function of tumour suppressors which can lead to the formation of cancer cells. However, in cells that are already cancerous that do not depend on extracellular signals for growth, further mutations can be detrimental to cell survivability. In response to DNA damage cells can activate several DNA repair pathways to prevent them undergoing cell cycle arrest or apoptosis. Tra2 proteins may therefore have a role in preventing cell cycle arrest or apoptosis by regulating alternative splicing. This idea was put forward by Dr Best upon discovering that Tra2 proteins are essential to maintain the correct splicing of *CHEK1*, an important checkpoint kinase that is involved in DNA repair and cell cycle checkpoint control. The use of MISO and MAJIQ has helped us identify a substantial number of additional alternatively spliced genes in the Tra2 depleted MDA-MB-231 cells involved in different DNA repair pathways (Figure 32). The splice changes affect genes involved in several pathways such as excision repair, mismatch repair and DNA single/double strand break repair. We hence investigated splice changes in some of the most important genes involved in these different repair pathways. In each case we consider the role of the gene in the DNA repair pathways, the amplitude of splice changes

measured and crucially the predicted outcome of the splice switch on the function and structure of the protein.

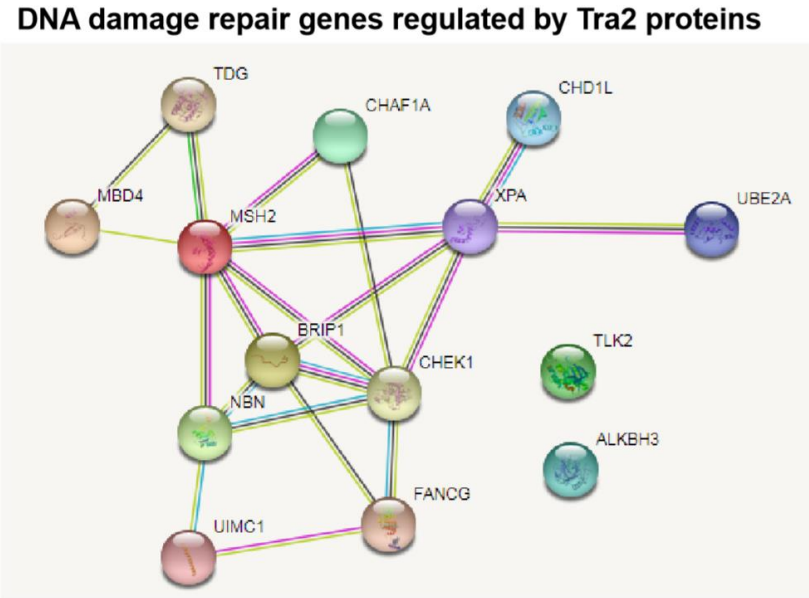

Figure 32 A subset of the DNA damage repair genes that are alternatively spliced in the Tra2 double knockdown cells. Diagram produced by STRING database of protein-protein interactions

3.4.2.1 Tra2 proteins control several genes in the DNA Strand Break Repair pathways

DNA double and single strand breaks can occur due to a range of endogenous and exogenous factors. The Non-Homologous End Joining (NHEJ) and Homology Directed Repair (HDR) pathways are mechanisms used by the cells to repair DNA strand breaks. The CHK1 protein is an important kinase that regulates cell cycle progression in response to DNA damage via the ATR-CHK1 pathway. This pathway is essential for the survival of many cell types (Smith, Tho, Xu, & Gillespie, 2010), however the role of CHK1 is mainly as an effector to DNA damage. Among the alternatively spliced targets regulated by Tra2 proteins in MDA-MB-231 cells are many other genes involved in the NHEJ and HDR, such as *MTA1* and *PARPBP* (Figure 33). Initially identified through a screen of genes in metastatic cancer cells, *MTA1* (*Metastasis Associated 1*) encodes a transcription factor that acts as a transcriptional corepressor and coactivator, but also plays a role in p53-dependent and -independent DNA repair following genotoxic stress (D.-Q. Li et al., 2010). The alternatively spliced exon 4 in *MTA1* is one of the few repressed exon cases in our dataset. In the Tra2 double knockdown cells exon 4 of *MTA1* is included 21% more than in the control cells. The additional 51 nucleotides are predicted to include an extra 17 amino acids within the Bromo-Adjacent Homology (BAH) domain - a domain that is well conserved across different MTA genes and their isoforms (Sen, Gui, & Kumar, 2014). While this is a small change to the protein sequence with a relatively low dPSI, it is

important to notice that it affects the sequence of the BAH domain and can potentially affect the role of *MTA1*.

PARP1 Binding Protein (PARPBP) gene encodes a protein required to suppress inappropriate homologous recombination. There are two alternatively spliced exons in *PARPBP* that refer to the same isoform switch. Exon 3 and exon 5 of *PARPBP* (NM_017915.4) are found to be skipped upon Tra2 double knockdown at -31% and -35% dPSI respectively. The resulting predicted transcript is anticipated to utilise a different start codon downstream within exon 7 that expresses a shorter protein of 294 amino acids. Interestingly, there are no predicted protein domains or motifs in either of the PARPBP protein isoforms, making it very difficult to predict the effect of the splice change on the function of the protein.

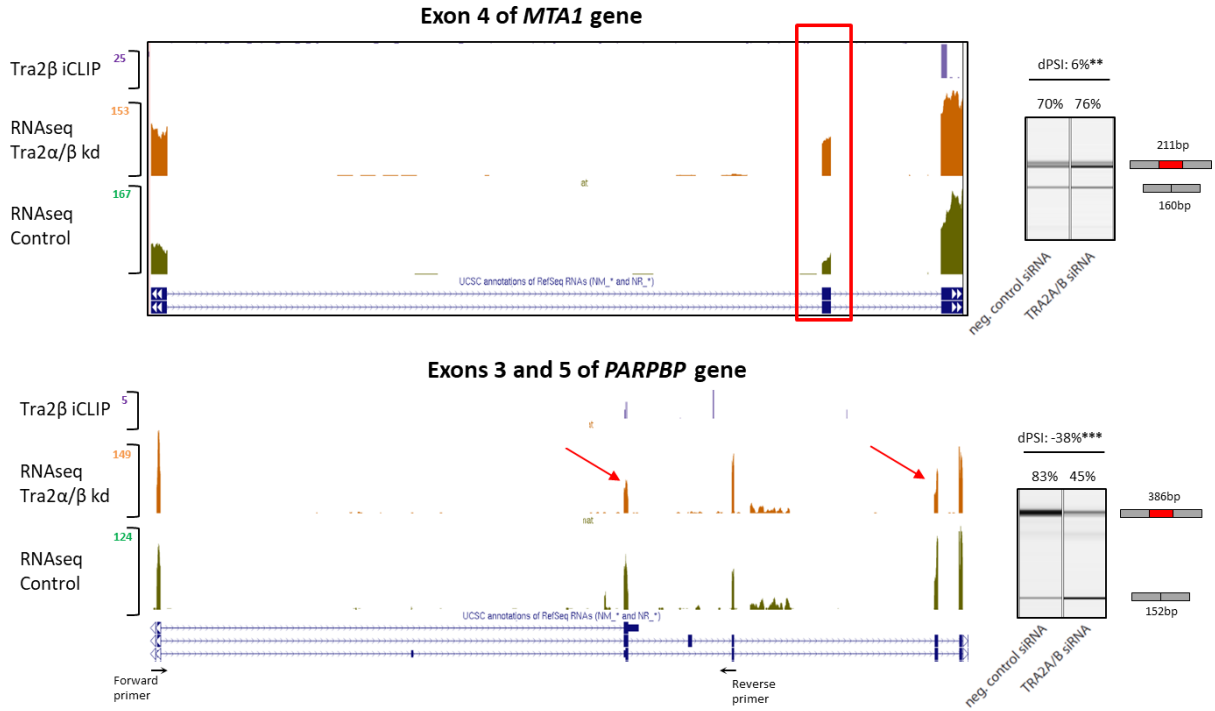

Figure 33 Alternative events in *MTA1* and *PARPBP* genes. *MTA1* exon 4 (red box) is included at a higher percentage in Tra2α/β siRNA treated MDA-MB-231 cells as shown by the RNAseq data (UCSC browser) and RT-PCR validation. *PARPBP* shows a complex set of potential isoforms, one of which was validated by RT-PCR (Primers indicated by black arrows and capillary gel electrophoresis shown). dPSI calculation was based on data from three biological repeats. ** represents $p<0.01$, *** $p<0.001$. Student's *t*-test.

3.4.2.1.1 Skipping of three consecutive exons in the *UIMC1* gene leads to a shift in the reading frame

One of the genes involved in DNA strand break repair and alternatively spliced upon Tra2 double knockdown was the *Ubiquitin Interaction Motif Containing 1* (*UIMC1*) (Figure 34). The *UIMC1* gene

encodes for the RAP80 protein which is required for the recognition and recruitment of BRCA1 to DNA damage sites (J. Wu et al., 2009). The RAP80 protein contains a C-terminal zinc finger domain and an N-terminal UIM (ubiquitin terminating motif), the latter being important for the DNA damage-induced focus formation. The MAJIQ analysis was able to detect three consecutive exons being skipped in the *UIMC1* gene. The dPSI calculated by MAJIQ varies depending on the local splicing variation (LSV) used and ranges from -54% to -21%. We tested the alternative splice junction using primers on the flanking exons and measured an average dPSI change of -61% in the Tra2 double knockdown MDA-MB-231 cells (Figure 35a). The alternatively spliced exons in this event include 968 coding nucleotides, which is not divisible by 3 and causes a shift in the reading frame. As a result, the alternative *UIMC1* mRNA expressed in the Tra2 depleted cells contains a premature stop codon and is anticipated to be targeted by nonsense mediated decay.

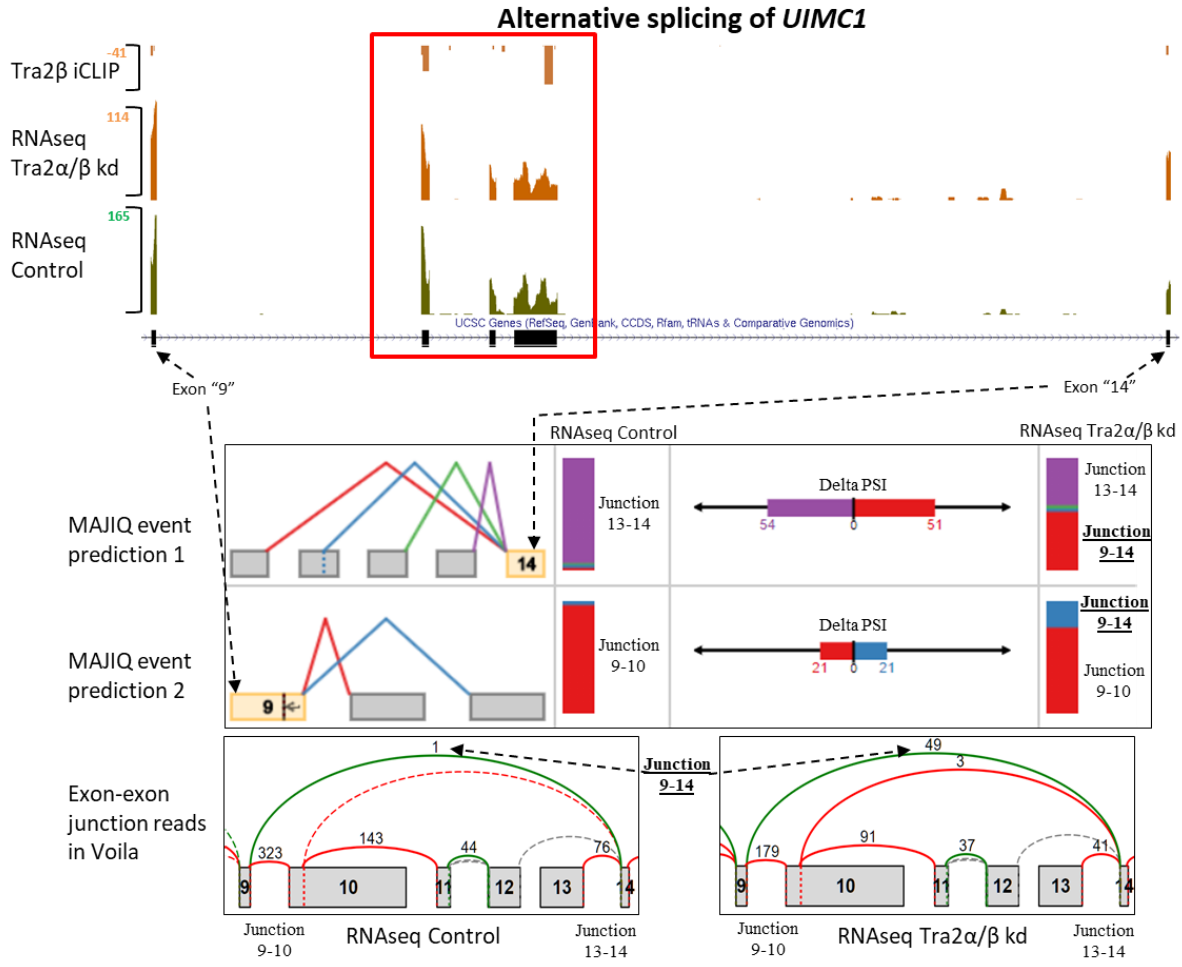

Figure 34 Alternative splicing of *UIMC1.* RNAseq data for *UIMC1* shown on UCSC browser does not evidently indicate a splice event (red box), however MAJIQ predicts skipping of 3 consecutive exons from two 'viewpoints': first by looking at junction read changes downstream of exon "14" and by comparing exon junctions connected upstream of exon "9". Note: "#" for exon numbers on MAJIQ/Voila are used as these are not necessarily the same as coding exons for the gene mRNA but rather are a count of all known exons within the open reading frame of the gene.

3.4.2.1.2 Tra2 protein depletion causes reduced expression of the full-length *UIMC1* (RAP80) protein

To test the effect of the alternative splice switch in the *UIMC1* gene we investigated changes in the levels of RAP80 protein (Figure 35). Western immunoblotting was performed on protein samples from control and Tra2α/β siRNA treated MDA-MB-231 cells extracted after 72 hours. The antibody against RAP80 detected a band just above 100kDa which significantly drops in intensity in the knockdown samples (p-value 0.001, -61%). Absence of Tra2 proteins is clearly having a direct effect on the expression of RAP80 via alternative splicing. The result measured by Western blotting show that a strong switch in dPSI is important in order to detect a significant change at the protein level. In the case of *UIMC1* the alternative splice change is clear and has a definitive effect on the stability of the mRNA isoform.

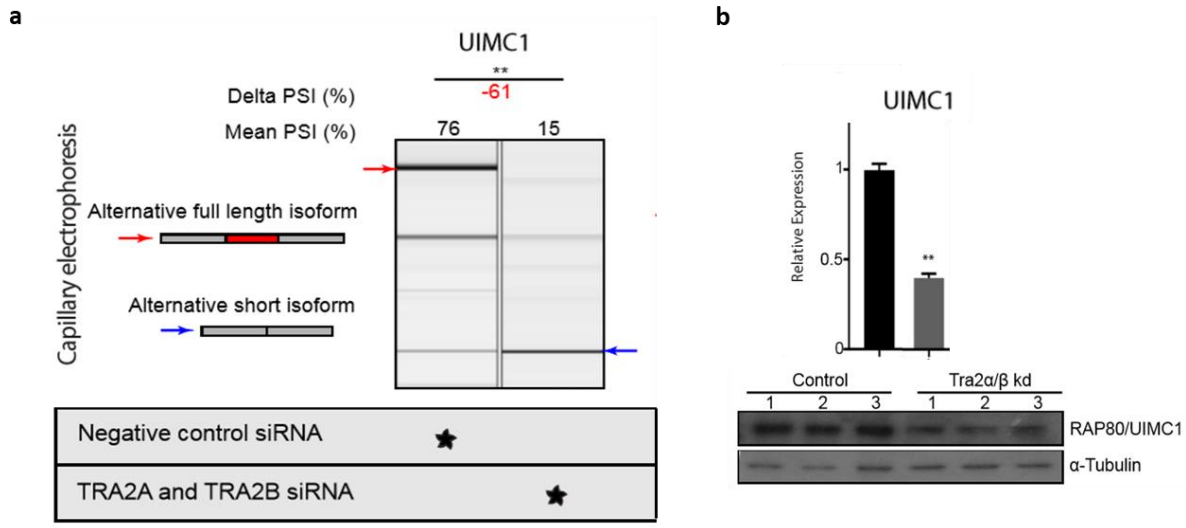

Figure 35 Validation of *UIMC1* splice switch and protein expression change. Alternative splicing verified by a) RT-PCR and an apparent reduction in protein expression by b) Western Immunoblotting. Data shown is representative of three biological repeats. Student's *t*-test, **p<0.01

3.4.2.2 Components of the Fanconi anaemia group involved in Interstrand Crosslink Repair are alternatively spliced upon Tra2 depletion

Interstrand crosslinks (ICL) are by-products of endogenous metabolic reactions or exogenous chemotherapeutic reagents that result in the covalent bonding of DNA strands and obstruct DNA replication (Muniandy, Liu, Majumdar, Liu, & Seidman, 2010). ICLs are categorised as cytotoxic lesions and it is estimated that 20-40 unrepaired ICLs can kill a mammalian cell (Dronkert and Kanaar, 2001; McHugh et al., 2001; Lawley and Phillips, 1996). The Fanconi Anaemia (FA) proteins are the most important class of proteins involved in the repair of ICLs (Andreassen & Ren, 2009).

Interestingly, Tra2 protein knockdown causes a difference in the splicing of two FA genes: *FANCG* and *BRIP1* (aka. *FANCJ*). The penultimate exon of *FANCG* is skipped in the Tra2 double knockdown cells with a dPSI of -23%. This leads to a frameshift in the shorter isoform of *FANCG* which causes a premature stop codon. However, since this stop codon would still be within the last exon of the gene it is unlikely that the mRNA will be degraded by NMD. The alternative event in *BRIP1* is also a cassette exon (exon 5/20) not divisible by 3 (Figure 36). The frameshift causes a premature stop codon very early in the mRNA sequence. Interestingly, the Fanconi Anaemia proteins do not share sequence similarity, but instead are related by their assembly into a common nuclear complex. This is evident when investigating the protein domains of the two affected genes: *FANCG* has no known protein domains in its sequence whereas *BRIP1* contains two important motifs, DEAD-2 and Helicase C-terminal domain. However, the importance of each component of the FA complexes cannot be underestimated since the repair of inter-strand crosslinks involves the recruitment of other pathways such as Nucleotide Excision Repair (NER) and Homologous Recombination (HR) (Andreassen & Ren, 2009). This means that genes such as *FANCG* with no enzymatic motifs can be crucial for interaction with other proteins from different repair pathways.

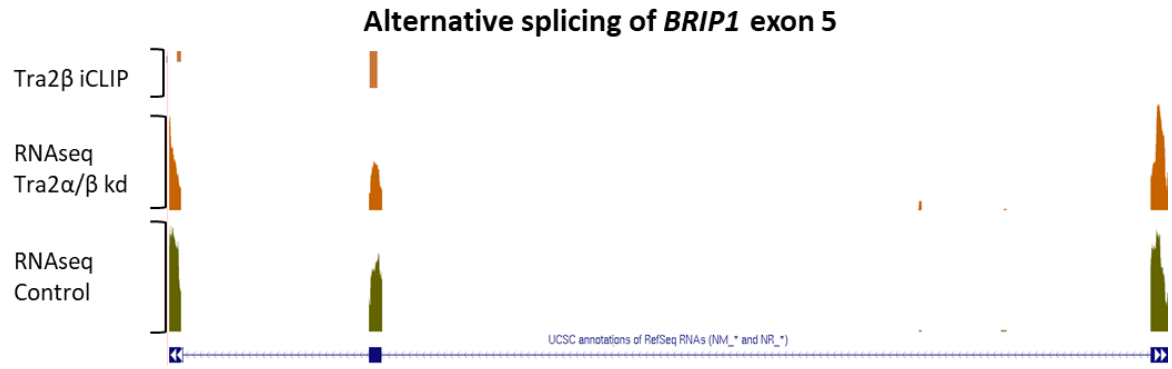

Figure 36 Alternative splicing of *BRIP1* (aka. *FANCJ*). UCSC browser view of the RNAseq data showing reduction in the reads covering exon 5 of *BRIP1* associated with Tra2β iCLIP tags. Exon 5 is a constitutive exon and when spliced out causes a shift in the ORF

3.4.2.2.1 Tra2α/β knockdown strongly influences the splicing of *BRIP1* and reduces protein expression

Although the alternative event in *BRIP1* changes splicing with high amplitude after Tra2α/β depletion, it was only MAJIQ that was able to identify this splice change. The dPSI change measured by MAJIQ was -43% but the RT-PCR results show a stronger splice switch of -65%. This is possibly due to a different knockdown efficiency and amplicon bias, or may be a difference between *in silico* prediction and experimental confirmation. A large number of iCLIP tags mapped to the target cassette exon suggesting direct binding of Tra2β. Considering the strong splice switch upon Tra2

double knockdown and the predicted mRNA degradation of the shorter isoform we tested the levels of BRIP1 protein. Western immunoblotting showed that there was a decrease in the expression of the full length BRIP1 protein in MDA-MB-231 cells treated with Tra2α/β siRNA compared to cells treated with control siRNAs (Figure 37).

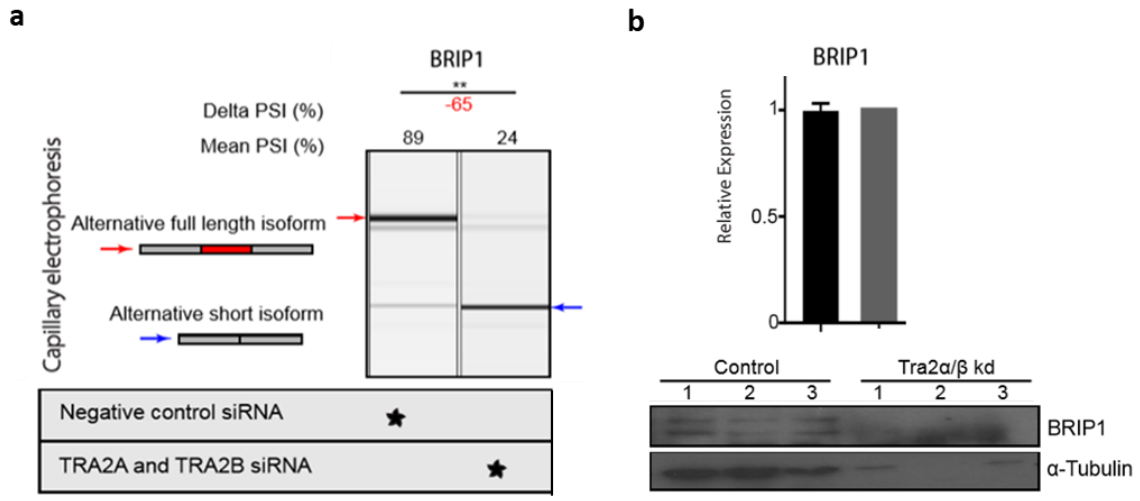

Figure 37 Regulation of *BRIP1* gene at RNA and protein level. a) Capillary gel electrophoresis of the splice junction surrounding exon 5 of *BRIP1*. b) Western immunoblotting of the BRIP1 protein in MDA-MB-231 cells treated with control and Tra2α/β siRNA. Statistical significance measured across three biological replicates using Student's *t*-test. **, $p<0.01$

3.4.2.3 MSH2 and POLD3 are genes involved in the DNA Mismatch Repair pathway that are alternatively spliced upon Tra2 double knockdown

The Mismatch Repair (MMR) pathway is a post-transcriptional repair mechanism that recognises mismatched nucleotides in the newly formed DNA double strand and differentiates between the parental and daughter strand in order to excise and replace misincorporation errors. At least two important genes involved in MMR were identified as alternatively spliced targets of Tra2 proteins: *MSH2* and *POLD3*.

DNA Polymerase Delta 3 (*POLD3*) codes for a key component of the trimeric DNA polymerase delta complex involved predominantly in lagging strand synthesis and repair. Pol-delta3 interacts with PCNA and RFC1-replication factor C complex in order to recruit POLD1, the catalytic subunit of DNA polymerase delta complex, to the DNA damage sites and carry out nucleotide base excision (Ogi et al., 2010). The alternative event that occurs in the Tra2 knockdown cells implicates the use of an alternate last exon of *POLD3* with a dPSI of -20%, which was confirmed by RT-PCR. The new shorter isoform has a considerably shorter 3'UTR and the amino acid sequence clearly affects the CDC27 motif of the protein predicted by the Pfam database.

MSH2 is a primary member of the MutSα (*MSH2*/*MSH6*) and MutSβ (*MSH2*/*MSH3*) complexes that recognise mismatched nucleotides in the DNA double strand. The alternative event in *MSH2* is one of the 25 cases of intron retention identified by MAJIQ (Figure 38). Several Tra2β iCLIP tags associated with the final exon of *MSH2,* suggesting binding of Tra2β is required for the definition of the 3' splice site of the last exon. In the absence of Tra2 proteins the intron is potentially not well defined leading to intron retention in this case. RT-PCR using primers binding on the flanking exons and an internal primer targeting the intron confirm the RNAseq analysis results. The intron introduces a 'premature' stop codon into the mRNA, but since there are no other exon-exon junctions downstream it is unlikely that this mRNA isoform will be targeted by nonsense mediated decay (Nagy & Maquat, 1998; J. Zhang et al., 1998). The alternative isoform is likely to express a protein isoform with few different amino acids in the C-terminal of the protein that are not predicted to affect any of the protein domains on Pfam.

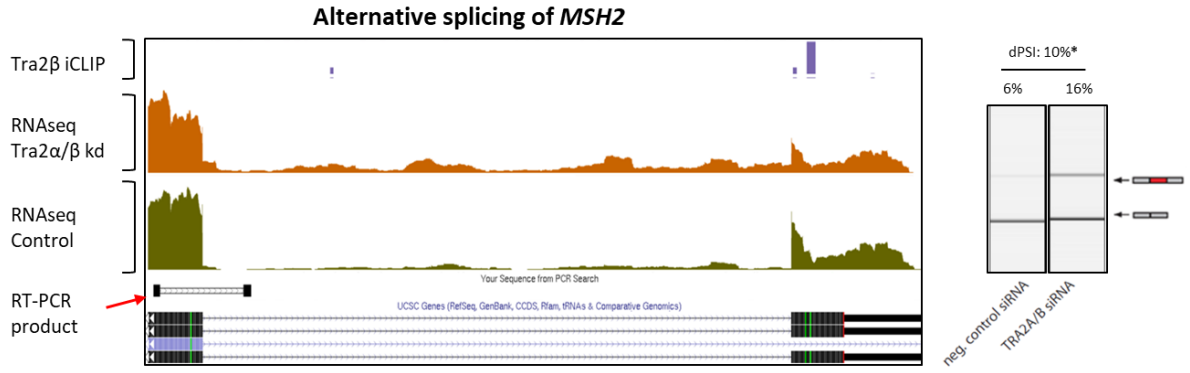

Figure 38 Alternative splicing of *MSH2*. The last intron of *MSH2* is retained in the Tra2α/β kd cells as indicated by the RNAseq and validated by RT-PCR. CGE image representative of three biological replicates. Student's *t*-test, *p<0.05

3.4.2.3.1 Tra2 depleted cells do not show sign of microsatellite instability

Defects in the mismatch repair activity of the cell are known to cause microsatellite instability, a phenomena affecting more than 15% of colorectal cancers (Boland & Goel, 2010). Considering the key importance of *MSH2* and *POLD3* in MMR we predicted that the Tra2 double knockdown may have an effect in microsatellite instability. We assayed the microsatellite instability of MDA-MB-231 cells treated with control and Tra2α/β siRNAs for 72 hours. Extracted genomic DNA was tested using the Promega MSI analysis kit which assesses the length of 5 mononucleotide repeats (microsatellites) commonly affected in breast cancer tissues. The analysis was performed by Dr Ghanim Alhilal (Newcastle University). The electrophoretically separated PCR products did not show any significant differences between the control and the knockdown samples, indicating the changes

in MMR genes induced by the Tra2 knockdown are not inducing microsatellite instability, at least not in the timeframe of the experiment (Figure 39).

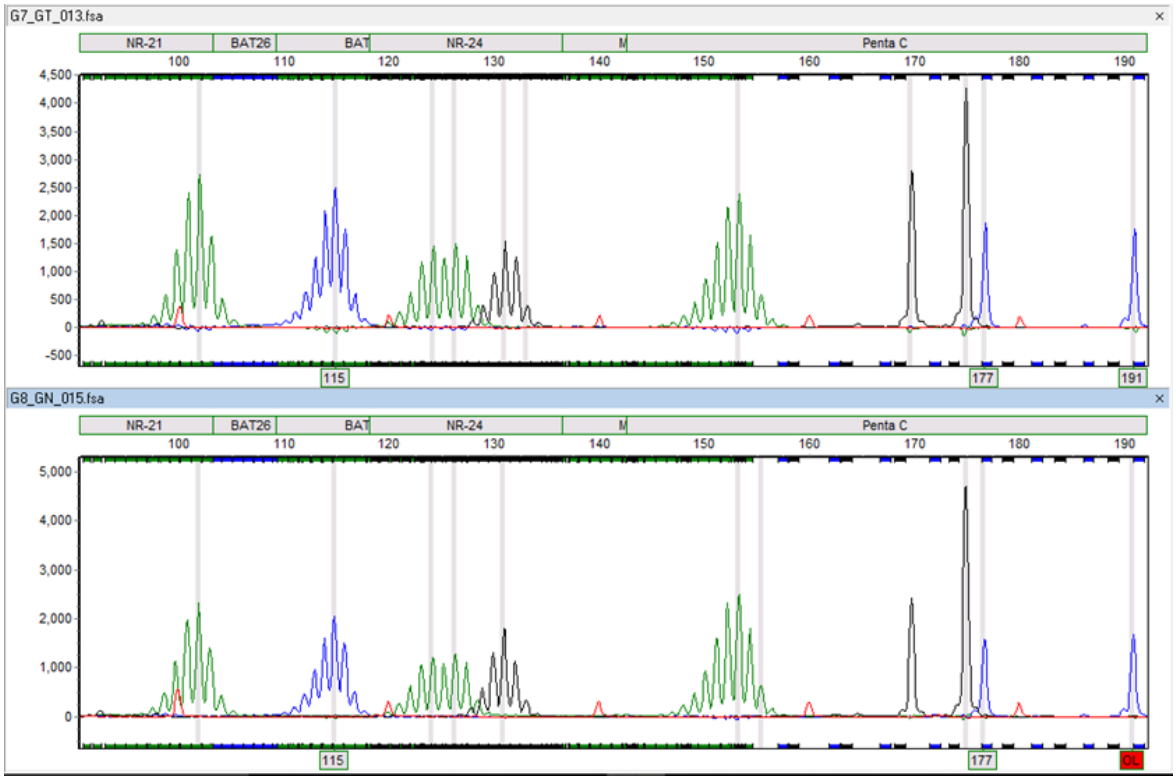

Figure 39 Microsatellite instability assay. Genomic DNA from control (bottom) and Tra2 double knockdown (top) was tested for microsatellite instability, however no change was visible as indicated by the presence of only one peak at 150 bp. Assay detects variations in the length of five mononucleotide repeats: NR-21, BAT26, BAT25, NR-24 and MONO27.

3.4.2.4 Tra2 proteins regulate alternative splicing of crucial Base Excision Repair genes

In our gene list we also identified two important genes involved in base excision repair (BER): *MBD4* (Methyl-CpG Binding Domain 4) and *TDG* (Thymine DNA Glycosylase). Interestingly, these genes are known to interact with each other and compensate for each other's role to some extent in the repair of mismatched DNA (Sjolund, Senejani, & Sweasy, 2013). The Methyl Binding Domain 4 (*MBD4*) gene which was identified by MISO and MAJIQ, but not by the DexSeq analysis, shows a very strong change in PSI upon Tra2 depletion (dPSI, -74%) (Figure 40). The *MBD4* gene encodes a methyl-CpG binding protein that has glycosylase activity. The commonly expressed isoform of *MBD4* encodes a 66kDa protein that is known to specifically recognise G:T mismatches in DNA and is able to remove the mismatched thymine via BER. The RNAseq data clearly shows a dramatic change in the splicing of *MBD4* that involves the use of an alternative 5' splice site within exon 2. The alternative 5' splice site use in exon 2 correlates with skipping of the downstream exon 3, producing an mRNA isoform that is 954 nucleotides shorter. Interestingly, the skipped sequence is divisible by three, meaning the

splice event does not create a frameshift. The shorter mRNA isoform has been previously identified by Owen et al. 2007 in HeLa cells and is predicted to produce a 30kDa protein (Owen, Baker, Bader, Dunlop, & Nicholl, 2007). The Pfam protein domain prediction online tool clearly indicates that the shorter protein isoform of *MBD4* does not contain the crucial Methyl-CpG binding domain to recognise the DNA. Importantly, the shorter isoform also is missing the 'intervening region', the amino acid sequence following the MBD motif. The intervening region was shown to be necessary for the interaction of MBD4 and UHRF1 complex and the subsequent localisation to chromocenters in heterochromatin replication and formation.

Upon Tra2 double knockdown exon 2 or exon 2 and 3 together of *TDG* are spliced out at -29% and -27% dPSI, respectively. The alternative mRNA isoform in both cases is predicted to utilise a methionine start codon in exon 4 and anticipated to express a protein 143 amino acids shorter than the control expressed protein. Although the splice switch substantially reduces the size of the protein, the UDG (Uracil DNA glycosylase) motif is predicted to remain intact.

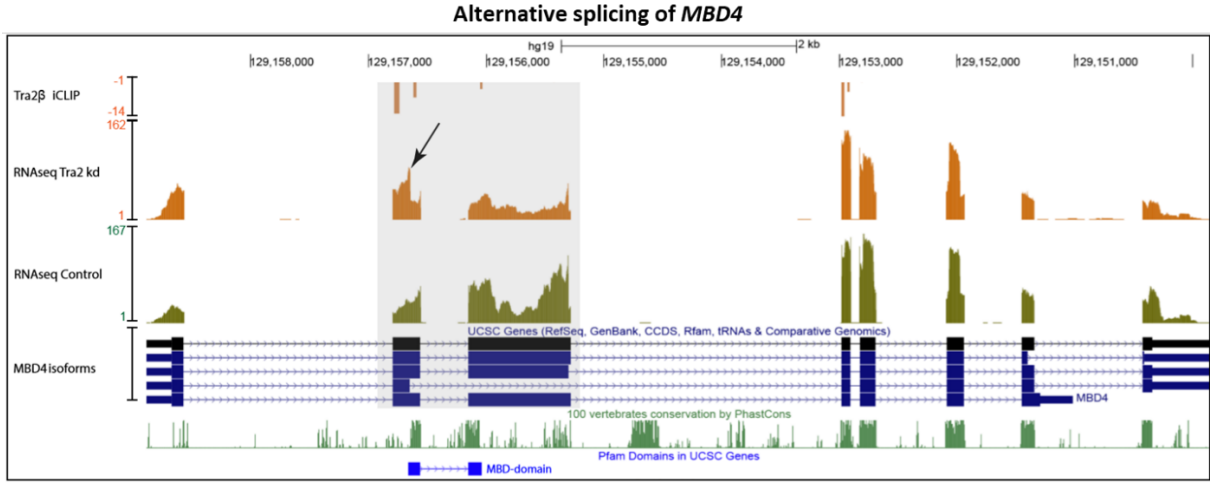

Figure 40 Alternative 5' splice site of exon 2 in *MBD4* causes skipping of exon 3 also. UCSC browser view of RNAseq indicating alternative splice site selection (arrow) and subsequent skipping of exon 3. Pfam domain track (bottom of image) indicates the MBD-domain is encoded by the spliced sequence. This sequence is also highly conserved across species – green track of conservation across 100 vertebrates.

3.4.2.4.1 *MBD4* expresses a shorter protein isoform that is exclusively localised in the cytoplasm

To assess if the alternative splice change induced by the Tra2 double knockdown is affecting the MBD4 protein we analysed the control vs Tra2α/β siRNA treated samples using Western blotting (Figure 41). The polyclonal antibody raised against the full length MBD4 protein was able to detect

the shorter isoform expressed in the knockdown samples but not in the control. Interestingly, the intensity of the band corresponding to the full-length isoform does not appear to be reduced upon Tra2 knockdown. Considering the role of MBD4 in DNA binding and repair we tested the localisation of the two protein isoforms. We then used nuclear fractionation on control and Tra2 double knockdown MDA-MB-231 cells. Western immunoblotting of the separated nuclear and cytoplasmic compartments shows exclusive expression of the full-length isoform in the nucleus and exclusive expression of the short isoform in the cytoplasm, suggesting the shorter isoform of MBD4 has lost the nuclear localisation signalling motif (Figure 42). We then attempted to visualise the localisation of the endogenous short and full-length protein isoforms of MBD4 using immunofluorescence microscopy. MDA-MB-231 cells were seeded onto coverslips and treated with control and Tra2α/β siRNA for 72 hours. However, the antibody appeared to stain all the cells uniformly and we were unable to see any differences in the Tra2 double knockdown cells.

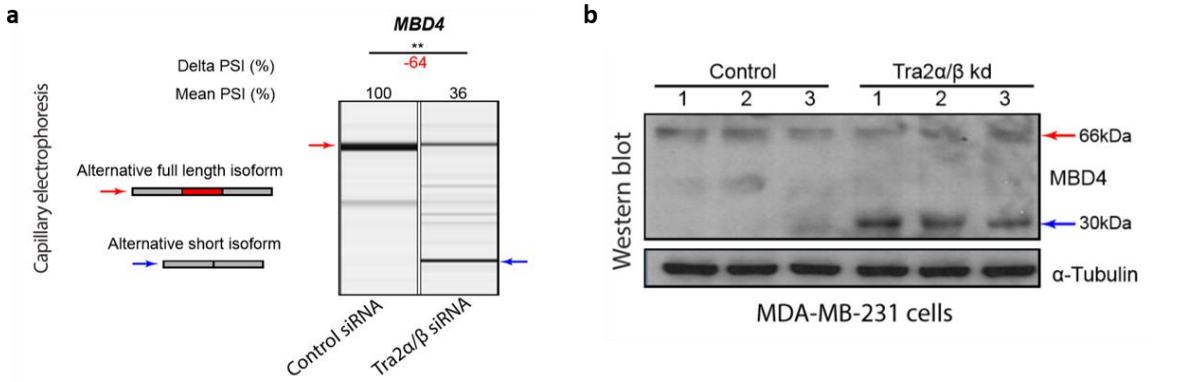

Figure 41 MBD4 alternative splicing is validated by a) capillary gel electrophoresis and the short protein isoform detected by b) Western Immunoblotting. CGE image representative of three biological replicates. Student's *t*-test, **p<0.01

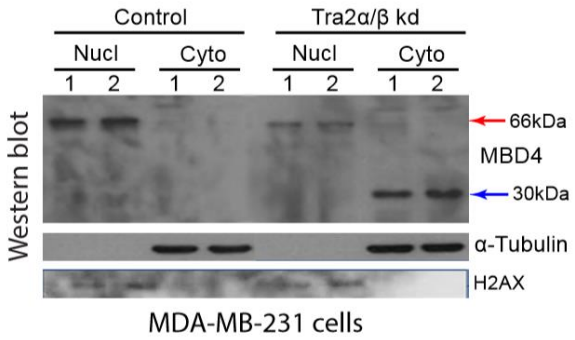

Figure 42 Nuclear fractionation experiment. Separated nuclear extract and cytoplasmic extracts show distinct expression of the shorter isoform of *MBD4* in the cytoplasm by Western Immunoblotting

3.4.2.4.2 GFP- and Flag-tagged *MBD4* short protein isoforms show predominantly cytoplasmic localisation

Although nuclear fractionation shows the two protein isoforms are clearly localised into different compartments, it is possible that the short MBD4 isoform is simply 'leaking' into the cytoplasm during the nuclear fractionation experiment. To validate the cytoplasmic localisation results obtained above we constructed a set of DNA plasmids to overexpress the short and full length MBD4 isoform with associated GFP and Flag tags. The GFP tagged plasmids were transfected into HEK293T cells, whereas the Flag tagged plasmids were transfected into HEK293T and Saos2 cells seeded onto coverslips. Immunocytochemistry clearly demonstrate that the Flag-tagged short isoform is localised in the cytoplasm and the full-length isoform of MBD4 is localised predominantly in the nucleus (Figure 43). The exclusive localisation was clearly visible in the transfected Saos2 cells as well (Figure 44). The GFP-tagged isoforms show a similar pattern but there is some overspill from the fluorescence indicating the GFP tag may be affecting the localisation of the protein, or the plasmid is expressed at a high rate.

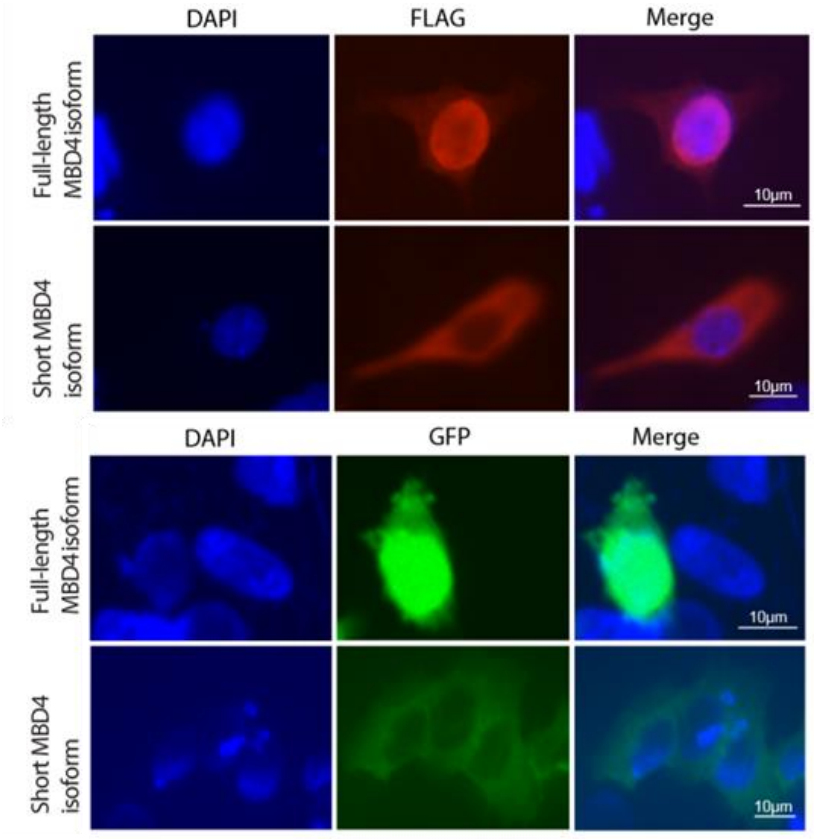

Figure 43 Fluorescence immunocytochemistry in HEK293T cells. Flag-tagged and GFP-tagged *MBD4* expression plasmids transfected into HEK293T cells show the different localisation pattern for the short and full-length isoforms. DAPI is a blue fluorescent DNA stain used to highlight the nucleus. Here, the hollowed appearance of cells expressing the short *MBD4* isoform indicates reduced presence of the protein in the nucleus. Scale bars: 10µm

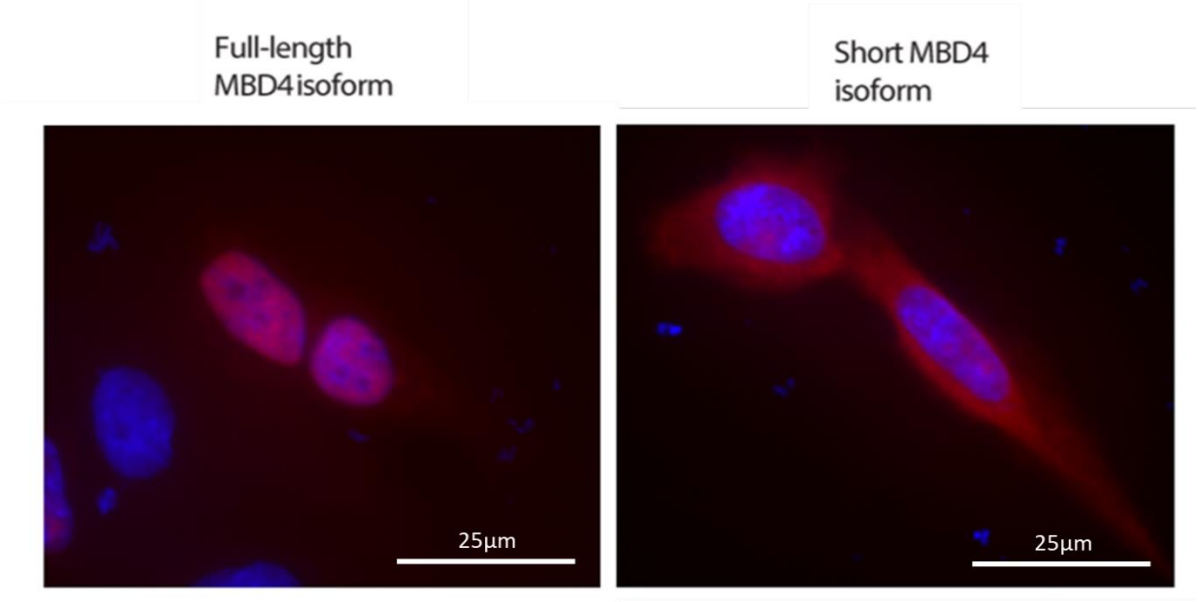

Figure 44 FLAG-*MBD4* plasmid expression in Saos2 cells. Immunocytochemistry of Saos2 cells shows a clear differentiation in the localisation of the two *MBD4* isoforms. This is highlighted by the abundantly bright overlay of the DAPI stain (nucleus) in the short *MBD4* plasmid, but not in the full-length plasmid. Merged image. Scale bars: 25µm

To try to replicate the results by Western immunoblotting the DNA plasmids were transfected into HEK293T cells seeded onto 6-well plates. After 72 hours cells were harvested, and the nuclear and cytoplasmic compartments were separated using nuclear fractionation. Western blots of the Flag tagged short isoform of MBD4 demonstrated exclusive cytoplasmic localisation (Figure 45). However, both GFP- and Flag-tagged full-length plasmids demonstrate expression in the nucleus and in the cytoplasm, indicating that the added peptide sequences may have some effect on the localisation of MBD4.

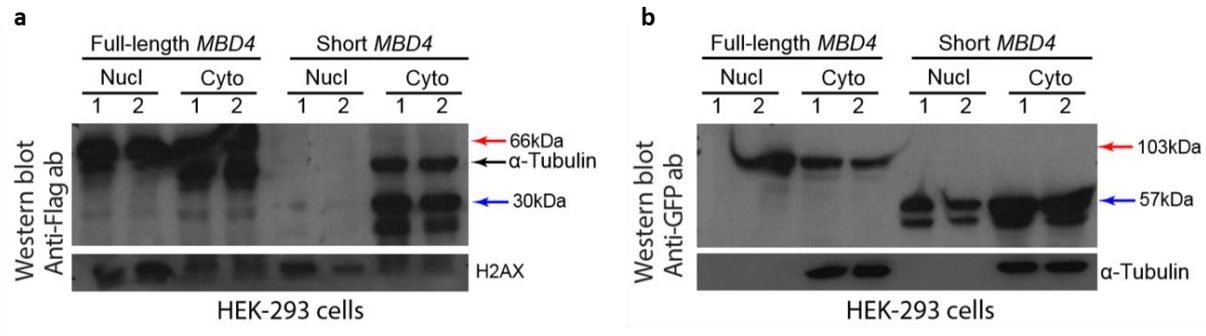

Figure 45 Nuclear fractionation of HEK293T cells transfected with a) Flag- and b) GFP-tagged *MBD4*.

3.4.2.4.3 Analysis of *MBD4* splicing control *in vitro* using a minigene

The mechanism of alternative splicing of *MBD4* is likely to be dependent on Tra2 proteins binding onto exon 2, as indicated by the high number of Tra2β iCLIP tags mapped there. To test this, we

designed minigene plasmids using the pXJ41 vector which harbours a human CMV promoter allowing the expression of the minigene in human cell lines. The *MBD4* insert was cloned into the vector within intron 2 of the rabbit β-globin gene flanked by constitutive exons. The introduction of the *MBD4* exons between these two constitutive exons was designed to test the strength of the splice sites and their response to Tra2β. This minigene construct was transfected into HEK293T cells alongside expression vectors encoding either Tra2β-GFP or GFP alone. Cells were harvested for RNA purification after 24 hours. One-Step RT-PCR analysis detected two amplification products from the HEK293T cells co-transfected with the GFP-only plasmid and the minigene. The shorter PCR amplification product (representing the short *MBD4* isoform) is not detected in the Tra2β-GFP plasmid transfected cells (Figure 46).

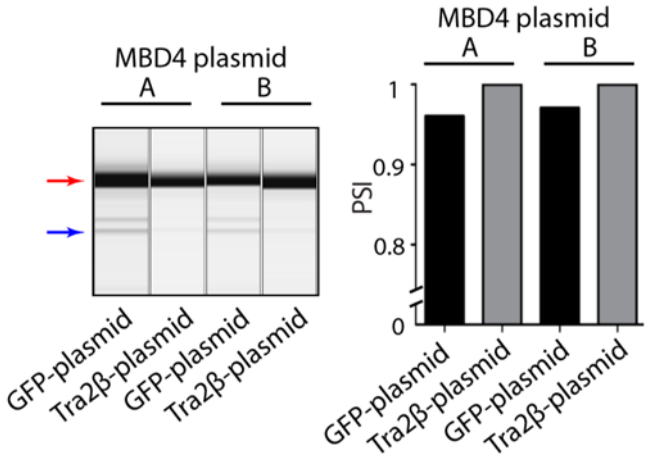

Figure 46 MBD4 minigene plasmid A and B in HEK293T cells. The introduction of the Tra2β overexpressing plasmid eliminates the expression of the shorter isoform (blue arrow) as detected by capillary electrophoresis.

3.4.2.5 Tra2 proteins regulate the splicing of XPA and other Nucleotide Excision Repair genes

Nucleotide Excision Repair (NER) is one of the most important DNA repair pathways that targets non-specific DNA lesions induced by UV light, environmental mutagens or chemotherapeutic drugs (Schärer, 2013). Deficiencies in NER are linked to xeroderma pigmentosum which predisposes sufferers to developing UV-induced skin cancer in their lifetime (Dworaczek & Xiao, 2007). In our panel of genes alternatively spliced upon Tra2 depletion there are at least three genes known to be involved in the NER pathway: *CHD1L, HMGN1* and *XPA.* The alternative cassette exon in *CHD1L* is skipped 27% more in the Tra2 double knockdown cells compared to control siRNA depletion. The presence of a few Tra2β iCLIP tags on the target exon are consistent with regulation by Tra2 proteins. The alternative shorter isoform of *CHD1L* is predicted to express a protein 28 amino acids shorter than the control isoform, and the Pfam database suggests that this shorter isoform disrupts (to some extent) the SNF2_N domain which is important in DNA repair. The alternative event in

HMGN1 includes a switch in the splicing of two mutually exclusive exons at -41% dPSI (Figure 47). However, both exons are only included in less than 5% of the total transcripts of the *HMGN1* gene making it difficult to suggest a role for these isoforms.

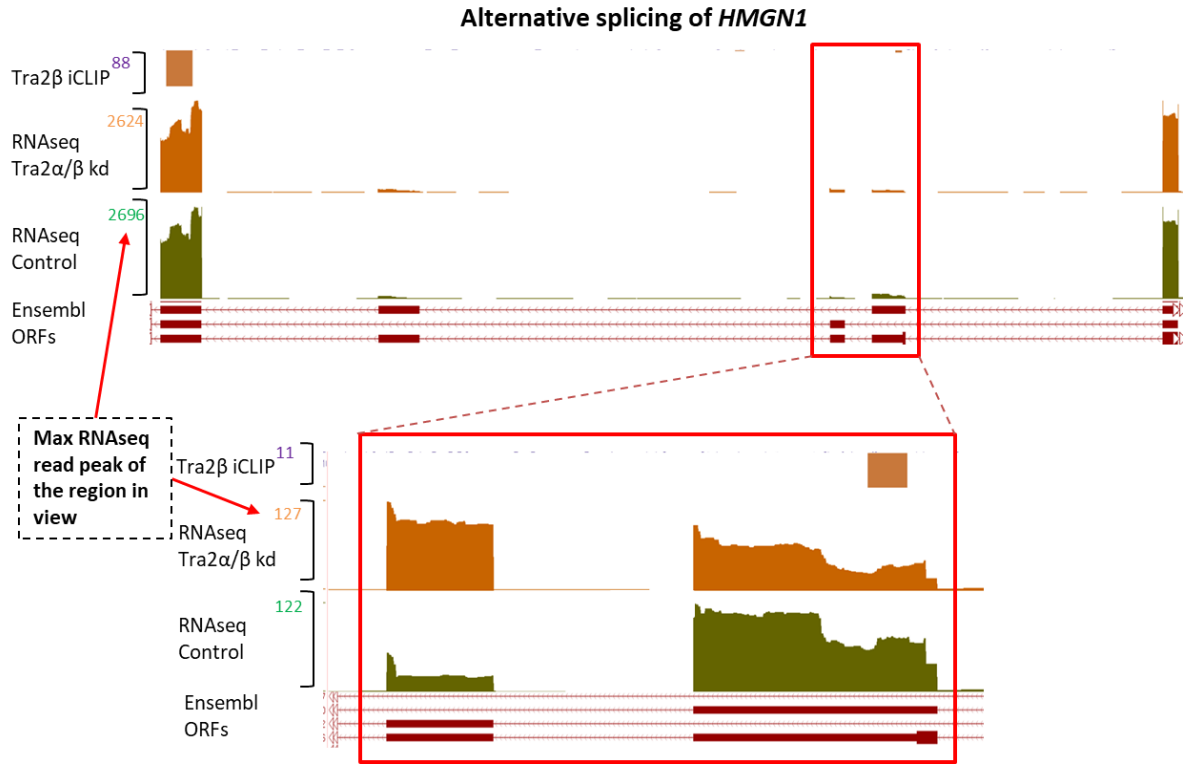

Figure 47 Alternative splicing of two cassette exons of *HMGN1.* UCSC genome browser view of RNAseq shows the isoforms including the two cassette exons are expressed far less than the main isoform of the gene (red arrows highlighting the different RNAseq depth coverage of the region viewed)

An interesting target gene that changes splicing upon Tra2 double knockdown is the *Xeroderma Pigementosa group A* (*XPA*) gene. The *XPA* gene codes for an essential component of the NER pathway that is required for the recognition of DNA damage lesions and nucleotide excision. MDA-MB-231 cells treated with siRNA against Tra2α/β express a short isoform of *XPA* that excludes three consecutive and constitutive exons of the gene (dPSI, -63%) (Figure 48). Interestingly, skipping of all three exons does not affect the reading frame of *XPA*, and a shorter protein isoform of 143 amino acids (273aa full length) is anticipated to be expressed. The Pfam domain prediction tool shows that the shorter isoform is missing both domains in the gene: the XPA_N and XPA_C domain. Moreover, previous reports have shown that the region between amino acid 74-114 is essential for the interaction of XPA and ERCC1, important for the nucleotide excision step (L. Li, Peterson, Lu, & Legerski, 1995), and the short isoform is missing a significant part of this sequence. This suggests that the NER function of the short *XPA* isoform may be severely disrupted.

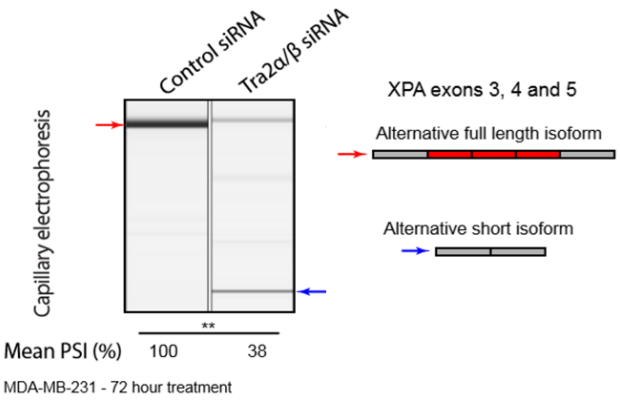

Figure 48 Capillary gel electrophoresis of XPA, showing all three exons are simultaneously skipped in the Tra2α/β knockdown cells. CGE image representative of three biological replicates. Student's *t*-test, **p<0.01

3.4.2.5.1 Tra2 double knockdown cells express the short protein isoform of *XPA*

The alternative switch in the *XPA* gene is a complex event and was only identified by MAJIQ. To validate the splice change observed in the RNAseq data we designed PCR primers to amplify between exon 2 and 6 of the *XPA* mRNA. RT-PCR from Tra2 double knockdown mRNA samples detected the expression of the shorter isoform at the expected size, while the control samples only express the full-length isoform. Considering that skipping these exons is not predicted to cause a frameshift in the *XPA* mRNA, we tested the XPA protein expression using a polyclonal antibody raised against the full length XPA protein. As predicted, the Tra2 double knockdown cells expressed the 16kDa short isoform of XPA but the control treated cells did not (Figure 49). Interestingly, the intensity of the band corresponding to the full-length isoform of XPA does not go down upon Tra2α/β knockdown.

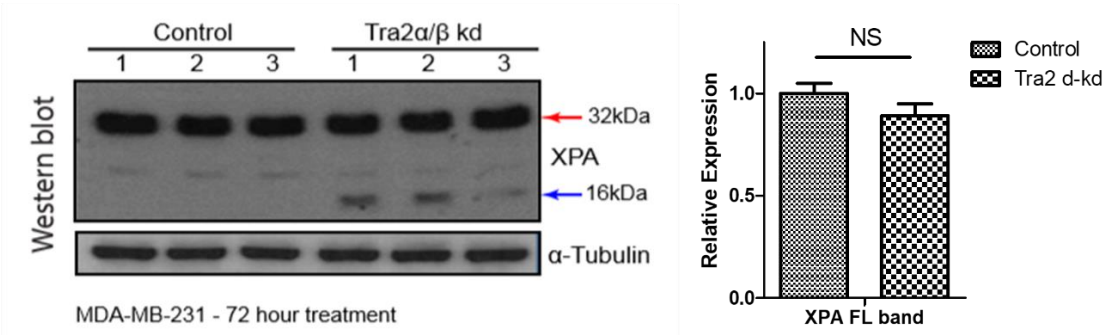

Figure 49 Western immunoblotting of XPA. Short isoform appears only in the TRA2 double knockdown cells. Quantification of the upper band detected by the anti-XPA antibody shows no significant reduction(*p*=0.08) relative to α-Tubulin. Statistical calculation based on three biological repeats. FL band, Full length band.

3.4.2.5.2 The endogenous short isoform of *XPA* is exclusively localised in the cytoplasm

Like the alternative event in the *MBD4* gene we were able to detect a shorter protein isoform of XPA in the Tra2 double knockdown cells, which prompted us to test the localisation of this isoform. Nuclear fractionation of MDA-MB-231 cell extracts (2x control vs Tra2 double knockdown) were used for western immunoblotting. The shorter protein isoform of XPA was detected only in the cytoplasmic compartment of the knockdown samples. Interestingly, the full length isoform of XPA was detected in the nuclear and cytoplasmic compartment of both the control and knockdown samples (Figure 50). The loading controls (tubulin and the nuclear H2AX) clearly show the separation of the compartments was adequate, indicating full length XPA is predominantly localised in the cytoplasm.

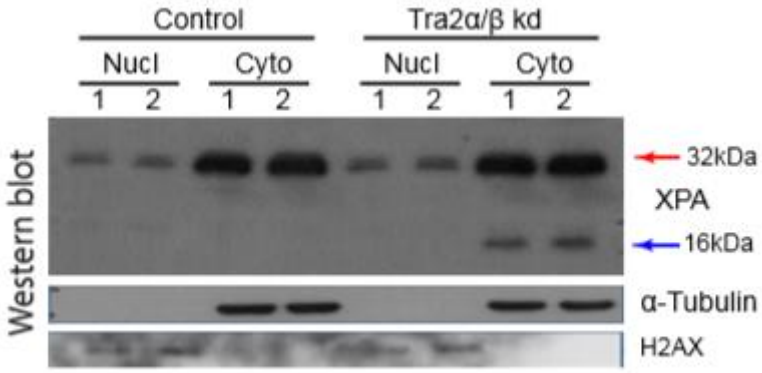

Figure 50 Nuclear fractionation of MDA-MB-231 cells. Short protein isoform of *XPA* is exclusively detected in the cytoplasmic extract by western immunoblotting.

3.4.3 Using modified oligonucleotides to induce Tra2-specific splice changes

3.4.3.1 Antisense oligonucleotides targeting the alternative events in BRIP1, CHEK1 and XPA induce the same splice changes as those observed in the Tra2 knockdown

Knocking down both Tra2 proteins in the MDA-MB-231 cells influences the splicing of many genes involved in a range of cellular processes. The impact of the Tra2 double knockdown on the cell viability, DNA damage and nuclear structure is detrimental (Best, 2014). However, it is difficult to associate a specific splice change with the global effects of the Tra2 knockdown on the cell since these splice changes are occurring all at the same time. Therefore, we tried to induce specific splice changes observed in the Tra2 knockdown cells individually by introducing antisense oligonucleotides (AO) to the cells. 2'O Methyl Phosphorothioate (2'OMePS) AOs are a commonly used form of oligonucleotides to target mRNAs and sterically block splice sites or splice signals in order to

influence the splice pattern of a certain gene (Yoo, Bochkareva, Bochkarev, Mou, & Gray, 2004). Therefore, we designed 2'OMePS AOs against some of the Tra2 responding targets from our panel of DNA repair genes. To select the best candidate genes, we considered the strength of dPSI, evidence of an effect on the protein and importantly the complexity of the splice change. Using these parameters, we designed AO targeting exon 5 of *BRIP1*, exon 3 of *CHEK1*, and exon 5 of *XPA*, at the 5' end of the exon in each case. MDA-MB-231 cells were transfected with a control AO and the gene specific AOs and were incubated for 72 hours. The splicing profile changes were measured using the same primer sets that were previously used to validate the endogenous splice changes upon Tra2 double knockdown. RT-PCR and capillary gel electrophoresis shows strong target exon skipping in the case of *BRIP1* and *CHEK1*. To our surprise, the AO against exon 5 of *XPA* induced skipping of all three exons (exon 3, 4 and 5) simultaneously, albeit at a much lower dPSI than the other two genes (Figure 51). This suggests that exon 5 inclusion/definition may be essential for the inclusion of the other two exons.

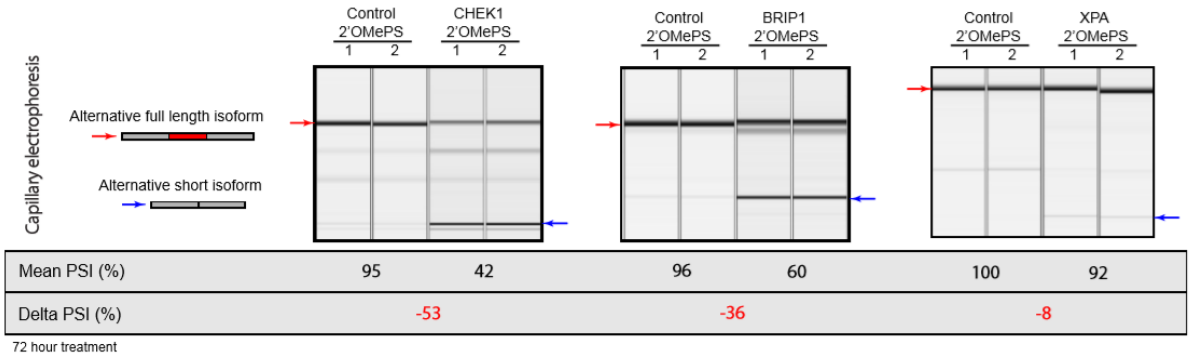

Figure 51 Capillary electrophoresis of 2'OMePS treated MDA-MB-231 cells after 72 hours. CGE image show two biological replicates. Student's *t*-test performed for all three experiments showed $p<0.05$

3.4.3.2 2'OMePS AOs affect the CHEK1 and XPA proteins but not BRIP1

The results from 2'OMePS treated cells were promising at recapitulating the splice changes occurring in the Tra2 double knockdown cells. To test if the splice changes induced by 2'OMePS AOs affect the proteins of these genes we treated the MDA-MB-231 cells over the course of 72 hours and harvested the protein extracts. Western immunoblotting of the CHK1 protein shows a significant decrease in expression after 72 hours (p-value 0.02) in MDA-MB-231 cells treated with the 2'OMePS against exon 3 of *CHEK1* (Figure 52). Although the AOs against exon 5 of *BRIP1* can successfully induce a splice change at the mRNA, Western blotting of the protein could not conclusively determine a change in the protein level. Finally, we assayed the protein expression of XPA. Western blotting was able to detect the previously identified shorter protein isoform of *XPA* confirming that the 2'OMePS AO can induce a change at the protein level.

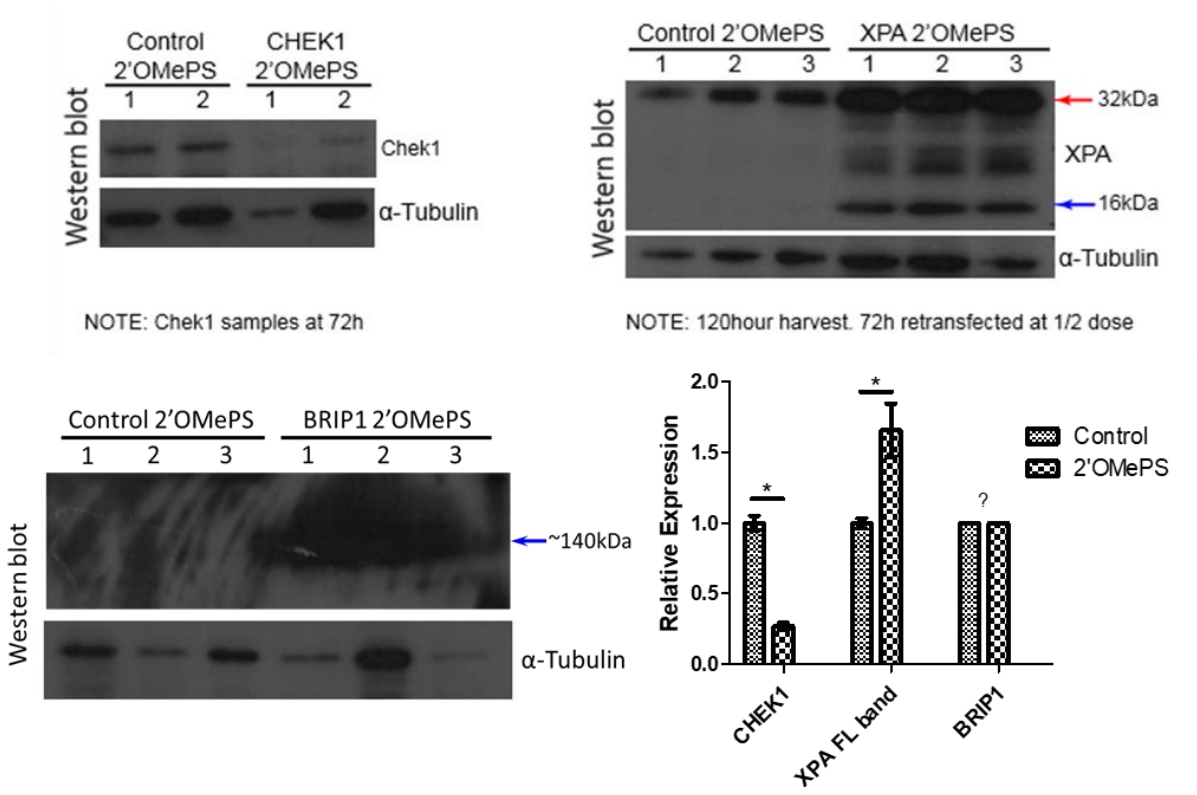

Figure 52 Western immunoblotting of 2'OMePS treated MDA-MB-231 cells. Figure shows antisense oligos targeting *CHEK1* and *XPA* successfully induce changes at the protein level. The western blot of BRIP1 proved to be difficult to conduct due to the abnormally large size of the target protein, but also the non-specific binding of the antibody. Bottom, right corner shows quantification of the visible bands relative to tubulin expression (quantified using ImageJ). * Represents $p<0.05$.

3.4.3.3 WST-1 cell viability assay of 2'OMePS treated cells

Tra2 double knockdown has a detrimental effect on the survivability of the cells. Therefore, we tested the effect of the individual 2'OMePS antisense oligos on the viability of the transfected cells. Cells were seeded onto 96 well plates for 24 hours and then treated with control AO or one of the gene specific AOs (*BRIP1, CHEK1* or *XPA*) initially and then again (half the dose) after 72 hours. The WST-1 based colorimetric assay was used to measure the cell confluency every 24 hours after the initial transfection with AOs. The results show the cells treated with *CHEK1* and *BRIP1* AOs are more viable than those treated with the control and *XPA* AOs. After day 4 cells treated with the antisense oligo against exon 5 of *XPA* show decreased viability (Figure 53).

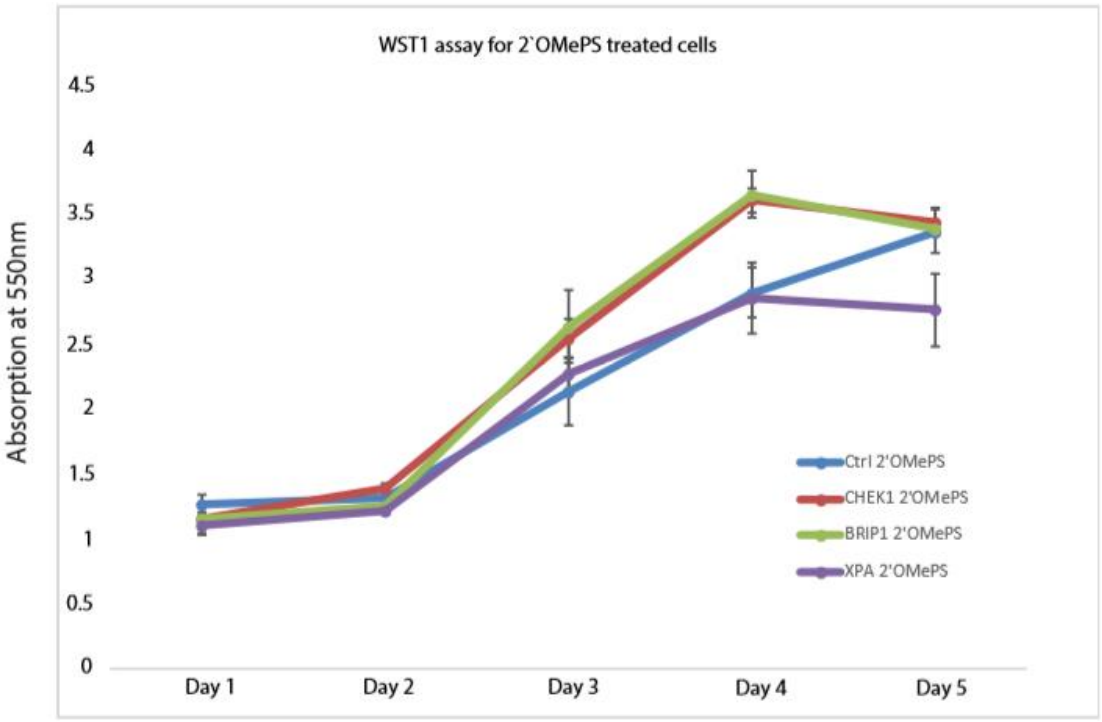

Figure 53 WST1 assay used to assess the viability of the cells after transfection with 2'OMePS. An increase in absorption at 550nm wavelength indicates cell proliferation. Bars represent Standard error, calculated based on three biological replicates at each time point.

3.4.3.4 Phosphorylated Histone H2AX levels increase in the cells treated with 2'OMePS against exon 5 of XPA

Phosphorylated histone H2AX (γH2AX) is a well-established marker of DNA damage and was shown to accumulate in Tra2 double knockdown cells (Best, 2014). The critical role of BRIP1, CHK1 and XPA in DNA repair lead us to investigate the accumulation of γH2AX in cells treated with 2'OMePS AOs. We considered that the accumulation of γH2AX would require more time therefore we treated the MDA-MB-231 cells over a course of 5 days with 2'OMePS AOs (control and gene specific). Western immunoblotting of protein extracted from the 2'OMePS treated cells using an antibody against γH2AX detects a significant increase in the total accumulation of phosphorylated Histone H2AX for *BRIP1* and *XPA* targeting AOs, indicating the induced expression of the shorter mRNA isoforms of these genes has a direct impact on DNA damage accumulation (Figure 54).

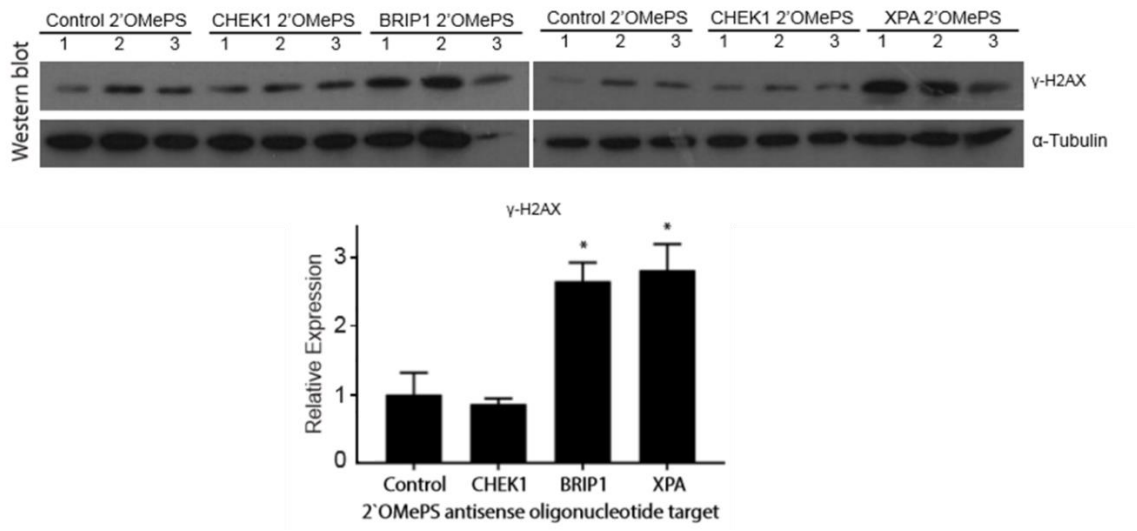

Figure 54 Western immunoblotting of MDA-MB-231 cells treated with 2'OMePS targeting *CHEK1, BRIP1* and *XPA.* *p<0.05

3.5 Discussion

3.5.1 Tra2 proteins regulate the alternative splicing of DNA repair genes important to cancer cell survival

In chapter 2 we attempted to validate protein isoform changes of genes from enriched GO terms connected with chromatin, but western immunoblotting did not identify a significant change for any of the proteins tested. In order to infer a role for the transcriptomic changes identified by the bioinformatics pipelines, protein changes must be detected in the MDA-MB-231 cells by following 72 hours of siRNA transfection. The gene ontology results from the differential gene expression changes are an indicator of the necessity of the cells to adapt to the Tra2 double knockdown. The enrichment of 'DNA repair' and 'Response to radiation' GO terms for differentially expressed genes directed our attention to alternative splice changes in genes involved in DNA damage repair. 47 components of the DNA damage repair pathway were identified as potential targets of Tra2 proteins, showing strong dPSI changes that in most cases directly affect the ORF (including the *CHEK1* gene) and sometimes protein domains. Interestingly, components of different DNA repair pathways are differentially spliced upon Tra2 double knockdown, suggesting potential compensatory roles of these pathways may be prevented in the case of Tra2 double knockdown (Figure 55). The implication of Tra2 proteins in regulating DNA repair genes in breast cancer cells indicates a potential role of Tra2 proteins in cancer progression. While the double Tra2 knockdown experiment has a detrimental effect on the breast cancer cell survivability, the overexpression of Tra2 proteins reported in various cancer types (Fischer et al., 2004b; Watermann et al., 2006) would suggest the two homologs act to protect the cells from further DNA damage accumulation *in vivo*, allowing them to sustain their already established phenotypes and prevent DNA damage induced apoptosis.

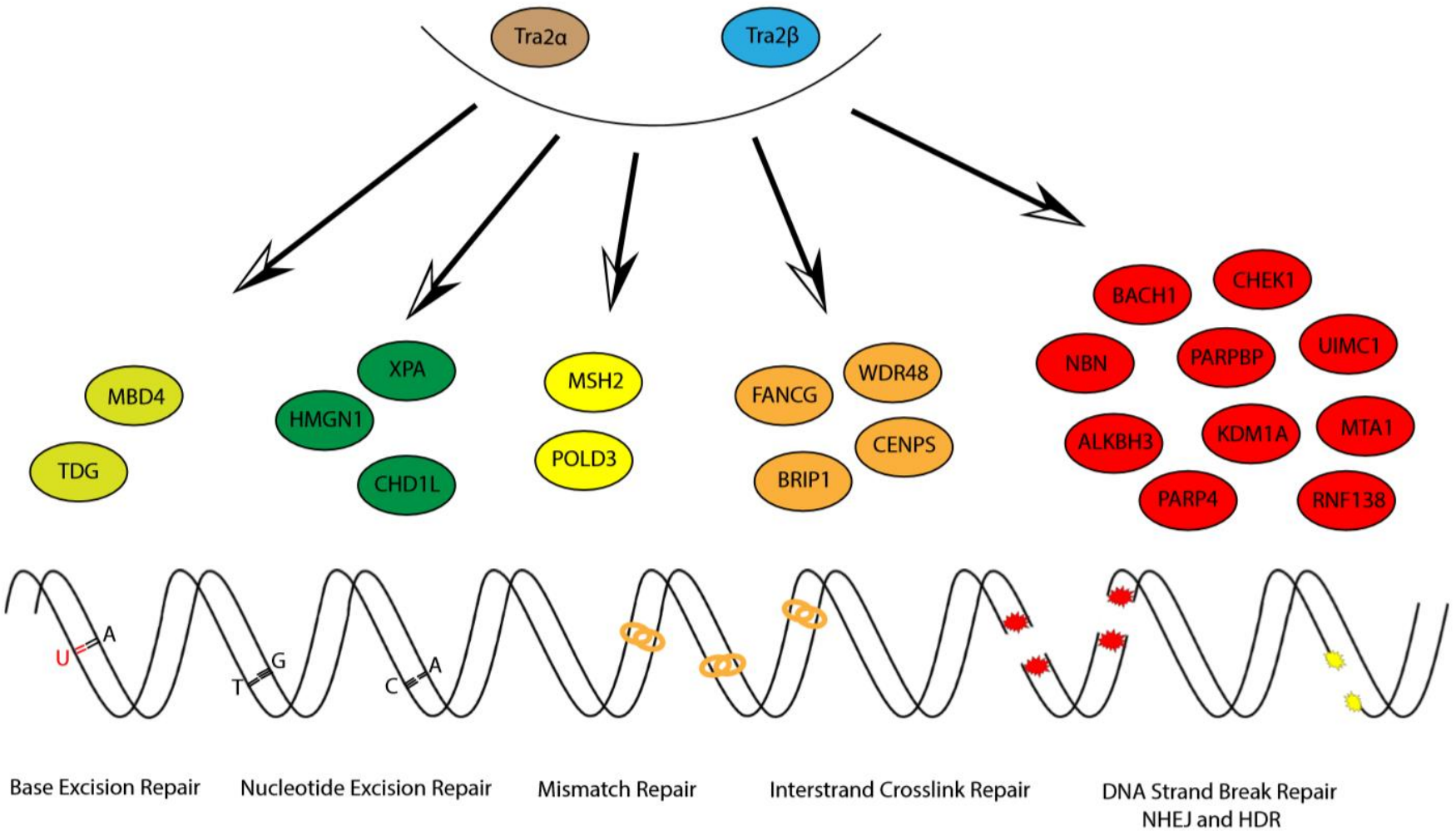

Figure 55 Tra2α and Tra2β regulate alternative splicing of various components of different DNA repair pathways in breast cancer cells.

3.5.2 Tra2 proteins regulate the expression of protein isoforms of DNA repair genes

Among the 47 alternatively spliced genes of the DNA damage repair pathways we investigated protein expression in four genes (*BRIP1*, *MBD4*, *UIMC1* and *XPA*) and detected changes in protein isoforms in all of them. Together with changes in the protein level of *CHEK1*, previously reported by our lab (Andrew Best et al., 2014), our data suggests that Tra2 proteins indeed regulate the protein expression of genes involved in DNA repair via splicing control. GO enrichment analysis of alternatively spliced genes did not highlight 'DNA repair/damage response' as an enriched term, suggesting such analyses must be used in conjunction with other methods to identify regulated pathways. Importantly, the tested genes were carefully selected based on strong dPSI changes and predicted effect on the protein structure or translation, factors that must be considered when investigating protein expression.

3.5.2.1 Protein expression of BRIP1 and its implication in cancer biology

Alternative splicing of *BRIP1* exon 5 in Tra2 double knockdown cells results in the expression of an mRNA isoform predicted to contain a premature stop codon and likely to be targeted by NMD. The strong dPSI change and abundant cluster of Tra2β iCLIP tags indicate a direct role of Tra2 proteins regulating expression of this gene via alternative splicing. Several population and case studies suggest pathogenic germline variants in *BRIP1* are associated with increased susceptibility of breast cancer (Couch et al., 2015; Kuusisto, Bebel, Vihinen, Schleutker, & Sallinen, 2011) and triple negative breast cancer (Buys et al., 2017). However, a meta-analysis of 41 studies and over 90 000 patients specifically concludes that truncating mutations in *BRIP1* are not associated with a substantial increase in breast cancer risk (Easton et al., 2016), suggesting only deleterious mutations - on par with the alternative splicing of exon 5 - may contribute to the risk factor. Previous reports indicate *BRIP1* overexpression in breast cancer, including triple negative breast cancer, is associated with poor survival rate (Gupta et al., 2018). Furthermore, *BRIP1* is shown to be overexpressed in MDA-MB-231 and MCF7 cells compared to benign breast tissue, but is downregulated in HeLa cells compared to cervical tissue (Gupta et al., 2018; Zou et al., 2016). *In vitro* studies in HeLa cells followed by *in vivo* mouse xenograft experiments show induced overexpression of *BRIP1* enhances the antitumour activity of cisplatin by inhibiting Rac1 GTPase activity (Y. Liu et al., 2016).This falls in line with the principle mechanism of action of cisplatin whereby DNA repair pathways unable to cope with the increased DNA damage causes cell cycle arrest and apoptosis, suggesting patients with overexpressed *BRIP1* would show better response to cisplatin treatment. Our data suggests Tra2 proteins promote the inclusion of *BRIP1* exon 5 in MDA-MB-231 cells and targeting *BRIP1* exon 5 with antisense oligonucleotides can cause skipping of this exon. However, further validation of

protein expression affected by the knockdown of Tra2α/β and by the AO targeting *BRIP1* exon 5 is necessary to conclude this isoform switch directly interferes with DNA damage repair or affects the cells' susceptibility to antitumour drugs such as cisplatin, a hypothesis worth testing in the future.

3.5.2.2 A cytoplasmic location of the short MBD4 isoform suggests loss of the nuclear localisation motif

The alternative event of *MBD4* revolves around the expression of a previously identified alternative isoform that expresses a truncated *MBD4* protein maintaining its uracil DNA glycosylase activity (Owen et al., 2007). The selection of an alternative 5' splice site within exon 2 of *MBD4* consequently causes the skipping of exon 3 in Tra2 double knockdown cells. Expression of the shorter protein isoform was confirmed by Western blotting. Immunocytochemistry and nuclear fractionation showed the short *MBD4* isoform is exclusively localised in the cytoplasm. Previous reports have identified at least four putative nuclear localisation signals located within or near the Methyl-CpG Binding Domain (abbreviated MBD) (Bellacosa et al., 1999; Owen et al., 2007), indicating the short *MBD4* protein isoform expressed in Tra2 double knockdown cells lacks these localisation signals and is therefore unable to enter the nucleus post-translation. Furthermore, reports from colorectal cancer cells show nuclear localisation of a truncated *MBD4* protein that is missing the C-terminal glycosylase domain but retains the nuclear localisation signals near the MBD (Suzuki et al., 2016). *MBD4* was originally identified and classified due to the presence of the "Methyl-CpG Binding Domain", which is conserved across all four members of the MBD family and *MeCP2* (Brian Hendrich & Bird, 1998). This suggests a role for *MBD4* in transcription repression by binding to methylated CpG islands in promoter regions. Several studies have indeed identified transcriptional regulation of genes such as *CDKN1A, CDKN2A, MLH1* and *MSH4* through binding of *MBD4* to their methylated promoters (Kondo, Gu, Horii, & Fukushige, 2005; Laget et al., 2014). However, it is worth noting that these reports aim to reinforce the idea that all MBD family members are solely involved in gene transcription while disregarding substantial evidence that *MBD4* plays a key role in DNA mismatch repair (Cortellino et al., 2003; B. Hendrich, Hardeland, Ng, Jiricny, & Bird, 1999; Millar et al., 2002; Petronzelli et al., 2000). While it is likely that *MBD4* has a role in transcription control, there is strong evidence to suggest the methyl-CpG binding domain is important to guide the activity of the glycosylase. A study investigating the affinity of *MBD4* to mismatched DNA substrate identified increased affinity for methylated m5CpG-TpG and suggested that truncated *MBD4* lacking the MBD, retains the glycosylase activity *in vitro* but loses the m5CpG-TpG specificity (B. Hendrich et al., 1999). The study further suggests the function of *MBD4* is to counteract the mutability of m5C due to hydrolytic deamination by initiating conversion of m5CpG-TpG mismatches back to m5CpG-CpG

through base excision (Figure 56). An accurate distinction of the role of *MBD4* is important to understand the significance of our finding of the alternative splice switch in breast cancer cells. The shorter isoform expressed in the Tra2 double knockdown cells lacks the MBD and therefore is likely to have reduced affinity for methylated CpG-TpG mismatches. In addition, the cytoplasmic localisation of the short MBD4 isoform suggests it will be unable to interact with nuclear DNA. Therefore, it is difficult to assign a role to this isoform whether that is as a transcription regulator or as base excision repair protein since both require direct interaction with DNA.

The *MBD4* minigene assay used in this chapter showed a decrease in the short isoform of *MBD4* expressed in HEK293T cells overexpressing Tra2β, providing further evidence of regulation of the splice event by Tra2 proteins. However, in order to conclusively determine direct binding and regulation by Tra2 proteins, *in vitro* examination of minigenes designed by site directed mutagenesis to specifically target and interrogate the regulatory sequence elements within exon 2 of *MBD4* could be used in future experiments.

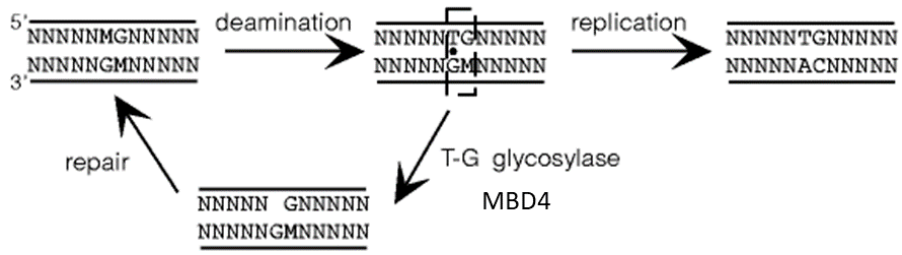

Figure 56 Proposed mechanism of action for MBD4 in repairing G-T mismatches due to deamination. Image adapted from Hendrich *et al.* 1999

3.5.2.3 Tra2 proteins indirectly regulate splicing of two upstream exons in the XPA gene

The data in this chapter indicates that double knockdown of Tra2 proteins in MDA-MB-231 cells cause three consecutive and constitutive exons of *XPA* to be spliced out of the mRNA. The resulting in-frame mRNA isoform was shown to express a corresponding truncated protein that is exclusively located in the cytoplasmic compartment of the cells as demonstrated by nuclear fractionation experiments. The *XPA* gene codes for an essential protein of DNA damage recognition and repair. Germline mutations in the *XPA* gene are a known cause for Xeroderma Pigmentosum, a rare autosomal recessive disorder with characteristic UV hypersensitivity and a 1000-fold increased frequency of cancer on sun-exposed skin (Cleaver, 1968; Kraemer et al., 1990). The XPA protein is primarily involved in the nucleotide excision repair (NER) pathway where it interacts with a range of proteins to perform nucleotide excision and replacement in DNA adducts (mainly cyclobutane pyrimidine dimers (CPD)) caused by UV irradiation of DNA (Cleaver, Lam, & Revet, 2009). Processing of damaged DNA by NER is essential for the activation of checkpoint pathways regulated by the

protein kinase Ataxia telangiectasia and Rad3-related (*ATR*). ATR phosphorylates Checkpoint kinase 1, another splice target of Tra2 proteins, and other kinases in order to arrest cell cycle progression and allow DNA repair (Zhao & Piwnica-Worms, 2001). Hela cells deficient in XPA and XPA-depleted *Xenopus* egg extracts exhibit defects in CHK1 phosphorylation during S phase (Bomgarden et al., 2006), indicating a direct relationship between XPA-directed NER and CHK1-dependent cell cycle arrest. Investigations into direct protein-protein interactions of XPA with Cep164, a chromosomal protein involved in ATR-mediated checkpoint activation upon UV damage, show amino-acids 4-97 of XPA are essential for direct binding and colocalisation at DNA damage sites (Y. R. Pan & Lee, 2009). Furthermore, UV-induced CHK1 phosphorylation was deficient in fibroblasts expressing an XPA (delta 10-88) mutant. However, these amino acid residues (4-97) are not spliced out in the alternative isoform we observe in MDA-MB-231 cells treated with Tra2α/β siRNAs, suggesting the short XPA isoform may still be recruited to CPD sites via interactions with Cep164. The nuclear fractionation experiment in this chapter indicates high abundance of the full length XPA protein in the cytoplasm of both control and Tra2 knockdown cells. This is in line with previous reports showing XPA (full length) is localised in the cytoplasm of synchronised G1 and S phase Hela cells, but is only imported into the nucleus of S phase cells upon exposure to UV irradiation (Z. Li, Musich, Cartwright, Wang, & Zou, 2013; Zhengke Li, Musich, Serrano, Dong, & Zou, 2011). The nuclear import of XPA upon UV-induced DNA damage is regulated by ATR and p53 (X. Wu, Shell, Liu, & Zou, 2007), but not by affected by expression of CHK1 or MAPKAPK2, both of which are downstream targets of ATR (Zhengke Li et al., 2011). ATR binds to XPA via the Lys188 and Ser196 residues (of XPA) in its helix-turn-helix (HTH) motif, important for the efficient repair of CPD adducts (Musich, Li, & Zou, 2017). These residues and the HTH motif are spliced out of the short XPA protein isoform expressed in Tra2 double knockdown MDA-MB-231 cells, indicating that the short XPA protein isoform is missing key interaction motifs that would allow its nuclear import by ATR. This suggests that any possible NER functions of the short XPA isoform may not be active, however, possible non-NER roles of XPA in the cytoplasm may still take place (Musich et al., 2017).

In order to investigate the individual impact of the short *XPA* isoform on cell viability and DNA damage in this chapter we used antisense oligonucleotides to interfere with the splicing of *XPA* in MDA-MB-231 cells. The 2'OMePS AO targeting the 3' splice site of intron 4 and exon 5 of *XPA* induced the expression of an mRNA product (detected by PCR) of the same size as the one observed in Tra2 double knockdown cells. Consequent Western blots of XPA and gammaH2AX indicate expression of the short XPA protein isoform correlated with an increase in DNA damage accumulation. However, there are certain aspects of the AO based experiments that must be discussed before making any definitive conclusions. Firstly, the mRNA splice change observed in the

XPA gene in the AO treated cells shows a far weaker band than Tra2α/β siRNA treated cells, yet the Western blots show a prominent band corresponding to the short isoform as well as new intermediary bands not previously detected (Figure 57). While it is possible that these intermediary bands correspond to other possible isoforms of *XPA* induced by the AO, no other PCR bands that could explain such isoforms are detected. Secondly, the cytotoxicity induced by control AOs is too high to be neglected. Both gammaH2AX levels as well as WST1 data show evidence of cytotoxicity in the control AO, suggesting that any effects on cell viability or DNA damage by the splice switch in *XPA* may be masked by the side effect of AOs. Although the data demonstrating an effect of the XPA protein isoform on DNA damage may not be conclusive, the splice change induced at the mRNA level is strongly suggestive of a regulatory mechanism where Tra2 proteins act transversely to regulate the splicing of exons 3 and 4 (Figure 58). A future potential genetic modification using a CRISPR Cas9 system to either abolish Tra2β binding sites or delete exon 5 altogether would directly answer the question of transverse regulation of splicing of exons 3 and 4 by Tra2 proteins. However, creating a stable cell line harbouring a deletion of an exon of an important DNA repair gene such as *XPA* may prove challenging. In chapter 2 we discussed other cases such as *DROSHA* where Tra2 proteins act to repress the inclusion of upstream exons. Although the exact mechanism of repression or definition for upstream exons by Tra2 proteins may not be clear, further cases confirming such regulatory junctions could change the portrayed role of Tra2 proteins in alternative splicing.

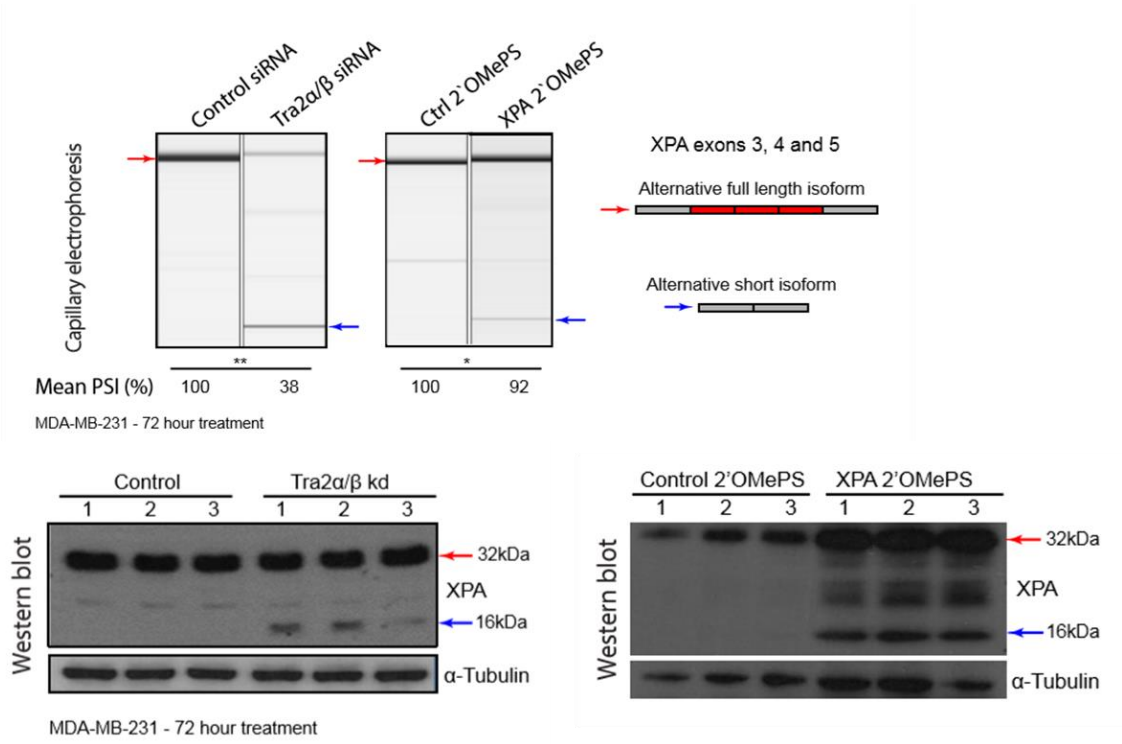

Figure 57 Detection of the short protein and mRNA isoform of XPA. Comparison of Tra2 double knockdown cells against XPA-AO treated cells showing weaker splice switch at the mRNA level but stronger protein band corresponding the shorter

isoform. Quantification of the upper bands of full length XPA protein relative to loading control are shown in figures 49 and 52. * Represents $p<0.05$

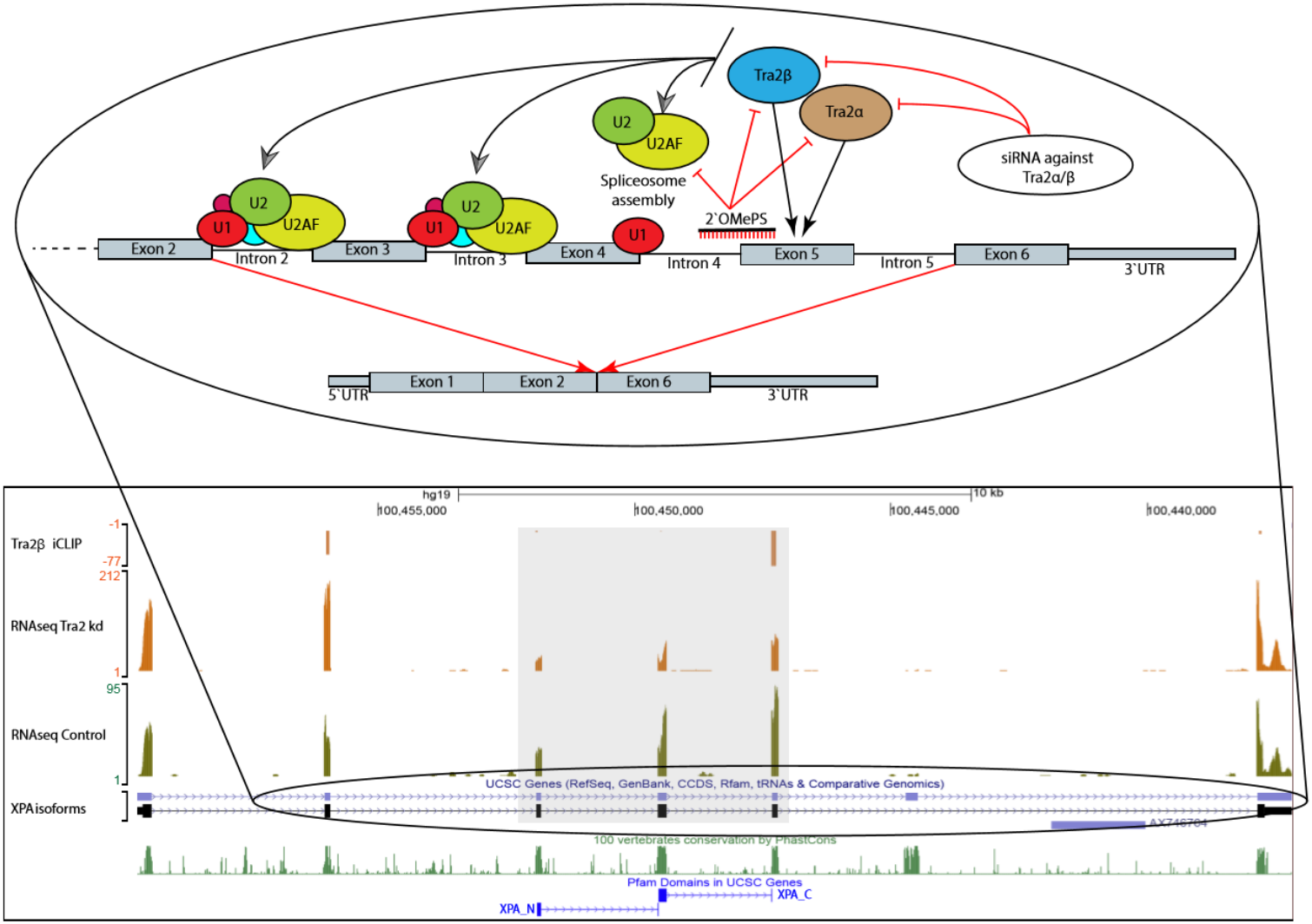

Figure 58 Alternative splicing of *XPA*. RNAseq data viewed on the UCSC browser clearly indicating skipping of three consecutive exons in the *XPA* gene upon Tra2 double knockdown (bottom panel). Diagram of the potential interplay between Tra2 proteins and splicing machinery to regulate splicing/definition of exons 3 and 4.

3.5.3 Alternatively spliced Mismatch Repair genes and their potential role in microsatellite instability

The alternative splicing events in *POLD3* and *MSH2*, two crucial genes involved in MMR (Edelbrock, Kaliyaperumal, & Williams, 2013; Nicolas, Golemis, & Arora, 2016), are located near the end of the genes. A large body of evidence, particularly in colorectal cancer, show germline mutations in MMR components such as *POLD3* and *MSH2* lead to microsatellite instability in early stage cancer development, which results in progressive late stage cancers harbouring mutations in many other tumour suppressor genes (Bubb et al., 1996; Vilar & Gruber, 2010; Zhou et al., 2018). However, our data shows no detectable microsatellite instability in the Tra2 double knockdown cells. The absence of microsatellite instability in our results could be due to the limited timeframe of the knockdown experiment which does not allow enough time for the cells to undergo multiple cell division cycles and show microsatellite instability. The alternative isoform of *POLD3* expressed at a higher level in the Tra2 knockdown cells is predicted to code for a protein with a disrupted CDC27 motif. Yeast two-hybrid experiments show the final ten residues of the C-terminal comprise the DNA sliding clamp, suggesting this isoform of *POLD3* may potentially have altered affinity for DNA (F. C. Gray, Pohler, Warbrick, & MacNeill, 2004). Moreover, the alternative final exon of *POLD3* also contains a different 3' UTR sequence. The intron retention detected in *MSH2* after Tra2 double knockdown is not predicted to affect the functional motifs in the final protein. However, it would introduce a different 3' UTR sequence, like the case of *POLD3* (Figure 59). A study investigating the mechanism by which *Helicobacter pylori* induces gastric cancer identified increased expression of miRNA-150-5p and miRNA-155-5p that bind to *POLD3* and *MSH2* mRNAs, respectively (binding confirmed by luciferase reporter assay)(Santos et al., 2017). The consequential reduced expression of *POLD3* and *MSH2* in gastric cells suggests that their 3'UTRs are essential for the endogenous regulation of expression via miRNAs. Therefore, it is possible the alternative events regulated by Tra2 proteins may cause small but important changes in the expression of *POLD3* and *MSH2*. Such splice changes *in vivo* may prove essential to the early stages of cancer development via microsatellite instability, pointing to Tra2 proteins as culprits.

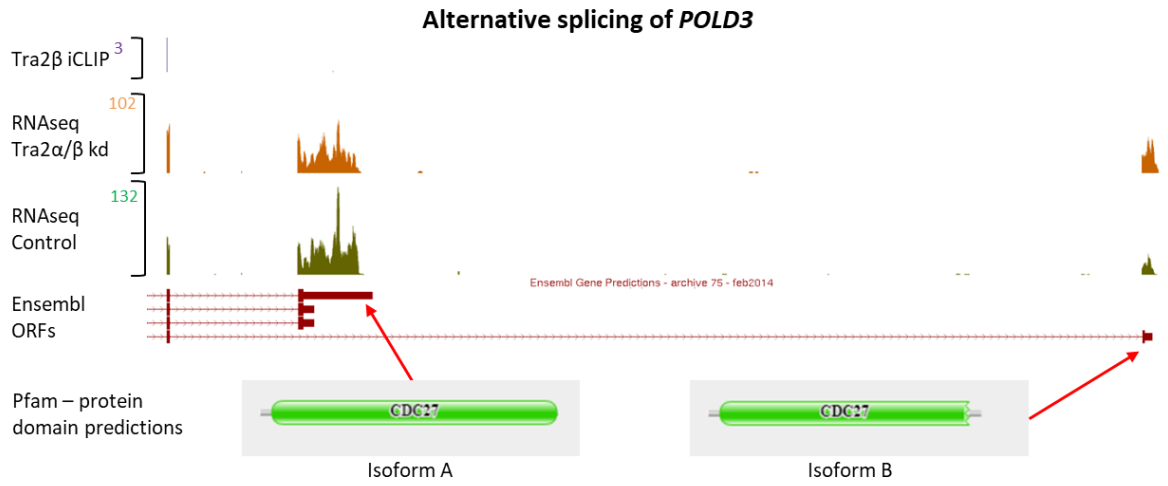

Figure 59 Alternative splicing of *POLD3*. UCSC browser view of RNAseq and iCLIP data over the alternatively selected last exon of *POLD3*. Isoform A, which is predominantly expressed in the control MDA-MB-231 cells, is predicted to produce the full structure of the CDC27 motif, whereas isoform B has a disruption at the C-terminal which is involved in DNA binding

3.6 Chapter summary

In this chapter, I utilised RNAseq differential gene expression data from Tra2 double knockdown cells which directed the attention towards a set of DNA repair genes that were not previously highlighted by GO enrichment analysis. Key components of different DNA repair pathways show strong dPSIs in Tra2 double knockdown cells and corresponding reductions in protein expression (*BRIP1* and *UIMC1*) or expression of a shorter protein isoform (*MBD4* and *XPA*). Nuclear fractionation experiments and immunocytochemistry show exclusive cytoplasmic localisation for the short protein isoforms of MBD4 and XPA. Antisense oligonucleotides targeting splice junctions of regulated by Tra2 proteins in *BRIP1, CHEK1* and *XPA* induce the same splice changes. In *XPA,* a single AO causes skipping of all three consecutive exons, followed by expression of the short protein isoform and apparent increased accumulation in DNA damage.

Chapter 4: Alternative splicing regulation by *RBMX* in breast cancer cells and the overlap with Tra2 proteins

4.1 Introduction

Regulation of alternative splicing is a complex process that involves a range of proteins and RNPs interacting with the pre-mRNA. While the role of Tra2 proteins in breast cancer cells' alternative splicing is the focus of the previous chapters, here we will investigate the potential function of the known co-regulator RBMX (also known as hnRNP G). RBMX is an hnRNP encoded by the gene *RBMX* which is located in the X chromosome and ubiquitously expresses a 43kDa protein with an N-terminal RNA recognition motif (RRM) (Figure 60) (David J. Elliott, Caroline Dalgliesh, Gerald Hysenaj, & Ingrid Ehrmann, 2019). *RBMX* has a homolog on the Y chromosome, *RBMXY.* Two other family members *RBMXL1* and *RBMXL2* are thought to have evolved via retrotransposition of *RBMX* early in mammalian evolution. Gene expression network analyses have identified *RBMX* as a key switch associated with cancer drivers (Climente-González, Porta-Pardo, Godzik, & Eyras, 2017). Interestingly, one study found RBMX is accumulated at DNA lesions through multiple domains in a PARP1-dependent manner thereby promoting DNA double-strand break repair though homologous recombination (Adamson, Smogorzewska, Sigoillot, King, & Elledge, 2012). Several previous reports have demonstrated association between expression of *RBMX* and several cancers (breast cancer (Martínez-Arribas et al., 2006), lung cancer (D. Zhang et al., 2018), endometrial cancer (Ouyang et al., 2011)).

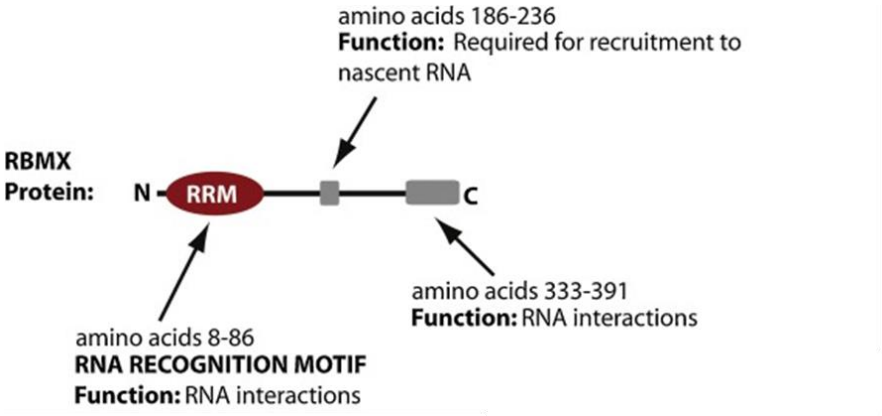

Figure 60 Protein structure RBMX highlighting the RRM and interacting motif of the protein. Image adapted from (David J. Elliott et al., 2019)

Crucially, RBMX has been known to interact with other RNA binding proteins, such as Tra2β, through the C-terminal motif (S. Grellscheid et al., 2011; Kanhoush et al., 2010). RBMX and Tra2β synergistically co-regulate the inclusion of exon 7 of *SMN2* which is of clinical importance for spinal

muscular atrophy patients (Cléry et al., 2011; Moursy et al., 2014). However, they have been shown to antagonistically regulate the skeletal muscle-specific exon of *TPM3* (Nasim et al., 2003). While the model for the synergistic regulation of *SMN2* exon 7 relies on protein-protein interaction between RBMX and Tra2β (Hofmann & Wirth, 2002), the antagonistic regulation of *TPM3* suggests they may also function as competitors (Nasim et al., 2003). Therefore, we hypothesised that some of the Tra2-regulated targets in breast cancer may also be regulated by RBMX. Furthermore, RBMX may function independently of Tra2 proteins in regulating splicing of other genes.

This chapter describes work that was chronologically performed alongside some of the Tra2 project, and therefore is at times difficult to justify the flow of experiments considering conclusions and observations made in the previous chapters. We initially identified by chance several Tra2-regulated genes to be under the control of *RBMX* too, particularly genes that were found to be repressed by Tra2 proteins. Following this discovery, we were prompted to replicate the strategy used for the investigation of Tra2 alternative splice regulation using a variety of packages to interrogate RNA sequencing, but due to time and resource limitations we could only analyse the RNAseq data using MAJIQ.

4.2 Chapter aims and objectives

The main objective of this chapter is to identify novel alternative splicing events regulated by RBMX in breast cancer cells, and any potentially co-regulated genes with Tra2 proteins. Specifically, we aim to:

- Identify alternatively spliced genes under the regulation of both RBMX and Tra2 proteins in the MDA-MB-231 cells
- Analyse RNAseq data from *RBMX* knockdown cells to search for alternative events under the regulation of RBMX independently of Tra2 proteins
- Investigate any potential autoregulatory mechanisms of RBMX expression
- Test the impact of RBMX on cell survivability

4.3 Materials and Methods

4.3.1 Cell culture and cell lines

MDA-MB-231 cells were maintained in DMEM (no phenol red) plus 10% fetal bovine serum and 1% Penicillin Streptomycin. HEK293T cells were maintained in DMEM with phenol red plus 10% fetal bovine serum.

4.3.2 siRNA knockdown

RBMX transient knockdown was established using two different siRNAs targeting *RBMX* mRNA transcripts (hs.Ri.RBMX.13.1 and hs.Ri.RBMX.13.2, from Integrated DNA Technologies). Negative control cells were transfected with scramble siRNA (IDT). Cells were seeded onto 6-well plates at a confluence of approximately $1x10^6$ and incubated for 24hours. After the incubation period, 6µl of 10µM siRNA was diluted in 150µl of Opti-MEM which was then mixed with 6µl of Lipofectamine RNAiMAX also diluted in 150µl. The combined reaction mix was incubated at room for 5 min and then added dropwise onto the seeded cells. Transfected cells were then incubated for 72h at 37°C before harvesting. RNA was extracted using standard Trizol extraction protocol and DNAse treated using DNA-free kit (Invitrogen). 2x Protein Loading buffer was used for protein extraction.

4.3.3 Tetracycline-controlled transcriptional activation of *RBMX* and *RBMXL2*

HEK293T tetracycline-inducible *RBMX*- and *RBMXL2*-overexpressing cells were established by Dr Mahsa Kheirollahi-Kouhestani. Cells were thawed and seeded for 24hr in standard DMEM with phenol red. Cells containing pcDNA empty vector, *RBMX*- and *RBMXL2*-overexpressing vectors were treated with Tetracycline at a final concentration of 1µg/ml alongside corresponding control untreated cells, which were then incubated for 24hours at 37°C. RNA was extracted using standard Trizol extraction protocol and DNAse treated using DNA-free kit (Invitrogen).

4.3.4 cDNA and PCR quantification

The RNA from siRNA-treated cells was extracted using standard Trizol RNA extraction (Life Technologies) following manufacturer's instructions. cDNA was synthesized from 500ng total RNA in 10µl reactions using Superscript VILO cDNA synthesis kit (Invitrogen) following manufacturer's instructions. To analyse the splicing profiles of the alternative events primers were designed using Primer 3 Plus and the predicted PCR products were confirmed using the UCSC *In-Silico* PCR tool. Primers were designed to have a melting temperature of 60°C (+/-1°C) often including an internal screening primer. All PCR reactions were performed using GoTaq G2 DNA polymerase kit from Promega. All PCR products were examined using the Qiaxcel capillary electrophoresis system

(Qiagen). Primers used for this chapter and corresponding coordinates of alternative events regulated by *RBMX* are listed in Appendix.

4.3.5 RNA sequencing

For the purpose of RNA sequencing, RNA was extracted from cells using RNeasy Plus Mini Kit (Qiagen) following manufacturer's instructions and re-suspended in nuclease-free water. RNA samples were DNase treated using DNA-free kit (Invitrogen). Pair-end sequencing was done initially for two samples, one of negative control siRNA treated MDA-MB-231 cells and one of *RBMX* siRNA treated cells, using Illumina HiSeq 2000. We then sequenced three biological repeats of negative control and *RBMX* siRNA treated MDA-MB-231 and mouse 3T3 fibroblasts using the same sequencing method.

4.3.6 RNAseq analysis

The base quality of raw sequencing reads was checked with FastQC. RNAseq reads were mapped to the human genome hg19 reference with Tophat2 (D. Kim et al., 2013). RNAseq mapping was performed by the bioinformatics unit at the Institute of Genetic Medicine, Newcastle University. Differential splicing analysis was determined using MAJIQ, performed by Matthew Gazzara, University of Pennsylvania (Vaquero-Garcia, Barrera, Gazzara, Gonzalez-Vallinas, et al., 2016).

4.3.7 WST-1 cell survivability assay

To test cell viability after *RBMX* knockdown we utilised the WST-1 colorimetric assay. After treatment with siRNA cells were incubated in solution for 3 hours at days 2, 3, 4 and 5 following manufacturer's protocol. Cells were seeded onto a 96-well plate for 24 hours prior to transfection at $4x10^3$ cells/well. Transfection mixture volume was adjusted accordingly maintaining the same ratios as in section 3.3.2 of this chapter. Every 24 hours after transfection 10µl WST-1 was added to each well containing cells and: a) negative control AO, b) targeting AO, c) DMEM media and d) DMEM media only (no cells). The absorbance of each well was measured at 450nm wavelength using spectrophotometer Varioskan (ThermoFisher).

4.4 Results

4.4.1 Tra2 proteins and *RBMX* have antagonistic effects on splicing of *SMN2* exon 7 in MDA-MB-231 cells

The splicing regulator RBMX coded by the *RBMX* gene has been previously identified as a co-regulator of alternative splicing working together with Tra2β. A classic example of their interaction is the co-regulation of exon 7 of the *SMN2* gene. Correct splicing of exon 7 is crucial for the production of active SMN protein in patients suffering from Spinal Muscular Atrophy (SMA) (Moursy et al., 2014). The alternative splicing of exon 7 is reliant on Tra2β binding an ESE within the exon. Clery *et al.* 2011 suggest that RBMX interacts with Tra2β to promote the inclusion of exon 7 of the *SMN2* gene (Cléry et al., 2011). Furthermore, they show that RBMX binds to CAA motifs and mutations in the binding motif of exon 7 minigenes greatly affected the binding affinity of RBMX, and subsequently influenced inclusion of the exon *in vitro.* To test if endogenous RBMX and Tra2 proteins have the same effect on *SMN2* exon 7 splicing we knocked down the expression of *RBMX* in MDA-MB-231 cells using two different siRNAs. After 72 hours of siRNA treatment cells were harvested for RNA and protein extraction. Firstly, we tested the efficiency of the *RBMX* knockdown using western immunoblotting (Figure 61). The results show a marked decrease in RBMX expression from both siRNAs. We then tested the inclusion of *SMN2* exon 7 in control, *RBMX* and *TRA2A/B* siRNA treated cells using RT-PCR. Contrary to the suggestions from Clery *et al*, 2011 we find that endogenous RBMX represses the inclusion of *SMN2* exon 7, whereas Tra2 proteins promote the inclusion of the exon (Figure 62). Our results suggest that RBMX and Tra2 proteins have antagonistic roles over the splicing inclusion of exon 7 of the *SMN2* gene.

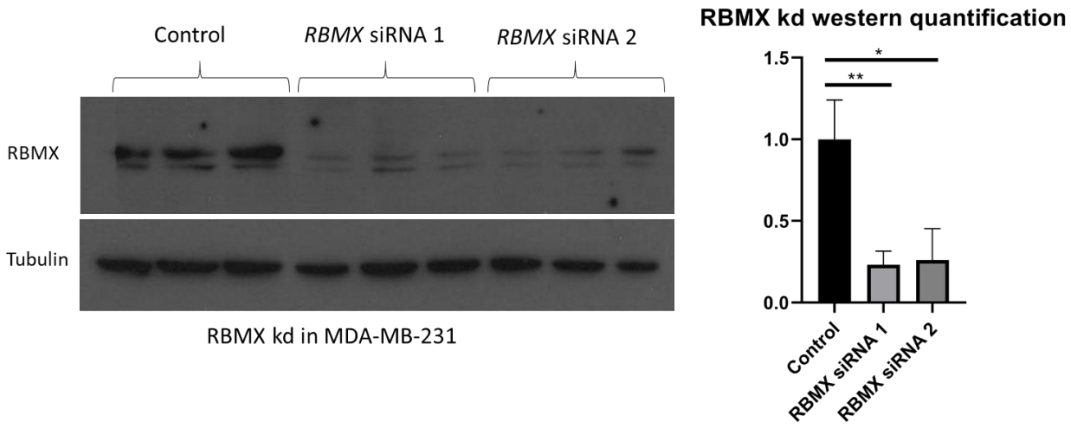

Figure 61 Western immunoblotting of MDA-MB-231 cells treated with two different siRNAs against *RBMX.* Western blot quantification was performed using ImageJ across three biological replicates. Student's *t*-test, *p<0.05, **p<0.01

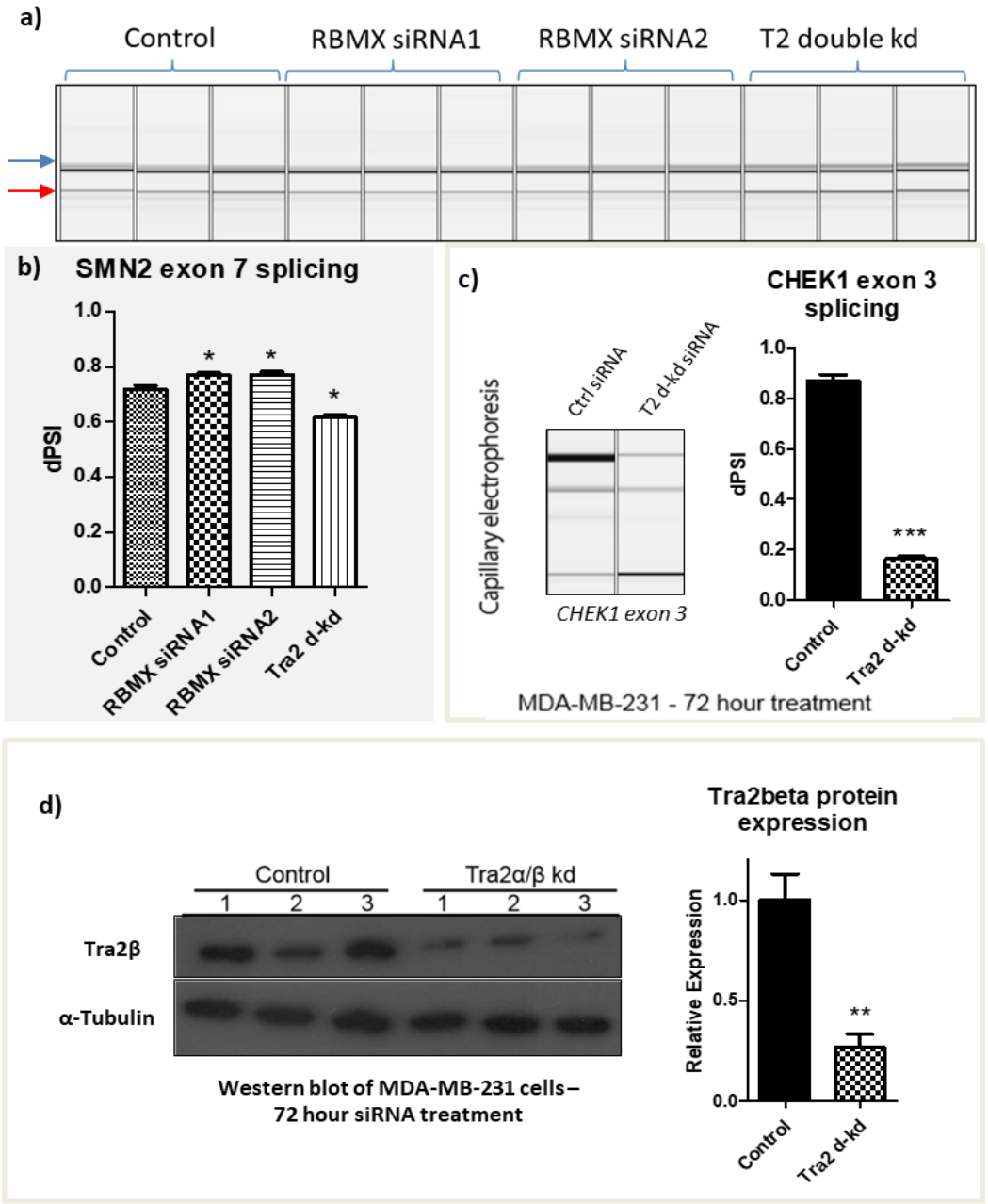

Figure 62 Alternative splicing control of *SMN2* exon 7. a) Capillary gel electrophoresis of siRNA treated MDA-MB-231 cells, blue arrow indicates exon 7 inclusion band, red arrow exon 7 skipping. b) Quantification of SMN2 exon 7 splicing in MDA-MB-231 cells. c) Indirect validation of Tra2α/β knockdown by quantification of *CHEK1* alternative splicing of exon 3. d) Western blot of Tra2α/β siRNA treated cells using anti-Tra2β antibody and corresponding quantification using ImageJ indicates clear reduction in Tra2β. All *p*-values are calculated from three biological replicates. * Represents p<0.05; ** represents *p*<0.01; and *** represents *p*<0.001.

4.4.2 Tra2-repressed exons are more likely to be regulated by *RBMX*

In line with the idea that Tra2β and RBMX interact to regulate alternative splicing we tested the splicing profiles of 19 genes from our Tra2 target panel in the *RBMX* knockdown cells by RT-PCR. These genes were selected based on the dPSI values and previous validation by RT-PCR in the Tra2

double knockdown MDA-MB-231 cells. Of the targets selected for testing, 10 exons were repressed by Tra2 proteins. Interestingly, five of the Tra2-repressed exons also respond to the *RBMX* knockdown in the same direction of change (Figure 63). Of the nine exons that are activated by Tra2 proteins only *FAM35A* exon 3 responds to the *RBMX* knockdown. Contrary to our findings in the regulation of *SMN2* exon 7 and previous reports suggesting antagonising roles of RBMX and Tra2β in alternative splicing (Nasim et al., 2003; Yan Wang, Wang, Gao, Stamm, & Andreadis, 2011), our results suggest the alternatively spliced targets regulated by RBMX and Tra2 proteins mainly change in the same direction. Moreover, repressed exons are more likely to be commonly regulated by RBMX and Tra2 proteins than promoted exons.

Interestingly, these results also indicated a clear correlation between *RBMX* knockdown efficiency and dPSI. We notice this in some examples such as *SS18* exon 8 and *HMGXB4* exon 2, where the splice change is stronger in the more efficient knockdowns. The two isoforms of *SS18* have both been previously described to interact with *RBM14* and function as nuclear receptors and transcriptional activators (Iwasaki, Koibuchi, & Chin, 2005). However, the *SS18* gene is often associated with synovial sarcoma due to it being commonly translocated and fused with other genes (particularly *SSX1* and *SSX2*)(Przybyl, van de Rijn, & Rutkowski, 2019), therefore the role of this alternative isoform in carcinogenesis may be difficult to determine. The *HMGXB4* gene on the other hand is associated with transcriptional regulation in primary myelofibrosis patients. Specifically the silencing of *HMGXB4* was noted to induce megakaryocyte differentiation and inhibit erythroid development in human stem cells (Salati et al., 2016). Upon *RBMX* knockdown exon 2 of *HMGXB4* is included 11% more. The exon 2-containing isoform codes for a protein 111 amino acids longer than the control expressed isoform (utilises a different start codon), however the two predicted domains of the *HMGXB4* gene are not disrupted as indicated by Pfam domain predictions.

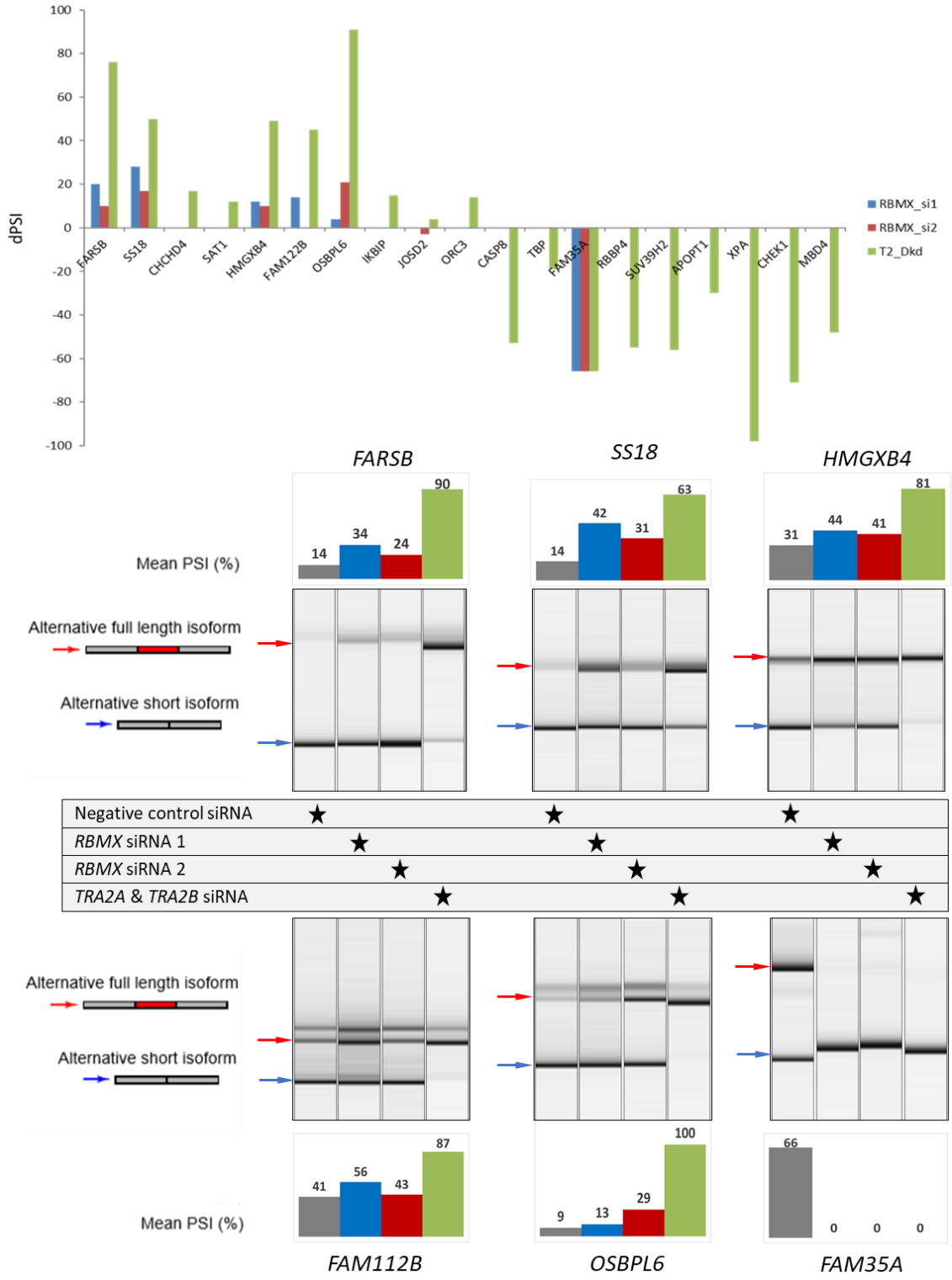

Figure 63 Delta PSI of previously identified Tra2 targets. *RBMX* knockdown MDA-MB-231 cells show significant changes in splicing of genes that are repressed by Tra2 proteins. All changes shown are statistically significant (p<0.05) and are calculated using Student's *t*-test on three biological repeats.

4.4.3 RNA sequencing of *RBMX* knockdown MDA-MB-231 cells and analysis using MAJIQ

Our preliminary results comparing regulated alternative events by RBMX and Tra2 proteins prompted us to continue searching for targets regulated by RBMX. To be able to globally compare our results from the *RBMX* knockdown with those from the Tra2 experiments we continued using the same cell line, as we had already mapped Tra2 target exons in MDA-MB-231 cells. Therefore, we knocked down *RBMX* expression in MDA-MB-231 cells as previously described using siRNA and harvested the RNA after 72 hours. The RNA was DNAse treated and prepared for RNA sequencing. We initially sequenced the total RNA from only 1 control and 1 *RBMX* knockdown sample using paired-end sequencing on the Illumina HiSeq 2000 platform. In our previous analysis of the RNAseq data from the Tra2 double knockdown cells it was evident that the MAJIQ pipeline can efficiently identify most alternatively spliced events in the dataset. Therefore, we also analysed the RNAseq data from the *RBMX* knockdown samples using the MAJIQ pipeline. The bioinformatics analysis was performed by Matthew Gazzara at the University of Pennsylvania comparing 1 control sample vs 1 knockdown sample. The analysis identified 596 unique local splicing variations (LSV) at a 20% dPSI minimum cut off from 505 different genes regulated by RBMX.

Given the success of MAJIQ at identifying a large subset of potential alternative splice targets regulated by RBMX in a 1-to-1 comparison, we decided to perform the RNA sequencing on three biological replicates (3 control vs 3 *RBMX* knockdown). The increased sample number made it increasingly easier to confidently determine if alternative events identified by MAJIQ are indeed being affected by the *RBMX* knockdown and are not RNAseq artefacts.

4.4.4 *RBMX* auto-regulates its expression via alternative splicing of the 3' UTR

4.4.4.1 RBMX knockdown causes a switch in alternative splicing of the RBMX isoforms

Tra2β has been reported to overlook the expression of itself and Tra2α by directly binding to poison exons within *TRA2A/B* that introduce premature termination codons (S. Grellscheid et al., 2011). Therefore, we reasoned that a similar autoregulatory mechanism may be in place in the *RBMX* gene. Unlike *TRA2A* or *TRA2B*, *RBMX* does not have a known poison exon that is alternatively spliced. However, there are at least 4 known alternative isoforms of the *RBMX* gene, one of which contains a splice junction within the 3'UTR. The presence of a splice junction means that an exon-exon junction complex (EJC) will be present in the mRNA downstream of the termination codon, potentially subjecting this isoform to nonsense mediated decay. To test this hypothesis, we examined our triplicate RNAseq data from the *RBMX* knockdown samples. A Sashimi plot of the RNAseq data for

the *RBMX* gene created using the IGV browser clearly shows the splice junction reads mapping to the spliced 3'UTR isoform was present only in the control samples (Figure 65). The total gene expression of *RBMX* has decreased more than 4-fold in the siRNA treated cells. The alternative isoform containing the splice junction within the 3'UTR is absent from the knockdown samples. This result suggests that in an attempt to sustain the expression of the functional *RBMX* isoform, the alternative splicing of *RBMX* has switched to the production of the more stable protein coding mRNA isoform to compensate for the knockdown. To validate these results by PCR we designed primer combinations that interrogate the expression of each known *RBMX* isoform. RNA extracted from *RBMX* knockdown MDA-MB-231 cells and control cells was tested by RT-PCR to compare the relative expression of four potential *RBMX* isoforms. Capillary gel electrophoresis shows the expression of the T1 isoform (corresponding to the isoform containing a splice junction within the 3'UTR) was reduced drastically with respect to all other three isoforms (Figure 64). This data confirms the results from RNAseq, that splicing switches to reduce levels of the potentially NMD degraded mRNA after *RBMX* depletion. These data are thus consistent with the existence of a concentration dependent *RBMX* autoregulatory feedback pathway.

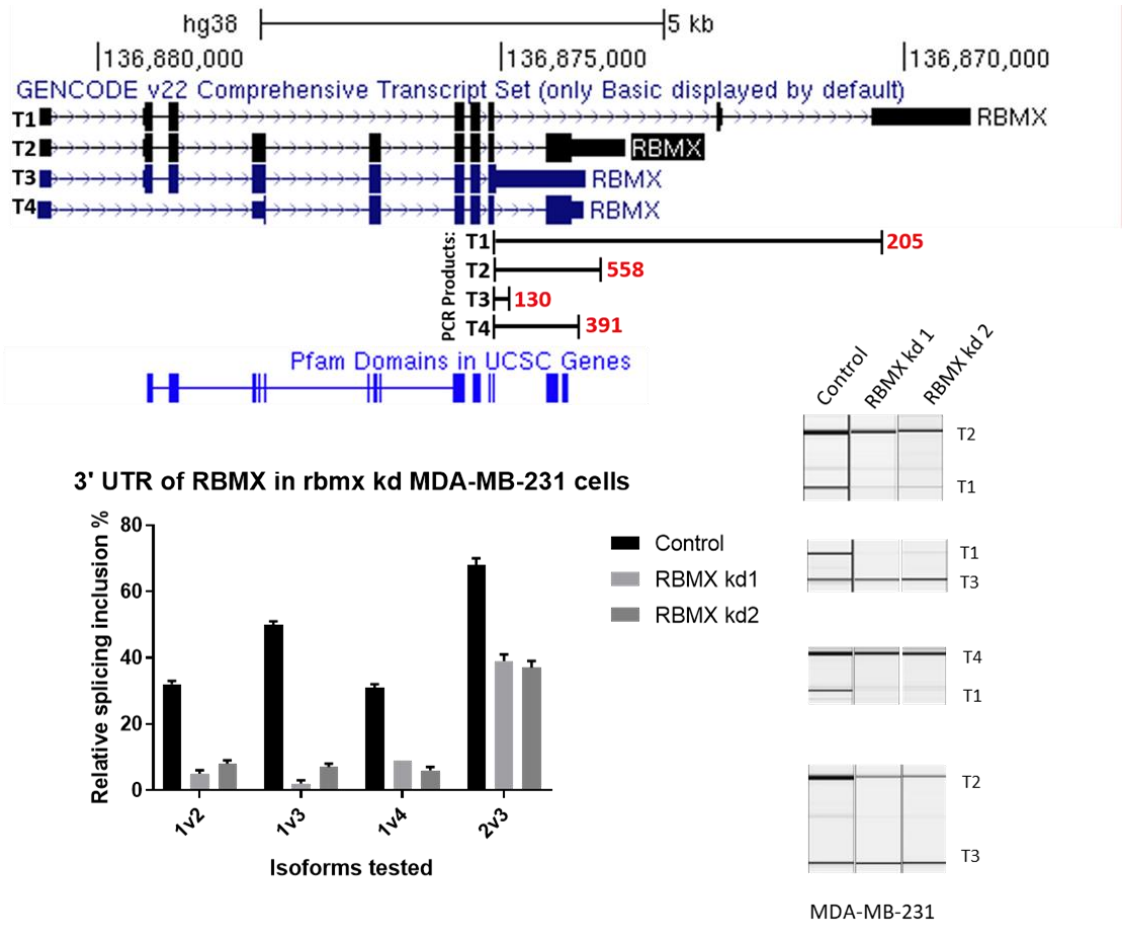

Figure 64 Autoregulation of *RBMX* expression. UCSC transcripts (top) tested using different primer combinations to assess their expression by RT-PCR. CGE image representative of three biological replicates. Bars represent standard error.

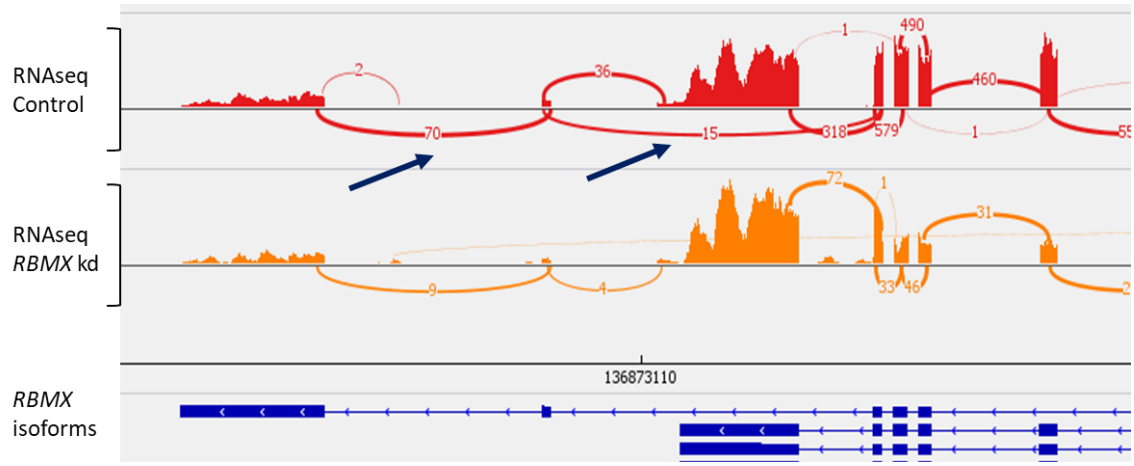

Figure 65 Sashimi plot of *RBMX* 3'UTR region. Sashimi plot of RNAseq data using IGV clearly indicates a drastic drop in expression of 'T1' isoform of *RBMX*. Splice junction reads mapping to the T1 isoform pointed out by blue arrows.

4.4.4.2 UPF1 knockdown allows accumulation of the NMD-targeted RBMX isoform

The above data suggest that the human *RBMX* gene can switch its own splice pattern in order to maximise the expression of the functional isoform if necessary. Under this hypothesis the *RBMX* isoform containing an exon junction within the 3'UTR is degraded by nonsense mediated decay (NMD). To test this, we knocked down the expression of *UPF1*, a crucial RNA helicase required for NMD, in HEK293T cells for 24 hours (knockdown experiment performed by Dr Ingrid Ehrmann, Newcastle University). We extracted the RNA and performed RT-PCR using the same primer combinations as previously interrogating the 3'UTR isoforms of *RBMX*. Capillary gel electrophoresis showed the relative abundance of isoform T1 significantly increases with respect to all three other isoforms in the absence of Upf1 (Figure 66). The increased stability of the T1 isoform after *UPF1* depletion is consistent with this isoform normally being targeted by NMD.

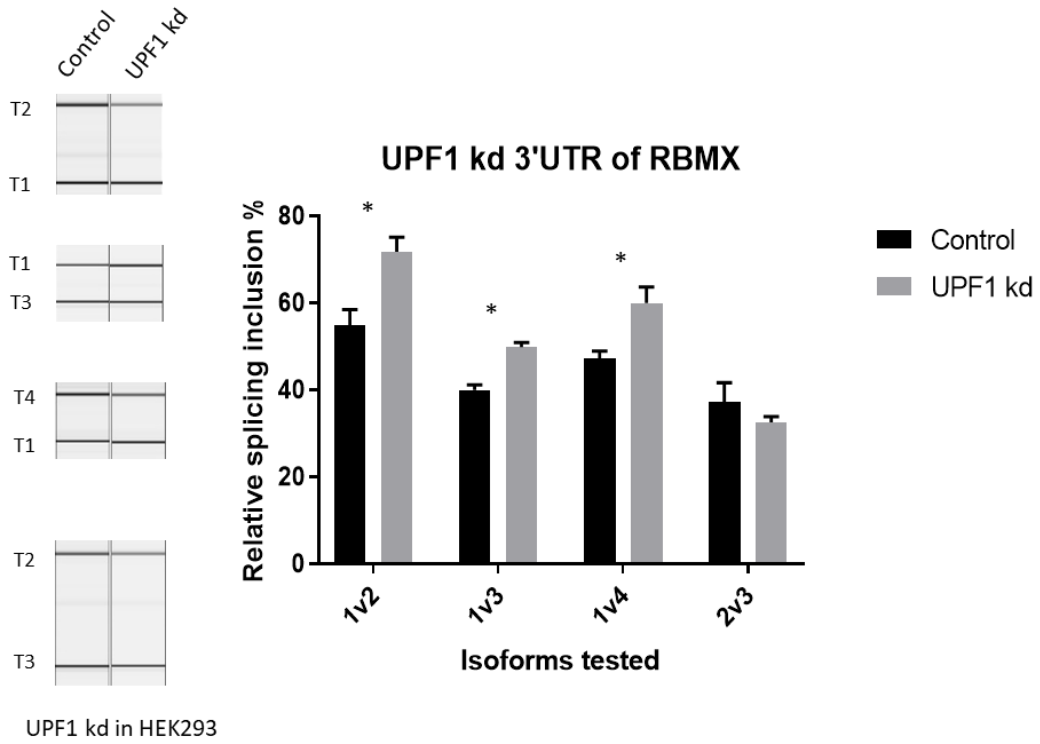

Figure 666 UPF1 knockdown HEK293T cells show increased expression of the T1 isoform of *RBMX* relative to the other isoforms. No significant difference was measured in the relative expression of T2 vs T3 isoforms. Capillary gel electrophoresis (left) and corresponding quantification. * indicates $p<0.05$

4.4.4.3 Autoregulation of the mouse Rbmx may be conserved

To see if the autoregulation of *RBMX* is conserved in mice we tested the alternative splicing of *Rbmx* (the mouse *RBMX* homologue) in 3T3 mouse fibroblasts. Cells were treated for 72 hours with two different siRNAs against *Rbmx* and one control. RNA was extracted, DNAse treated and analysed by RT-PCR to evaluate the splice changes in the *Rbmx* gene, using specific primers that were able to detect the respective isoforms. Capillary gel electrophoresis showed a significant drop in the relative expression of the T1 isoform (containing the splice junction) compared to the other isoforms. However, the presence of non-specific PCR bands and detection of just a small dPSI prompted us to continue the investigation. We determined the efficiency of the knockdown using quantitative-PCR and used the samples with the most efficient knockdown for RNA sequencing. However, the RNAseq data and Sashimi plots of the *Rbmx* gene show no visible change to the alternative splicing of the gene. Moreover, the expression of *Rbmx* is reduced only by approximately 0.5-fold by the siRNA treatment (low compared to the 4-fold expression change in the human breast cancer cells) (Figure 67). This could be the reason why there is no clear difference in the splice pattern of *Rbmx* in 3T3 mouse cells after *Rbmx* knockdown.

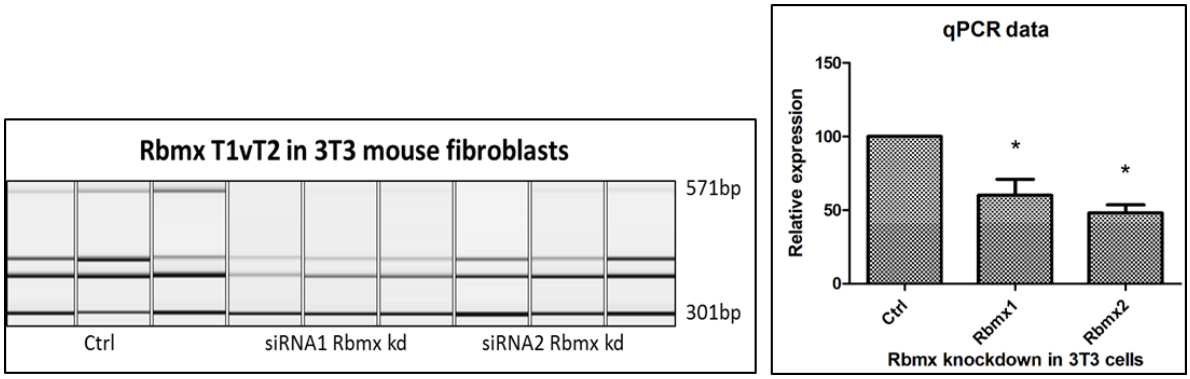

Figure 677 Rbmx knockdown in mouse 3T3 fibroblast cells shows a switch in the expected T1-like isoform at 571 bp, but two intermediary bands that were not expected to amplify indicate a potentially more complex splice junction at the 3'UTR of the mouse *Rbmx.* Right panel shows qPCR quantification of the knockdown, * indicates $p<0.05$

4.4.4.4 RBMX and RBMXL2 overexpression both repress the expression of endogenous functional isoform of RBMX

To investigate further the implications of gene expression changes for *RBMX* in the regulation of its own splicing we used an *RBMX* overexpressing HEK293T cell line previously established by Dr Mahsa Kheirollahi-Kouhestani in the group. *RBMX* overexpression was induced using tetracycline alongside cells containing the empty vector pcDNA, and with and without tetracycline used as controls. Capillary gel electrophoresis showed a clear downregulation of the *RBMX* isoforms 2 and 4 relative to isoform 1 containing the 3'UTR splice junction. The *RBMXL2* (*RBMX*-Like-2) gene codes for hnRNP-GT, a member protein of the *RBMX* family thought to have arisen from retrotransposition of *RBMX.* As such the *RBMXL2* gene contains no introns and shares 73% sequence homology with *RBMX* at the protein level. *RBMXL2* is an autosomal gene expressed only during and immediately after meiosis, but not in the preceding spermatogonial cells of male mice (Ehrmann et al., 2008). Interestingly, HEK293T cells overexpressing *RBMXL2* show clear downregulation of the functional isoforms of *RBMX* through capillary gel electrophoresis, suggesting *RBMXL2* can also regulate the alternative splice pattern of *RBMX* (Figure 68).

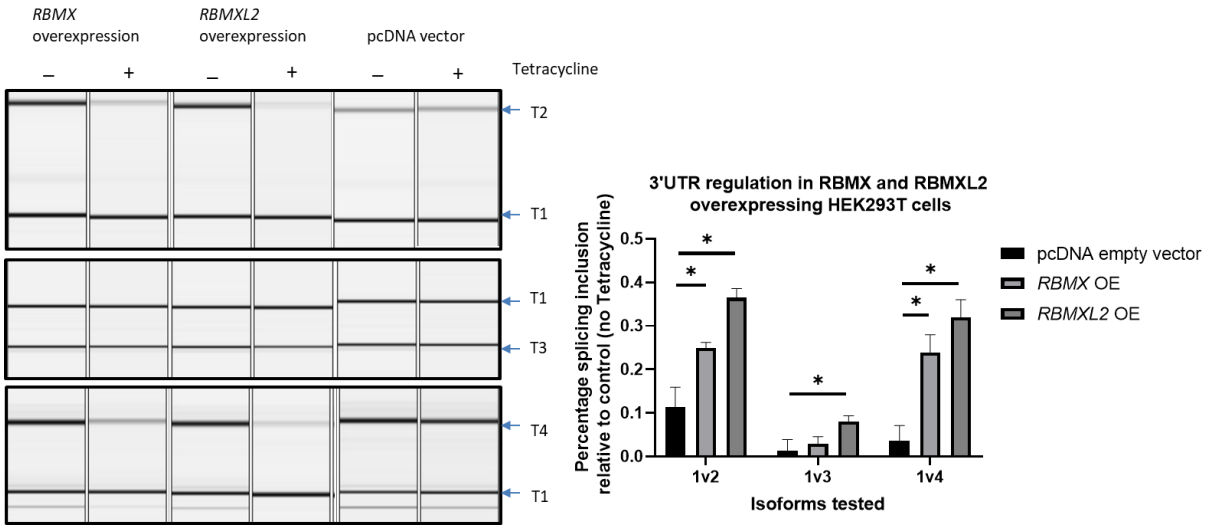

Figure 688 Tetracycline induced overexpression of *RBMX* and *RBMXL2* in HEK293T cells shows an increase of the relative abundance of isoform 'T1' of *RBMX.* CGE image representative of three biological replicates. Student's *t*-test, *p<0.05

4.4.5 Investigating alternatively spliced targets in the *RBMX* knockdown MDA-MB-231 cells

4.4.5.1 Only 51% of targets selected through MAJIQ are validated by RT-PCR

The 1-to-1 comparison of *RBMX* knockdown to control RNAseq data using MAJIQ identified 596 unique LSVs above the minimum 20% dPSI. However, the results obtained from the 1-to-1 analysis may contain false positive splice changes. Such false positives would be difficult to spot due to the absence of biological repeats. Therefore, many LSVs were manually inspected using the RNAseq data from the 3-to-3 knockdown experiment visualised on the UCSC browser. Subsequently, 70 alternative events were selected to be tested by RT-PCR as the most likely true positive candidate targets in the panel. Primers designed to differentiate between each isoform were used to test the splicing profiles of all 70 target genes by RT-PCR and capillary gel electrophoresis. 36 of the tested genes (51%) show alternative splice changes upon *RBMX* knockdown in MDA-MB-231 cells (Appendix table 3). The validation rate of *RBMX* regulated genes is considerably lower than the Tra2 protein-regulated genes in Chapter 2, clearly indicating the need of a 3-to-3 comparison in the bioinformatics analysis.

4.4.5.2 Majority of RBMX regulated targets are cassette exons but also include a substantial proportion of alternative polyadenylation sites

Heterogeneous nuclear ribonucleoproteins such as RBMX generally bind to intronic splicing silencer (ISS) sequences and mask the splice sites of downstream exons causing skipping of such exons. However, this is a general rule for hnRNPs and in our dataset we observed a variety of alternative events occurring upon *RBMX* knockdown. The majority of alternative events regulated by RBMX are

cassette exons comprising nearly 60% of the total events. Interestingly, more than 20% of the target events regulated by *RBMX* are alternative polyadenylation (11 genes) sites or alternative promoter/start exons (four genes). While alternative polyadenylation often only affects the length of the 3'UTR, a few target genes lose a substantial number of coding exons upon *RBMX* knockdown. In the cases of *ASPH* and *KLHL7* the shorter mRNA isoforms are predicted to code for much shorter protein isoforms that are missing important protein domains in their C-terminal (Figure 69). The short isoform of *ASPH* is missing the TPR_16 domain, a tetratricopeptide repeat predominantly found in transglutaminase enzymes, and Aspartyl/Asparaginyl β-hydroxylase domain which is crucial for the catalysis of oxidative reactions in a range of metabolic processes. Interestingly, the functions of both isoforms of *ASPH* have been previously investigated and it was shown the shorter isoform plays a role in Ca2+ sensing (Srikanth et al., 2012) while the full length isoform is involved in hydroxylation of Asp and Asn residues of epidermal growth factor-like proteins (Dinchuk et al., 2002). *KLHL7* encodes a gene recognised to be involved in ubiquitination. The short isoform of *KLHL7* is missing four Kelch motifs which form a β-propeller tertiary structure that is often required for interaction with Cullin 3 ubiquitin ligases (Kigoshi, Tsuruta, & Chiba, 2011). The role of *RBMX* in alternative promoter selection or transcription has not been previously described. The alternative promoter of *TCEA2* leads to the expression of a longer mRNA isoform upon *RBMX* knockdown. Interestingly, the longer isoform introduces a second Med26 protein domain at the N-terminus which is often found in transcription initiation or elongation factors (Soutourina, 2018). The MAJIQ analysis of RNAseq data from the Tra2-double knockdown in MDA-MB-231 cells showed approximately 6% of the target exons were repressed by Tra2 proteins. The role of *RBMX* in exon repression is more persistent as data shows RBMX represses the splicing of more than 13 of its target exons (28% of validated targets) (Figure 70).

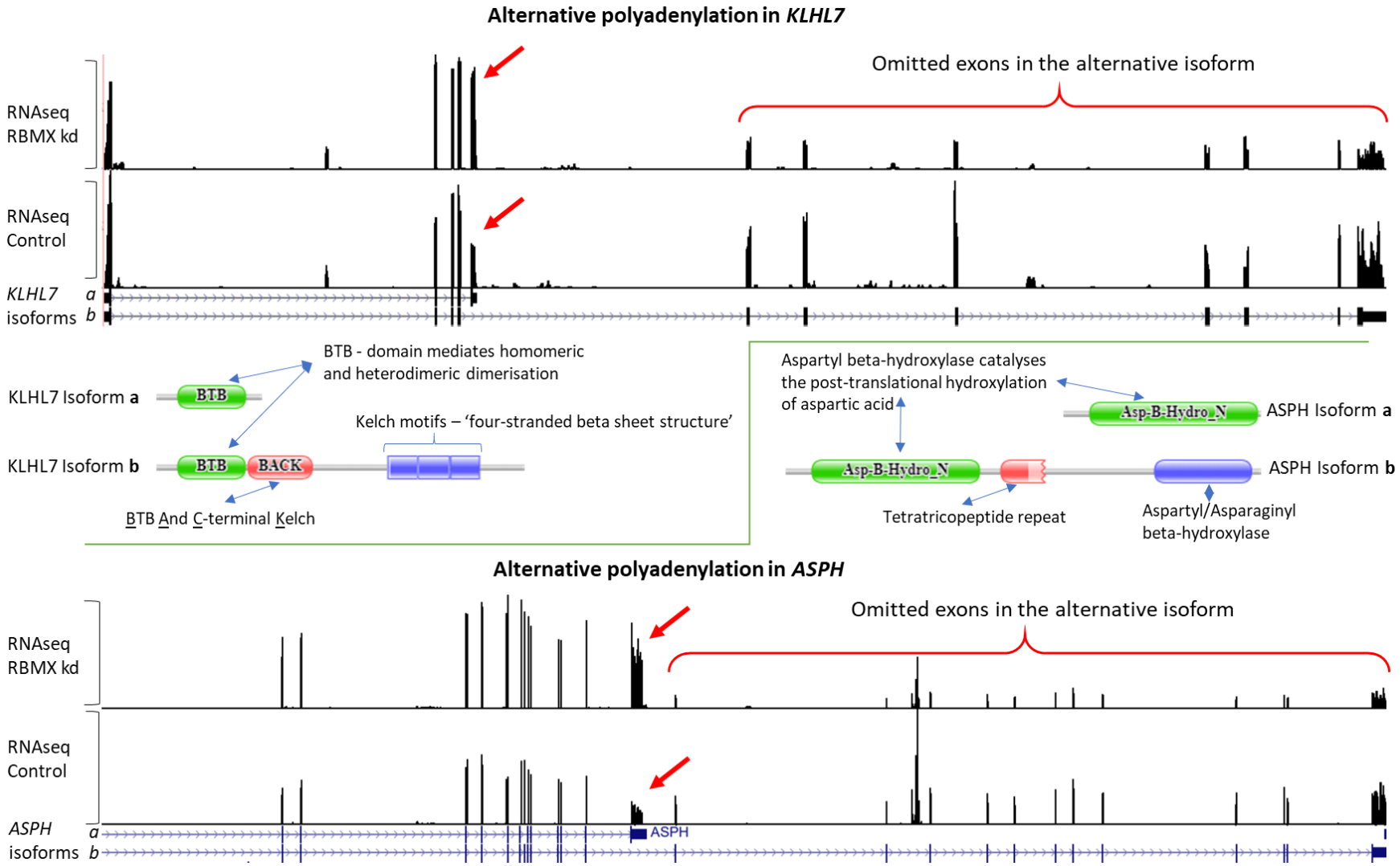

Figure 699 Alternative polyadenylation events in KLHL7 and ASPH. UCSC genome browser views show clear isoform switch which results in the likely expression of two truncated proteins missing essential protein domains as illustrated by the Pfam domain prediction data

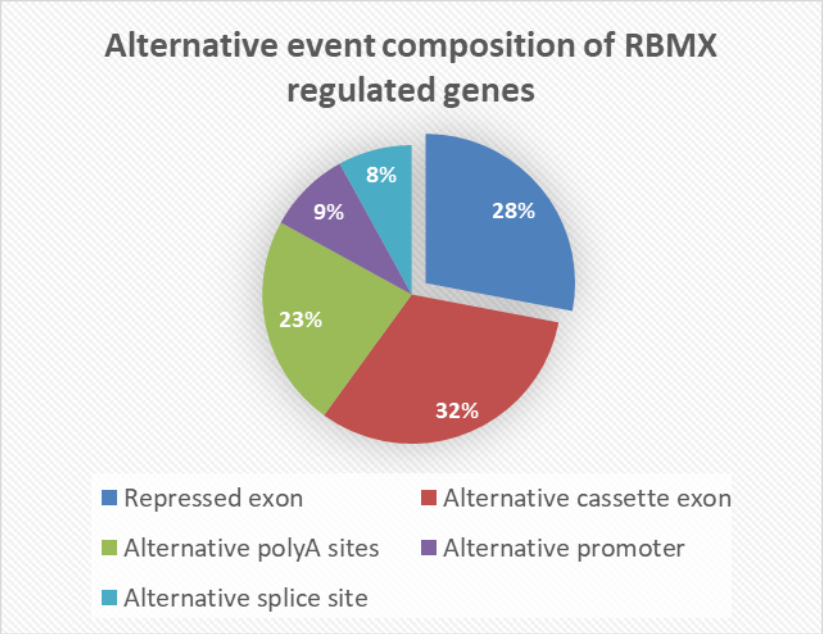

Figure 7070 Pie chart of alternative events regulated by RBMX. Unsurprisingly, RBMX acts as a repressor relatively more often than Tra2 proteins. The diagram represents proportions of different alternative event types from the 36 validated splice events identified by the 1-to-1 comparison.

4.4.6 Reciprocal splice changes in six target genes indicate direct regulation by *RBMX*

Unlike Tra2 proteins, *RBMX* has not been reported to be overexpressed in cancers relative to normal tissue. Therefore, we tested the effect of *RBMX* overexpression on the splice variations identified by RNAseq. During her PhD, Dr Mahsa Kheirollahi-Kouhestani established a stable HEK293T cell line overexpressing *RBMX* which we used to test the splicing profiles of 16 target genes. Of these, 6 alternative events show statistically significant changes in PSI in the *RBMX* overexpressing cells vs pcDNA empty vector/control cells. *ETAA1* codes for a protein kinase that binds to *ATR* and is accumulated at DNA damage sites (Figure 71). Upon *RBMX* knockdown exon 5 of *ETAA1* uses an alternative 3' splice site which leads to the expression of an mRNA isoform encoding for a shorter protein (724aa shorter) that is missing a large portion of the Pfam predicted Ewing's tumour-associated antigen motif. Only exon 10 of *STX3* gene showed a PSI change in the *RBMX*-overexpressing cells that is in the same direction as in the *RBMX* knockdown cells. In the *ATRX* gene the knockdown of *RBMX* causes the formation of an exitron – a sequence within the exon is spliced out as an intron – however there was no measurable change to the exitron splicing in overexpressing cells. Although we do not have iCLIP data for the binding sites of RBMX, the reciprocal splice changes in the genes above indicate that *RBMX* is directly regulating their splice patterns.

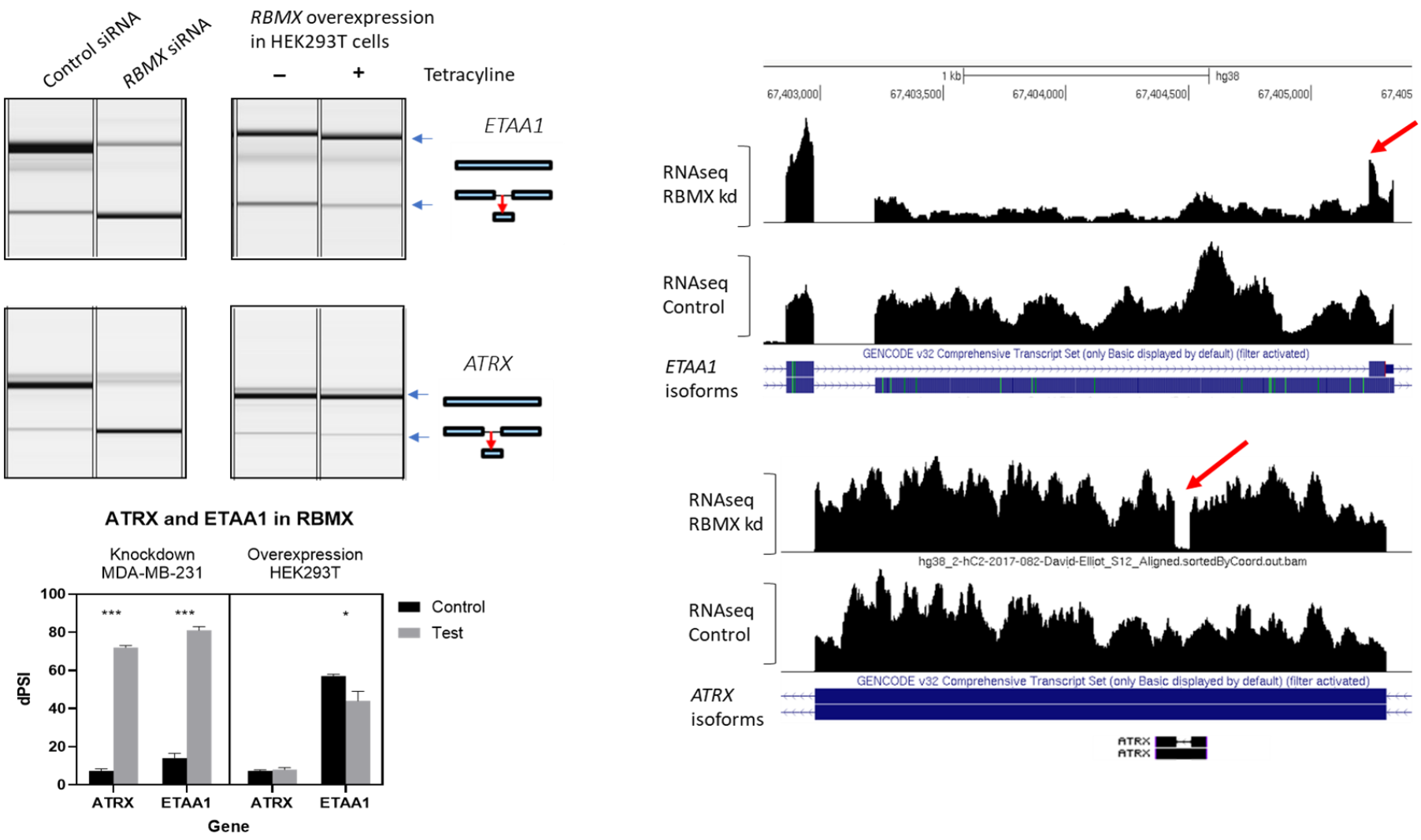

Figure 7171 The role of *RBMX* in regulating splicing of *ETAA1* and *ATRX*. Capillary gel electrophoresis of knockdown and overexpressing cells showing reciprocal switch in *ETAA1*, but not *ATRX*. Right panel shows UCSC browser view of RNAseq data for each gene clearly indicating (red arrows) read coverage drop in *RBMX* knockdown cells. * p<0.05

4.4.7 WST-1 assay indicates the *RBMX* knockdown does not have detrimental effects on cell viability

To assess the role of RBMX in the viability of the cells we used WST-1 based colorimetric assay. We measured the confluency of the cells every 24 hours after 2 days of siRNA transfection. The RBMX knockdown efficiency was estimated based on the strength of splice switch within the 3'UTR of RBMX. Unlike Tra2 protein double knockdown, the RBMX knockdown does not dramatically affect the viability of the cells. However, the confluency of the cells is initially reduced but the cells continue to grow after day 2 (Figure 72). This indicates that RBMX is not an essential gene for cell survivability and that the overall changes in alternative splicing are not detrimental to the cell.

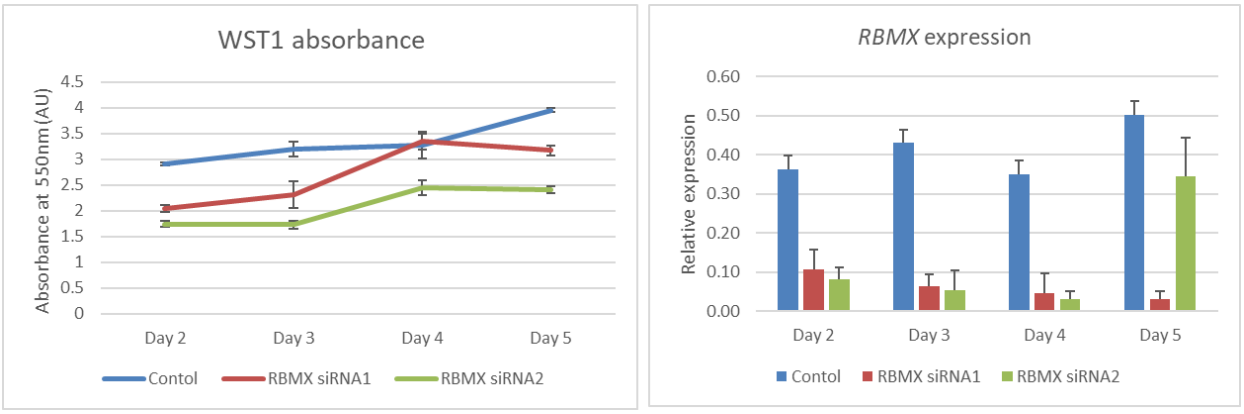

Figure 72 WST1 assay assessing MDA-MB-231 cell survivability upon *RBMX* knockdown. Relative expression of *RBMX* was measured from RNA of corresponding wells using 3'UTR splice junction primers. Bars show standard error calculated from three biological replicates.

4.5 Discussion

4.5.1 RNAseq analysis of single *RBMX* knockdown data identifies too many false positive splice events

We initially sequenced a single biological replicate of the *RBMX* knockdown in MDA-MB-231 cells against negative control siRNA treated cells. Analysis of the RNAseq data using MAJIQ identified 596 local splicing variations indicating the *RBMX* knockdown has a strong impact on alternative splicing. However, the validation rate of the splice events (51%) was considerably lower than that of Tra2-regulated genes (over 88%). The primary reason for the low validation rate is the use of a single biological replicate for RNAseq. Previous reports investigating differential gene expression changes detectable from a simple control vs treated experiment show that three biological replicates can detect between 20-40% of differentially expressed changes without setting a minimum threshold for fold-change (Schurch et al., 2016). Recently developed and improved bioinformatics packages such as SUPPA2 exploit the variability between biological replicates to determine the uncertainty in the PSI values (Juan L. Trincado et al., 2018). Although we repeated the RNAseq experiment for the *RBMX* knockdown in triplicates, the data was not subsequently analysed for differential splicing during this project. Following our conclusions from chapter 2, the RNAseq data from the triplicate *RBMX* knockdown should be analysed using at least three bioinformatics pipelines. Therefore, the identified and validated RBMX-regulated alternative events in this chapter do not represent the full spectrum of genes regulated by RBMX. Hence, Gene Ontology enrichment analysis of this data set would be biased and misleading but robust analysis of the RNAseq data from three biological repeats is essential for the continuation of this project.

4.5.2 Joint regulation of alternative events by RBMX and Tra2 proteins show predominantly synergistic regulation

Previous reports into the regulation of alternative splicing by RBMX have suggested an interaction between RBMX and Tra2β is vital to the inclusion of *SMN2* exon 7, rather than the non-specific interaction of RBMX with the pre-mRNA (Hofmann & Wirth, 2002). Contrary to this conclusion, here we find that exon 7 inclusion of *SMN2* is repressed by RBMX but activated by Tra2 proteins. However, of 13 alternative splice events regulated by both RBMX and Tra2 proteins, 11 show differential splice changes in the same directions suggesting RBMX and Tra2 proteins work synergistically to regulate splicing. This supports previous findings showing protein-protein interactions of RBMX and Tra2β and co-localisation in human spermatocytes (J. P. Venables et al., 2000). However, the identification of antagonism in exon 7 of *SMN2* remains puzzling as this

alternative event has been previously confirmed by several different reports (Cléry et al., 2011; Hofmann & Wirth, 2002; Moursy et al., 2014). Interestingly, analysis by MAJIQ (although in the 1-to-1 comparison) of the *RBMX* knockdown did not identify exon 7 of *SMN2* as differentially spliced, nor was the event identified in the Tra2 double knockdown cells. This is likely due to the low dPSI of the alternative event. The reason why an antagonistic regulation is observed in the splicing of *SMN2* could be due to the low dPSI and potentially to the diversity of other RNA binding proteins expressed in the MDA-MB-231 cells that might affect the interactions of RBMX and Tra2 proteins with RNA. While the splicing of exon 7 in *SMN2* is clinically important in relation to SMA patients, the role of RBMX and Tra2 proteins in alternative splicing of other genes is significantly stronger.

4.5.3 RBMX regulates alternative splicing independently of Tra2 proteins

Due to previous evidence of RBMX interaction with Tra2β, regulation of alternative splicing by RBMX has often been studied in conjunction with other splice regulators. However, in this chapter we validated a total of 47 alternative events, 36 of which are only regulated by RBMX suggesting interaction with Tra2 proteins is not necessary for alternative splice regulation. The other validated 11 target events are co-regulated by RBMX and Tra2β, six of these were tested prior to RNAseq analysis and shown in Figure 63. Mechanisms of alternative events regulated by RBMX involve some unique splice junctions such as the exitron in *ATRX* and 13 alternative polyadenylation events. By contrast, Tra2-regulated alternative splice events are mainly alternative cassette exons. The classical model of alternative splicing regulation by hnRNPs suggests binding to exonic or intronic splicing silencer *cis*-acting elements to suppress nearby exon recognition by the spliceosome (Busch & Hertel, 2012). The high number of alternative polyadenylation/last exon events suggest RBMX is potentially acting by masking the recognition of some of these exons which are only spliced when *RBMX* is knocked down. The same principle may also apply for the exitron in *ATRX.* Spanning only 87 nucleotides (29aa) within exon 9 of *ATRX,* the splice junctions flanking the exitron are only utilised in the *RBMX* knockdown cells, suggesting RBMX represses the recognition of these splice sites. A major drawback of the current study is the absence of a clear indication of RBMX binding. iCLIP data for RBMX in MDA-MB-231 cells would provide clear evidence of regulation by RBMX and help identify binding sites within the regulated genes.

4.5.4 Autoregulatory feedback mechanism of *RBMX* relies on an alternative polyadenylation signal

In this chapter we discuss the effect of *RBMX* knockdown on the expression of its own alternative isoforms. Although not initially identified as differentially spliced by MAJIQ, we reasoned that a

potential autoregulatory mechanism, like that between Tra2β and Tra2α (S. Grellscheid et al., 2011), might be in place. Similarly, previous reports have shown *RBM10* negatively regulates its own mRNA and protein expression via alternative skipping of exon 6 creating an out of frame isoform targeted by NMD (Sun et al., 2017). Furthermore, *RBM10* was shown to promote skipping of exon 6 or 16 of its paralog *RBM5*. The known isoform of *RBMX* (UCSC Id: NM_001164803) that contains a splice junction within the 3' UTR is a potential target for NMD. Primers testing the expression of four different mRNA isoforms of *RBMX* different at the 3' end showed a significant decrease in the relative expression of the NMD-targeted isoform (T1) (Figure 64). Furthermore, the same relative decrease was detected using two different siRNAs against *RBMX*, suggesting this effect is not due to biased isoform targeting by the siRNA. To strengthen the argument that the T1 isoform is regulated by and dependent on *RBMX* expression levels, we showed that *RBMX* overexpressing HEK293T cells also influence the relative expression of T1. Capillary gel electrophoresis indicates that induced *RBMX* overexpression causes a reduction in the relative expression of the endogenous isoforms T2, T3 and T4 rather than increase the expression of T1. Interestingly, we find the same results replicated when overexpressing the paralog *RBMXL2*, suggesting expression of *RBMX* may be under the regulation of hnRNP-GT (*RBMXL2)*. Ideally, we would test the effect of *RBMXL2* knockdown on *RBMX* splice pattern changes, but the germline/meiosis specific expression of *RBMXL2* restrict us from doing so. *UPF1* knockdown in HEK293T cells shows a relative increase in the expression of the T1 isoform suggesting this isoform is targeted by NMD. When comparing the relative expression changes between T2 and T3, no significant difference was measured suggesting these isoforms are not affected by *UPF1* knockdown. Furthermore, I searched through online publicly available differential gene expression data from a previous report of *UPF1/XRN1* knockdown in HEK293T cells (Lykke-Andersen et al., 2014) and found that the expression of the corresponding T1 isoform increases 3-fold, although this increase was marginally non-significant (p=0.08). In line with our observations above into the alternative events regulated by *RBMX*, the autoregulatory mechanism within *RBMX* is another example of alternative poly(A) site selection. However, we must note that the T1 isoform is not predicted to be targeted by NMD according to the UCSC or NCBI gene predictions. In fact, the splice junction within the 3'UTR is only 27 nucleotides downstream from the predicted stop codon, which does not conform with the general rule suggesting splice junctions must be more than 55 nucleotides downstream in order for NMD to be triggered (Nagy & Maquat, 1998; J. Zhang et al., 1998). Therefore, the isoform changes measured in this chapter should be considered with caution and must be backed with further evidence. Although the data from mouse *Rbmx* knockdown 3T3 fibroblasts is not definitive, improvements in knockdown efficiency and a better understanding of the mouse *Rbmx* splice patterns could potentially show similar results as in MDA-

MB-231 cells and would reinforce the evidence shown here for autoregulation via alternative polyadenylation.

4.6 Chapter summary

In this chapter, I investigated the effects of *RBMX* knockdown in regulating alternative splicing in MDA-MB-231 cells. I tested the splice changes in several genes previously identified as Tra2-regulated targets and detected several alternative events differentially regulated in the same direction in the *RBMX* knockdown cells. Interestingly, *RBMX* regulates a proportionally large number of alternative polyadenylation events suggestive of an inhibitory mechanism in splicing regulation. Similarly, an alternative polyadenylation event at the 3' end of *RBMX*'s own mRNA is under the regulation of *RBMX* and *RBMXL2*. The alternative isoform of *RBMX* is potentially targeted by nonsense mediated decay and may be used as a mechanism for autoregulation total RBMX expression.

Chapter 5: Concluding remarks and future work

The vast majority of human protein-coding multiexon genes are under the regulation of alternative splicing which greatly increases the coding capacity of the human genome. Classically, gene expression changes and somatic mutations have been the primary identifier of cancer hallmarks (Hanahan & Weinberg, 2011). However, alternative splicing has a much stronger impact on proteome diversification and as such has been increasingly linked to the central cancer hallmarks (Oltean & Bates, 2014). More recently RNA binding proteins involved in regulation of alternative splicing have emerged as a new class of oncoproteins and tumour suppressors (Urbanski, Leclair, & Anczukow, 2018). Exploring the orchestration of alternative splicing by various RNA binding proteins (RBPs) can help understand their role in cancer hallmarks. This thesis largely focuses on the specific splice changes regulated by Tra2α and Tra2β, but also includes the first global analysis of alternative splicing events regulated by RBMX.

In this thesis, I present results from three different bioinformatics pipelines analysing the same RNAseq data from Tra2α/β double knockdown breast cancer cells previously reported by Best *et al*. 2014 (Andrew Best et al., 2014). The results clearly indicate the necessity of using more than one bioinformatics pipeline to analyse alternative splice changes. Out of a total of 556 target genes identified to be alternatively spliced in Tra2 double knockdown cells, only 5 alternative events are detected by all three bioinformatics analyses (DexSeq, MISO and MAJIQ). This is particularly concerning when considering the relatively strong splice changes that are measured in many target genes (72 alternative events show dPSI greater than 40%). However, the high validation rates of differentially spliced events identified by both MISO and MAJIQ is indicative that these packages use very strict selection criteria in order to eliminate false positives. The added benefit of the new analysis methods revealed a range of alternative events that suggest Tra2 proteins regulate complex splicing events other than cassette exons. Unexpectedly, Tra2 proteins act by repressing splicing in important genes such as *DROSHA* and regulate intron retention in the mRNAs of genes such as *POLR3C.* These complex alternative events suggest different mechanistic regulation of splicing by Tra2 proteins compared to the classical model of ESE binding that leads to exon activation. Several bioinformatics packages, other than the ones used in this study, are available for the analysis of alternative splicing, such as SUPPA2 (J. L. Trincado et al., 2018), Psichomics (Saraiva-Agostinho & Barbosa-Morais, 2019), rMATS (Shen et al., 2014), LeafCutter (Y. I. Li et al., 2018). Future studies must consider analysing their data using more than one package to effectively identify differential splice changes.

Consequently, I investigated the protein expression of several Tra2-regulated genes from GO enriched pathways showing splice changes that were expected to impact protein structure or expression. However, protein expression of GO-enriched genes did not appear to change. Specifically, the protein expression of CENPQ, MAPKAP1 (Sin1) and SUV39H2 did not show significant isoform changes or reductions in line with the mRNA differential splicing measured. A potential explanation for this could be that these proteins may have long protein turnover rates. Turnover studies in mice found the average half-lives of proteins in the brain and blood to be between 3.5 and 9 days (Price, Guan, Burlingame, Prusiner, & Ghaemmaghami, 2010), whereas dividing mammalian cells show protein turnover rates up to a maximum of 35hr (Cambridge et al., 2011). Although different cell conditions affect the metabolic rate and overall protein turnover, some proteins will inevitably have shorter half-lives that closely relate to their individual role in the cell. In line with this idea, CHK1 is known to have a central role in cell cycle regulation. While phosphorylation and dephosphorylation of CHK1 is the main mean of regulation of cell cycle, the importance of rapid protein turnover is paramount when considering the impact of expression fluctuations on cell cycle and survivability (Andrew Best et al., 2014). Consequently, I identified protein expression changes in several genes involved in DNA repair pathways. Using cellular fractionation and immunocytochemistry I detected the short protein isoforms of MBD4 and XPA localised in the cytoplasm of MDA-MB-231 cells treated with Tra2α/β siRNAs. These shorter protein isoforms have either lost their nuclear localisation motifs (MBD4 (Bellacosa et al., 1999; Owen et al., 2007)) or are unable to interact with other proteins that regulate their nuclear import (XPA (X. Wu et al., 2007)). Although the iCLIP data clearly indicates Tra2β binds these target mRNAs, an additional experiment to verify direct binding and regulation of the *MBD4* and *XPA* alternative mRNAs by Tra2 proteins could be set up in the future. Such experiments could involve minigene constructs harbouring point mutations that interfere with Tra2α/β binding, or RNA electrophoretic mobility shift assay (RNA EMSA) to assess binding of Tra2α/β on specific target RNA sequences.

Tra2 double knockdown by siRNA causes MDA-MB-231 cells to die. The broad range of differentially spliced events orchestrated by Tra2 proteins present in the knockdown cells poses a problem in determining the role of each individual isoform. Therefore, I utilised 2'O methyl phosphorothioate-modified antisense oligonucleotides to target specific mRNA splice junctions and induce the expression of Tra2-regulated isoforms in *BRIP1, CHEK1* and *XPA*. Surprisingly, an AO targeting exon 5 of *XPA* caused skipping of three consecutive exons, like the splice event in Tra2 double knockdown cells, suggesting Tra2 proteins may play a key role in the definition and recognition of exon 5 by the spliceosome and inclusion of exon 5 is essential for the inclusion of the upstream exons 3 and 4. Subsequently, γH2AX was accumulated in cells treated with AOs targeting *XPA* and *BRIP1*, but not

CHEK1. Although very promising, the data presented here relating to the induction of DNA damage by AOs targeting DNA damage repair genes must be considered with caution, primarily due to the strong cytotoxic effect of the negative control antisense oligonucleotide. An alternative option to the use of AOs is genome editing using the CRISPR/Cas9 system. During my PhD, I attempted to establish a cell line harbouring a mutation in the *MBD4* gene that would potentially interfere with Tra2 protein binding (data not shown). Although I could induce the mutations at the specified site, I was unable to successfully grow single cell colonies with the genetic modification. It is possible that a cell line with mutated splice sites in the target genes may not be viable due to the importance of these isoforms to the cell.

In the final part of this thesis, I describe the first global analysis of differentially spliced genes under the regulation of RBMX. Tra2β and RBMX have been described to synergistically regulate the splicing control of *SMN2* exon 7 (Cléry et al., 2011; Hofmann & Wirth, 2002; Moursy et al., 2014). Unexpectedly, the endogenous splicing of *SMN2* was found to contradict previous reports, showing opposite changes in the inclusion of exon 7 from RBMX and Tra2α/β siRNA treated MDA-MB-231 cells. However, both conditions induce only a small change in PSI which could indicate weak overall effects of either RBMX or Tra2 proteins on exon 7 splicing. A future interesting experiment might involve knocking down RBMX and Tra2 proteins at the same time to assess the effect on the *SMN2* gene, but this may prove technically difficult. The RNAseq data from the RBMX knockdown experiment was analysed using MAJIQ only, but in concordance with my previous conclusion, a second analysis should be conducted. Interestingly, the data showed a range of alternative polyadenylation sites in RBMX-siRNA treated cells suggesting RBMX acts as a splicing silencer. RBMX was also found to regulate the alternative splice pattern of its own mRNA via an alternative event involving a potentially NMD-targeted isoform switch. This is a novel autoregulatory mechanism suggested for RBMX, but other RNA binding proteins such as Tra2β and RBM10 have been shown to self-regulate splicing of NMD-targeted isoforms confirming it as a common theme among RBPs (S. Grellscheid et al., 2011; Sun et al., 2017).

Global analysis of differential splicing generates a large number of potential target genes. A standard method of clustering important and related genes is by doing Gene Ontology enrichment analysis. GO is best tailored to identify enriched biological and cellular pathways where all the gene entries are known to be affected by the condition tested. However, alternative splicing events do not always infer a difference in the coding sequence of the final mRNA. Therefore, such gene entries may skew or dilute the data. In the GO enrichment analysis of Tra2-regulated genes, we identify several enriched pathways involved in chromosome biology. Although experimentally the GO enriched genes did not show changes at the protein level, long term alternative isoform expression changes

induced by Tra2 proteins could play a key role in the cells' fate *in vivo*, potentially initiating tumorigenesis. However, the data in this thesis suggests the role of Tra2 proteins is to regulate the alternative splicing of DNA repair pathways in order to maintain their functional isoforms in check and protect the cells from DNA damage-induced apoptosis or cell cycle arrest. Furthermore, alternative splice changes can influence the impact of drug responses, making them potential therapeutic targets for drug development (Urbanski et al., 2018).

Bibliography

Adamson, B., Smogorzewska, A., Sigoillot, F. D., King, R. W., & Elledge, S. J. (2012). A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat Cell Biol, 14*(3), 318-328. doi:10.1038/ncb2426

Andreassen, P. R., & Ren, K. (2009). Fanconi anemia proteins, DNA interstrand crosslink repair pathways, and cancer therapy. *Current cancer drug targets, 9*(1), 101-117. doi:10.2174/156800909787314011

Azubel, M., Habib, N., Sperling, R., & Sperling, J. (2006). Native Spliceosomes Assemble with Pre-mRNA to Form Supraspliceosomes. *Journal of Molecular Biology, 356*(4), 955-966. doi:https://doi.org/10.1016/j.jmb.2005.11.078

Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A. T., Matsumoto, Y., . . . Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proceedings of the National Academy of Sciences, 96*(7), 3969-3974. doi:10.1073/pnas.96.7.3969

Best, A., Dagliesh, C., Ehrmann, I., Kheirollahi-Kouhestani, M., Tyson-Capper, A., & 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*, 843781. doi:10.1155/2013/843781

Best, A., James, K., Dalgliesh, C., Hong, E., Kheirolahi-Kouhestani, M., Curk, T., . . . Elliott, D. J. (2014). Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons. *Nat Commun, 5*, 1-15. doi:10.1038/ncomms5760

Best, A., James, K., Hysenaj, G., Tyson-Capper, A., & Elliott, D. J. (2016). Transformer2 proteins protect breast cancer cells from accumulating replication stress by ensuring productive splicing of checkpoint kinase 1. *Frontiers of Chemical Science and Engineering, 10*(2), 186-195. doi:10.1007/s11705-015-1540-4

Bienroth, S., Keller, W., & Wahle, E. (1993). Assembly of a processive messenger RNA polyadenylation complex. *The EMBO journal, 12*(2), 585-594. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8440247

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC413241/

Boland, C. R., & Goel, A. (2010). Microsatellite instability in colorectal cancer. *Gastroenterology, 138*(6), 2073-2087.e2073. doi:10.1053/j.gastro.2009.12.064

Bomgarden, R. D., Lupardus, P. J., Soni, D. V., Yee, M.-C., Ford, J. M., & Cimprich, K. A. (2006). Opposing effects of the UV lesion repair protein XPA and UV bypass polymerase eta on ATR checkpoint signaling. *The EMBO journal, 25*(11), 2605-2614. doi:10.1038/sj.emboj.7601123

Bubb, V. J., Curtis, L. J., Cunningham, C., Dunlop, M. G., Carothers, A. D., Morris, R. G., . . . Wyllie, A. H. (1996). Microsatellite instability and the role of hMSH2 in sporadic colorectalcancer. *Oncogene, 12*(12), 2641-2649.

Busch, A., & Hertel, K. J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip Rev RNA, 3*(1), 1-12. doi:10.1002/wrna.100

Buys, S. S., Sandbach, J. F., Gammon, A., Patel, G., Kidd, J., Brown, K. L., . . . Daly, M. B. (2017). A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. *Cancer, 123*(10), 1721-1730. doi:10.1002/cncr.30498

Cambridge, S. B., Gnad, F., Nguyen, C., Bermejo, J. L., Kruger, M., & Mann, M. (2011). Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *J Proteome Res, 10*(12), 5275-5284. doi:10.1021/pr101183k

Chandler, D. S., Qi, J., & Mattox, W. (2003). Direct repression of splicing by transformer-2. *Molecular and Cellular Biology, 23*(15), 5174-5185. doi:10.1128/mcb.23.15.5174-5185.2003

Chen, Y., & Sanchez, Y. (2004). Chk1 in the DNA damage response: conserved roles from yeasts to mammals. *DNA Repair, 3*(8), 1025-1032. doi:https://doi.org/10.1016/j.dnarep.2004.03.003

Cleaver, J. E. (1968). Defective Repair Replication of DNA in Xeroderma Pigmentosum. *Nature, 218*(5142), 652-656. doi:10.1038/218652a0

Cleaver, J. E., Lam, E. T., & Revet, I. (2009). Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. *Nat Rev Genet, 10*(11), 756-768. doi:10.1038/nrg2663

Cléry, A., Jayne, S., Benderska, N., Dominguez, C., Stamm, S., & Allain, F. H. T. (2011). Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1. *Nature structural & molecular biology, 18*(4), 443-450.

Climente-González, H., Porta-Pardo, E., Godzik, A., & Eyras, E. (2017). The Functional Impact of Alternative Splicing in Cancer. *Cell Reports, 20*(9), 2215-2226. doi:https://doi.org/10.1016/j.celrep.2017.08.012

Colgan, D. F., & Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes & development, 11*(21), 2755-2766. doi:10.1101/gad.11.21.2755

Coller, J. M., Gray, N. K., & Wickens, M. P. (1998). mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. *Genes & development, 12*(20), 3226-3235. doi:10.1101/gad.12.20.3226

Coomer, A. O., Black, F., Greystoke, A., Munkley, J., & Elliott, D. J. (2019). Alternative splicing in lung cancer. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 1862*(11), 194388. doi:https://doi.org/10.1016/j.bbagrm.2019.05.006

Cortellino, S., Turner, D., Masciullo, V., Schepis, F., Albino, D., Daniel, R., . . . Bellacosa, A. (2003). The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. *Proceedings of the National Academy of Sciences of the United States of America, 100*(25), 15071-15076. doi:10.1073/pnas.2334585100

Couch, F. J., Hart, S. N., Sharma, P., Toland, A. E., Wang, X., Miron, P., . . . Fasching, P. A. (2015). Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. *J Clin Oncol, 33*(4), 304-311. doi:10.1200/jco.2014.57.1414

Cramer, P., Srebrow, A., Kadener, S., Werbajh, S., de la Mata, M., Melen, G., . . . Kornblihtt, A. R. (2001). Coordination between transcription and pre-mRNA processing. *FEBS Letters, 498*(2), 179-182. doi:https://doi.org/10.1016/S0014-5793(01)02485-1

Daneholt, B. (2001). Assembly and transport of a premessenger RNP particle. *Proceedings of the National Academy of Sciences of the United States of America, 98*(13), 7012-7017. doi:10.1073/pnas.111145498

de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., . . . Kornblihtt, A. R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell, 12*(2), 525-532. doi:10.1016/j.molcel.2003.08.001

Desmet, F. O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., & Beroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res, 37*(9), e67. doi:10.1093/nar/gkp215

Dinchuk, J. E., Focht, R. J., Kelley, J. A., Henderson, N. L., Zolotarjova, N. I., Wynn, R., . . . Friedman, P. A. (2002). Absence of post-translational aspartyl beta-hydroxylation of epidermal growth factor domains in mice leads to developmental defects and an increased incidence of intestinal neoplasia. *The Journal of biological chemistry, 277*(15), 12970-12977. doi:10.1074/jbc.M110389200

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., . . . Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature, 489*(7414), 101-108. doi:10.1038/nature11233

Dworaczek, H., & Xiao, W. (2007). Xeroderma pigmentosum: a glimpse into nucleotide excision repair, genetic instability, and cancer. *Crit Rev Oncog, 13*(2), 159-177. doi:10.1615/critrevoncog.v13.i2.20

Easton, D. F., Lesueur, F., Decker, B., Michailidou, K., Li, J., Allen, J., . . . Chenevix-Trench, G. (2016). No evidence that protein truncating variants in BRIP1 are associated with breast cancer risk:

implications for gene panel testing. *J Med Genet, 53*(5), 298-309. doi:10.1136/jmedgenet-2015-103529

Edelbrock, M. A., Kaliyaperumal, S., & Williams, K. J. (2013). Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities. *Mutation research, 743-744*, 53-66. doi:10.1016/j.mrfmmm.2012.12.008

Ehrmann, I., Dalgliesh, C., Tsaousi, A., Paronetto, M. P., Heinrich, B., Kist, R., . . . Elliott, D. J. (2008). Haploinsufficiency of the germ cell-specific nuclear RNA binding protein hnRNP G-T prevents functional spermatogenesis in the mouse. *Human Molecular Genetics, 17*(18), 2803-2818. doi:10.1093/hmg/ddn179

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., . . . Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic acids research, 47*(D1), D427-D432. doi:10.1093/nar/gky995

Elliott, David J., Best, A., Dalgliesh, C., Ehrmann, I., & Grellscheid, S. (2012). How does Tra2β protein regulate tissue-specific RNA splicing?. 2012, *Biochem Soc Trans*, 40(4), 784-8. doi: 10.1042/BST20120036.

Elliott, D. J., Dalgliesh, C., Hysenaj, G., & Ehrmann, I. (2019). RBMX family proteins connect the fields of nuclear RNA processing, disease and sex chromosome biology. *Int J Biochem Cell Biol, 108*, 1-6. doi:10.1016/j.biocel.2018.12.014

Elliott, D. J., Dalgliesh, C., Hysenaj, G., & Ehrmann, I. (2019). RBMX family proteins connect the fields of nuclear RNA processing, disease and sex chromosome biology. *The International Journal of Biochemistry & Cell Biology, 108*, 1-6. doi:https://doi.org/10.1016/j.biocel.2018.12.014

Erkelenz, S., Mueller, W. F., Evans, M. S., Busch, A., Schoneweis, K., Hertel, K. J., & Schaal, H. (2013). Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *Rna, 19*(1), 96-102. doi:10.1261/rna.037044.112

Ferrari, Karin J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., . . . Pasini, D. (2014). Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Molecular Cell, 53*(1), 49-62. doi:https://doi.org/10.1016/j.molcel.2013.10.030

Fischer, D. C., Noack, K., Runnebaum, I. B., Watermann, D. O., Kieback, D. G., Stamm, S., & Stickeler, E. (2004a). Expression of splicing factors in human ovarian cancer. *Oncology reports, 11*(5), 1085-1090.

Fischer, D. C., Noack, K., Runnebaum, I. B., Watermann, D. O., Kieback, D. G., Stamm, S., & Stickeler, E. (2004b). Expression of splicing factors in human ovarian cancer. *Oncol Rep, 11*(5), 1085-1090.

Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I. W., & Gilmartin, G. M. (1997). Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proceedings of the National Academy of Sciences of the United States of America, 94*(22), 11893-11898. doi:10.1073/pnas.94.22.11893

Fu, X. D., & Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet, 15*(10), 689-701. doi:10.1038/nrg3778

Furuichi, Y., & Shatkin, A. J. (2000). Viral and cellular mRNA capping: past and prospects. *Adv Virus Res, 55*, 135-184. doi:10.1016/s0065-3527(00)55003-9

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

Galante, P. A. F., Sandhu, D., de Sousa Abreu, R., Gradassi, M., Slager, N., Vogel, C., . . . Penalva, L. O. F. (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), 426-433.

Galarneau, A., & Richard, S. (2005). Target RNA motif and target mRNAs of the Quaking STAR protein. *Nature structural & molecular biology, 12*(8), 691-698.

Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P. M., . . . Biamonti, G. (2005). Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Mol Cell, 20*(6), 881-890. doi:10.1016/j.molcel.2005.10.026

Glover-Cutter, K., Kim, S., Espinosa, J., & Bentley, D. L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature structural & molecular biology, 15*(1), 71-78. doi:10.1038/nsmb1352

Gonatopoulos-Pournatzis, T., & Cowling, V. H. (2014). Cap-binding complex (CBC). *The Biochemical journal, 457*(2), 231-242. doi:10.1042/BJ20131214

Graveley, B. R., Hertel, K. J., & Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *Rna, 7*(6), 806-818. doi:10.1017/s1355838201010317

Gray, F. C., Pohler, J. R., Warbrick, E., & MacNeill, S. A. (2004). Mapping and mutation of the conserved DNA polymerase interaction motif (DPIM) located in the C-terminal domain of fission yeast DNA polymerase delta subunit Cdc27. *BMC Mol Biol, 5*(1), 21. doi:10.1186/1471-2199-5-21

Gray, N. K., Coller, J. M., Dickson, K. S., & Wickens, M. (2000). Multiple portions of poly(A)-binding protein stimulate translation in vivo. *The EMBO journal, 19*(17), 4723-4733. doi:10.1093/emboj/19.17.4723

Grellscheid, S., Dalgliesh, C., Storbeck, M., Best, A., Liu, Y., Jakubik, M., . . . Elliott, D. J. (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

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

Gupta, I., Ouhtit, A., Al-Ajmi, A., Rizvi, S. G. A., Al-Riyami, H., Al-Riyami, M., & Tamimi, Y. (2018). BRIP1 overexpression is correlated with clinical features and survival outcome of luminal breast cancer subtypes. *Endocr Connect, 7*(1), 65-77. doi:10.1530/ec-17-0173

Hall, M. P., Nagel, R. J., Fagg, W. S., Shiue, L., Cline, M. S., Perriman, R. J., . . . Ares, M., Jr. (2013). Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *RNA (New York, N.Y.), 19*(5), 627-638. doi:10.1261/rna.038422.113

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell, 144*(5), 646-674. doi:10.1016/j.cell.2011.02.013

Harrison, P. M., Kumar, A., Lang, N., Snyder, M., & Gerstein, M. (2002). A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res, 30*(5), 1083-1090. doi:10.1093/nar/30.5.1083

He, C., Zhou, F., Zuo, Z., Cheng, H., & Zhou, R. (2009). A global view of cancer-specific transcript variants by subtractive transcriptome-wide analysis. *PloS one, 4*(3), e4732-e4732. doi:10.1371/journal.pone.0004732

Hendrich, B., & Bird, A. (1998). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Molecular and Cellular Biology, 18*(11), 6538-6547. doi:10.1128/mcb.18.11.6538

Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., & Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature, 401*(6750), 301-304. doi:10.1038/45843

Hofmann, Y., Lorson, C. L., Stamm, S., Androphy, E. J., & Wirth, B. (2000). 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, 97*(17), 9618-9623.

Hofmann, Y., & 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), 2037-2049.

Hogg, R., McGrail, J. C., & O'Keefe, R. T. (2010). The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochemical Society transactions, 38*(4), 1110-1115. doi:10.1042/BST0381110

Hughes, T. A. (2006). Regulation of gene expression by alternative untranslated regions. *Trends in Genetics, 22*(3), 119-122. doi:https://doi.org/10.1016/j.tig.2006.01.001

Ip, J. Y., Schmidt, D., Pan, Q., Ramani, A. K., Fraser, A. G., Odom, D. T., & Blencowe, B. J. (2011). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res, 21*(3), 390-401. doi:10.1101/gr.111070.110

Iwasaki, T., Koibuchi, N., & Chin, W. W. (2005). Synovial sarcoma translocation (SYT) encodes a nuclear receptor coactivator. *Endocrinology, 146*(9), 3892-3899. doi:10.1210/en.2004-1513

Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C. Y., . . . Fu, X. D. (2013). SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell, 153*(4), 855-868. doi:10.1016/j.cell.2013.04.028

Kanhoush, R., Beenders, B., Perrin, C., Moreau, J., Bellini, M., & Penrad-Mobayed, M. (2010). Novel domains in the hnRNP G/RBMX protein with distinct roles in RNA binding and targeting nascent transcripts. *Nucleus, 1*(1), 109-122. doi:10.4161/nucl.1.1.10857

Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., . . . Wu, J. Y. (2011). RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5′ splice site. *Molecular and Cellular Biology, 31*(9), 1812-1821.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., & Stamm, S. (2013). Function of alternative splicing. *Gene, 514*(1), 1-30. doi:https://doi.org/10.1016/j.gene.2012.07.083

Kigoshi, Y., Tsuruta, F., & Chiba, T. (2011). Ubiquitin ligase activity of Cul3-KLHL7 protein is attenuated by autosomal dominant retinitis pigmentosa causative mutation. *The Journal of biological chemistry, 286*(38), 33613-33621. doi:10.1074/jbc.M111.245126

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology, 14*(4), R36. doi:10.1186/gb-2013-14-4-r36

Kim, M.-Y., Hur, J., & Jeong, S.-J. (2009). Emerging roles of RNA and RNA-binding protein network in cancer cells. *BMB reports, 42*(3), 125-130.

Konarska, M. M., Padgett, R. A., & Sharp, P. A. (1984). Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell, 38*(3), 731-736. doi:10.1016/0092-8674(84)90268-x

Kondo, E., Gu, Z., Horii, A., & Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. *Mol Cell Biol, 25*(11), 4388-4396. doi:10.1128/mcb.25.11.4388-4396.2005

Kozak, M. (1991). An analysis of vertebrate mRNA sequences: intimations of translational control. *The Journal of Cell Biology, 115*(4), 887. doi:10.1083/jcb.115.4.887

Kraemer, K. H., Seetharam, S., Seidman, M. M., Bredberg, A., Brash, D., Waters, H. L., . . . Sanford, K. (1990). Defective DNA Repair in Humans: Clinical and Molecular Studies of Xeroderma Pigmentosum. In B. M. Sutherland & A. D. Woodhead (Eds.), *DNA Damage and Repair in Human Tissues* (pp. 95-104). Boston, MA: Springer US.

Kuusisto, K. M., Bebel, A., Vihinen, M., Schleutker, J., & Sallinen, S. L. (2011). Screening for BRCA1, BRCA2, CHEK2, PALB2, BRIP1, RAD50, and CDH1 mutations in high-risk Finnish BRCA1/2-founder mutation-negative breast and/or ovarian cancer individuals. *Breast Cancer Res, 13*(1), R20. doi:10.1186/bcr2832

Laget, S., Miotto, B., Chin, H. G., Estève, P.-O., Roberts, R. J., Pradhan, S., & Defossez, P.-A. (2014). MBD4 cooperates with DNMT1 to mediate methyl-DNA repression and protects mammalian cells from oxidative stress. *Epigenetics, 9*(4), 546-556. doi:10.4161/epi.27695

Lander, E. S. (2011). Initial impact of the sequencing of the human genome. *Nature, 470*(7333), 187-197. doi:10.1038/nature09792

Lewis, B. P., Green, R. E., & 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), 189-192. doi:10.1073/pnas.0136770100

Li, D.-Q., Pakala, S. B., Reddy, S. D. N., Ohshiro, K., Peng, S.-H., Lian, Y., . . . Kumar, R. (2010). Revelation of p53-independent function of MTA1 in DNA damage response via modulation of the p21 WAF1-proliferating cell nuclear antigen pathway. *The Journal of biological chemistry, 285*(13), 10044-10052. doi:10.1074/jbc.M109.079095

Li, L., Peterson, C. A., Lu, X., & Legerski, R. J. (1995). Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair. *Molecular and Cellular Biology, 15*(4), 1993-1998. doi:10.1128/mcb.15.4.1993

Li, Y. I., Knowles, D. A., Humphrey, J., Barbeira, A. N., Dickinson, S. P., Im, H. K., & Pritchard, J. K. (2018). Annotation-free quantification of RNA splicing using LeafCutter. *Nature genetics, 50*(1), 151-158. doi:10.1038/s41588-017-0004-9

Li, Z., Musich, P. R., Cartwright, B. M., Wang, H., & Zou, Y. (2013). UV-induced nuclear import of XPA is mediated by importin-alpha4 in an ATR-dependent manner. *PloS one, 8*(7), e68297. doi:10.1371/journal.pone.0068297

Li, Z., Musich, P. R., Serrano, M. A., Dong, Z., & Zou, Y. (2011). XPA-mediated regulation of global nucleotide excision repair by ATR Is p53-dependent and occurs primarily in S-phase. *PloS one, 6*(12), e28326-e28326. doi:10.1371/journal.pone.0028326

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

Liu, Y., Li, H., Zhang, R., Dang, H., Sun, P., Zou, L., . . . Hu, Y. (2016). Overexpression of the BRIP1 ameliorates chemosensitivity to cisplatin by inhibiting Rac1 GTPase activity in cervical carcinoma HeLa cells. *Gene, 578*(1), 85-91. doi:10.1016/j.gene.2015.12.007

Long, Jennifer C., & Caceres, Javier F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem J.*, 417(1), 15-27. doi: 10.1042/BJ20081501. (Vol. 417).

Lukas, J., Gao, D.-Q., Keshmeshian, M., Wen, W.-H., Tsao-Wei, D., Rosenberg, S., & Press, M. F. (2001). Alternative and aberrant messenger RNA splicing of the mdm2 oncogene in invasive breast cancer. *Cancer research, 61*(7), 3212-3219.

Lykke-Andersen, S., Chen, Y., Ardal, B. R., Lilje, B., Waage, J., Sandelin, A., & Jensen, T. H. (2014). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes & development, 28*(22), 2498-2517. doi:10.1101/gad.246538.114

Maquat, L. E., Hwang, J., Sato, H., & Tang, Y. (2010). CBP80-promoted mRNP rearrangements during the pioneer round of translation, nonsense-mediated mRNA decay, and thereafter. *Cold Spring Harb Symp Quant Biol, 75*, 127-134. doi:10.1101/sqb.2010.75.028

Martínez-Arribas, F., Agudo, D., Pollán, M., Gómez-Esquer, F., Díaz-Gil, G., Lucas, R., & Schneider, J. (2006). Positive correlation between the expression of X-chromosome RBM genes (RBMX, RBM3, RBM10) and the proapoptotic Bax gene in human breast cancer. *Journal of Cellular Biochemistry, 97*(6), 1275-1282. doi:10.1002/jcb.20725

Matlin, A. J., Clark, F., & Smith, C. W. (2005). Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol, 6*(5), 386-398. doi:10.1038/nrm1645

Mende, Y., Jakubik, M., Riessland, M., Schoenen, F., Roßbach, K., Kleinridders, A., . . . 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. *Human Molecular Genetics, 19*(11), 2154-2167.

Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., MacDougall, E., Hendrich, B., . . . Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science, 297*(5580), 403-405. doi:10.1126/science.1073354

Mischo, H. E., & Proudfoot, N. J. (2013). Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochimica et biophysica acta, 1829*(1), 174-185. doi:10.1016/j.bbagrm.2012.10.003

Moursy, A., Allain, F. H. T., & Cléry, A. (2014). Characterization of the RNA recognition mode of hnRNP G extends its role in SMN2 splicing regulation. *Nucleic acids research, 42*(10), 6659-6672. doi:10.1093/nar/gku244

Muniandy, P. A., Liu, J., Majumdar, A., Liu, S.-t., & Seidman, M. M. (2010). DNA interstrand crosslink repair in mammalian cells: step by step. *Critical reviews in biochemistry and molecular biology, 45*(1), 23-49. doi:10.3109/10409230903501819

Musich, P. R., Li, Z., & Zou, Y. (2017). Xeroderma Pigmentosa Group A (XPA), Nucleotide Excision Repair and Regulation by ATR in Response to Ultraviolet Irradiation. *Advances in experimental medicine and biology, 996*, 41-54. doi:10.1007/978-3-319-56017-5_4

Nagy, E., & Maquat, L. E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci, 23*(6), 198-199. doi:10.1016/s0968-0004(98)01208-0

Nasim, M. T., Chernova, T. K., Chowdhury, H. M., Yue, B.-G., & Eperon, I. C. (2003). HnRNP G and Tra2β: opposite effects on splicing matched by antagonism in RNA binding. *Human Molecular Genetics, 12*(11), 1337-1348. doi:10.1093/hmg/ddg136

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*(2), 191-202.

Nicolas, E., Golemis, E. A., & Arora, S. (2016). POLD1: Central mediator of DNA replication and repair, and implication in cancer and other pathologies. *Gene, 590*(1), 128-141. doi:10.1016/j.gene.2016.06.031

Niwa, M., Rose, S. D., & Berget, S. M. (1990). In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes & development, 4*(9), 1552-1559. doi:10.1101/gad.4.9.1552

Ogi, T., Limsirichaikul, S., Overmeer, R. M., Volker, M., Takenaka, K., Cloney, R., . . . Lehmann, A. R. (2010). Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Mol Cell, 37*(5), 714-727. doi:10.1016/j.molcel.2010.02.009

Oltean, S., & Bates, D. O. (2014). Hallmarks of alternative splicing in cancer. *Oncogene, 33*(46), 5311-5318. doi:10.1038/onc.2013.533

Ouyang, Y. Q., zur Hausen, A., Orlowska-Volk, M., Jäger, M., Bettendorf, H., Hirschfeld, M., . . . Stickeler, E. (2011). Expression levels of hnRNP G and hTra2-beta1 correlate with opposite outcomes in endometrial cancer biology. *International Journal of Cancer, 128*(9), 2010-2019. doi:10.1002/ijc.25544

Owen, R. M., Baker, R. D., Bader, S., Dunlop, M. G., & Nicholl, I. D. (2007). The identification of a novel alternatively spliced form of the MBD4 DNA glycosylase. *Oncol Rep, 17*(1), 111-116.

Pabis, M., Neufeld, N., Steiner, M. C., Bojic, T., Shav-Tal, Y., & Neugebauer, K. M. (2013). The nuclear cap-binding complex interacts with the U4/U6.U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *Rna, 19*(8), 1054-1063. doi:10.1261/rna.037069.112

Pajares, M. J., Ezponda, T., Catena, R., Calvo, A., Pio, R., & Montuenga, L. M. (2007). Alternative splicing: an emerging topic in molecular and clinical oncology. *The lancet oncology, 8*(4), 349-357. doi:https://doi.org/10.1016/S1470-2045(07)70104-3

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet, 40*(12), 1413-1415. doi:10.1038/ng.259

Pan, Y. R., & Lee, E. Y. (2009). UV-dependent interaction between Cep164 and XPA mediates localization of Cep164 at sites of DNA damage and UV sensitivity. *Cell Cycle, 8*(4), 655-664. doi:10.4161/cc.8.4.7844

Papasaikas, P., & Valcárcel, J. (2016). The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends in Biochemical Sciences, 41*(1), 33-45. doi:https://doi.org/10.1016/j.tibs.2015.11.003

Pasini, D., Bracken, A. P., Jensen, M. R., Denchi, E. L., & Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *The EMBO journal, 23*(20), 4061-4071. doi:10.1038/sj.emboj.7600402

Patel, A. A., & Steitz, J. A. (2003). Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol, 4*(12), 960-970. doi:10.1038/nrm1259

Patil, M., Pabla, N., & Dong, Z. (2013). Checkpoint kinase 1 in DNA damage response and cell cycle regulation. *Cellular and Molecular Life Sciences, 70*(21), 4009-4021. doi:10.1007/s00018-013-1307-3

Pertea, M., Shumate, A., Pertea, G., Varabyou, A., Chang, Y.-C., Madugundu, A. K., . . . Salzberg, S. L. (2018). Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise. *bioRxiv*, 332825. doi:10.1101/332825

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., & Liuni, S. (2001). Structural and functional features of eukaryotic mRNA untranslated regions. *Gene, 276*(1), 73-81. doi:https://doi.org/10.1016/S0378-1119(01)00674-6

Petronzelli, F., Riccio, A., Markham, G. D., Seeholzer, S. H., Stoerker, J., Genuardi, M., . . . Bellacosa, A. (2000). Biphasic Kinetics of the Human DNA Repair Protein MED1 (MBD4), a Mismatch-specific DNA N-Glycosylase. *Journal of Biological Chemistry, 275*(42), 32422-32429. doi:10.1074/jbc.M004535200

Pilch, D. R., Sedelnikova, O. A., Redon, C., Celeste, A., Nussenzweig, A., & Bonner, W. M. (2003). Characteristics of gamma-H2AX foci at DNA double-strand breaks sites. *Biochem Cell Biol, 81*(3), 123-129. doi:10.1139/o03-042

Pitout, I., Flynn, L. L., Wilton, S. D., & Fletcher, S. (2019). Antisense-mediated splice intervention to treat human disease: the odyssey continues. *F1000Research, 8*, F1000 Faculty Rev-1710. doi:10.12688/f1000research.18466.1

Price, J. C., Guan, S., Burlingame, A., Prusiner, S. B., & Ghaemmaghami, S. (2010). Analysis of proteome dynamics in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America, 107*(32), 14508-14513. doi:10.1073/pnas.1006551107

Proudfoot, N. (1991). Poly(A) signals. *Cell, 64*(4), 671-674. doi:10.1016/0092-8674(91)90495-k

Proudfoot, N. J., & Brownlee, G. G. (1976). 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature, 263*(5574), 211-214. doi:10.1038/263211a0

Przybyl, J., van de Rijn, M., & Rutkowski, P. (2019). Detection of SS18-SSX1/2 fusion transcripts in circulating tumor cells of patients with synovial sarcoma. *Diagn Pathol, 14*(1), 24. doi:10.1186/s13000-019-0800-x

Qi, J., Su, S., & Mattox, W. (2007). The doublesex splicing enhancer components Tra2 and Rbp1 also repress splicing through an intronic silencer. *Mol Cell Biol, 27*(2), 699-708. doi:10.1128/mcb.01572-06

Ramanathan, A., Robb, G. B., & Chan, S.-H. (2016). mRNA capping: biological functions and applications. *Nucleic acids research, 44*(16), 7511-7526. doi:10.1093/nar/gkw551

Rappsilber, J., Ryder, U., Lamond, A. I., & Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res, 12*(8), 1231-1245. doi:10.1101/gr.473902

Richardson, R. T., Alekseev, O. M., Grossman, G., Widgren, E. E., Thresher, R., Wagner, E. J., . . . Michael, G. O. (2006). Nuclear autoantigenic sperm protein (NASP), a linker histone chaperone that is required for cell proliferation. *Journal of Biological Chemistry, 281*(30), 21526-21534.

Richardson, R. T., Batova, I. N., Widgren, E. E., Zheng, L.-X., Whitfield, M., Marzluff, W. F., & Michael, G. O. (2000). Characterization of the histone H1-binding protein, NASP, as a cell cycle-regulated somatic protein. *Journal of Biological Chemistry, 275*(39), 30378-30386.

Salati, S., Zini, R., Nuzzo, S., Guglielmelli, P., Pennucci, V., Prudente, Z., . . . Manfredini, R. (2016). Integrative analysis of copy number and gene expression data suggests novel pathogenetic mechanisms in primary myelofibrosis. *Int J Cancer, 138*(7), 1657-1669. doi:10.1002/ijc.29920

Santos, J. C., Brianti, M. T., Almeida, V. R., Ortega, M. M., Fischer, W., Haas, R., . . . Ribeiro, M. L. (2017). Helicobacter pylori infection modulates the expression of miRNAs associated with DNA mismatch repair pathway. *Molecular Carcinogenesis, 56*(4), 1372-1379. doi:10.1002/mc.22590

Saraiva-Agostinho, N., & Barbosa-Morais, N. L. (2019). psichomics: graphical application for alternative splicing quantification and analysis. *Nucleic acids research, 47*(2), e7-e7. doi:10.1093/nar/gky888

Schärer, O. D. (2013). Nucleotide excision repair in eukaryotes. *Cold Spring Harbor perspectives in biology, 5*(10), a012609-a012609. doi:10.1101/cshperspect.a012609

Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., . . . Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA (New York, N.Y.), 22*(6), 839-851. doi:10.1261/rna.053959.115

Schwartz, S., Hall, E., & Ast, G. (2009). SROOGLE: webserver for integrative, user-friendly visualization of splicing signals. *Nucleic Acids Res, 37*(Web Server issue), W189-192. doi:10.1093/nar/gkp320

Sebestyén, E., Singh, B., Miñana, B., Pagès, A., Mateo, F., Pujana, M. A., . . . Eyras, E. (2016). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome research, 26*(6), 732-744. Retrieved from http://genome.cshlp.org/content/26/6/732.abstract

Sen, N., Gui, B., & Kumar, R. (2014). Role of MTA1 in cancer progression and metastasis. *Cancer metastasis reviews, 33*(4), 879-889. doi:10.1007/s10555-014-9515-3

Shen, S., Park, J. W., Lu, Z.-x., Lin, L., Henry, M. D., Wu, Y. N., . . . Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences, 111*(51), E5593-E5601. doi:10.1073/pnas.1419161111

Shimizu-Motohashi, Y., Murakami, T., Kimura, E., Komaki, H., & Watanabe, N. (2018). Exon skipping for Duchenne muscular dystrophy: a systematic review and meta-analysis. *Orphanet J Rare Dis, 13*(1), 93. doi:10.1186/s13023-018-0834-2

Shuman, S. (1997). Origins of mRNA identity: capping enzymes bind to the phosphorylated C-terminal domain of RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America, 94*(24), 12758-12760. doi:10.1073/pnas.94.24.12758

Singh, K., Wolfe, A. L., Zhong, Y., Rätsch, G., & Wendel, H.-G. (2015). Abstract B23: The 5 UTR of many oncogenes and transcription factors encodes a targetable dependence on the eIF4A RNA helicase. *Molecular Cancer Therapeutics, 14*(7 Supplement), B23. doi:10.1158/1538-8514.PI3K14-B23

Sjolund, A. B., Senejani, A. G., & Sweasy, J. B. (2013). MBD4 and TDG: multifaceted DNA glycosylases with ever expanding biological roles. *Mutation research, 743-744*, 12-25. doi:10.1016/j.mrfmmm.2012.11.001

Smith, J., Tho, L. M., Xu, N., & Gillespie, D. A. (2010). The ATM-Chk2 and ATR-Chk1 pathways in DNA damage signaling and cancer. *Adv Cancer Res, 108*, 73-112. doi:10.1016/b978-0-12-380888-2.00003-0

Sorek, R., & Ast, G. (2003). Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res, 13*(7), 1631-1637. doi:10.1101/gr.1208803

Soutourina, J. (2018). Transcription regulation by the Mediator complex. *Nature Reviews Molecular Cell Biology, 19*(4), 262-274. doi:10.1038/nrm.2017.115

Srebrow, A., & Kornblihtt, A. R. (2006). The connection between splicing and cancer. *Journal of Cell Science, 119*(13), 2635. doi:10.1242/jcs.03053

Srikanth, S., Jew, M., Kim, K. D., Yee, M. K., Abramson, J., & Gwack, Y. (2012). Junctate is a Ca2+-sensing structural component of Orai1 and stromal interaction molecule 1 (STIM1).

Proceedings of the National Academy of Sciences of the United States of America, 109(22), 8682-8687. doi:10.1073/pnas.1200667109

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*(5), 509-524. doi:10.1093/hmg/ddh051

Sun, Y., Bao, Y., Han, W., Song, F., Shen, X., Zhao, J., . . . Wang, Y. (2017). Autoregulation of RBM10 and cross-regulation of RBM10/RBM5 via alternative splicing-coupled nonsense-mediated decay. *Nucleic Acids Res, 45*(14), 8524-8540. doi:10.1093/nar/gkx508

Suzuki, S., Iwaizumi, M., Tseng-Rogenski, S., Hamaya, Y., Miyajima, H., Kanaoka, S., . . . Carethers, J. M. (2016). Production of truncated MBD4 protein by frameshift mutation in DNA mismatch repair-deficient cells enhances 5-fluorouracil sensitivity that is independent of hMLH1 status. *Cancer biology & therapy, 17*(7), 760-768. doi:10.1080/15384047.2016.1178430

Tacke, R., Tohyama, M., Ogawa, S., & Manley, J. L. (1998). Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell, 93*(1), 139-148. doi:10.1016/s0092-8674(00)81153-8

Torres, I. O., & Fujimori, D. G. (2015). Functional coupling between writers, erasers and readers of histone and DNA methylation. *Current opinion in structural biology, 35*, 68-75. doi:10.1016/j.sbi.2015.09.007

Trincado, J. L., Entizne, J. C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D. J., & Eyras, E. (2018). SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol, 19*(1), 40. doi:10.1186/s13059-018-1417-1

Trincado, J. L., Entizne, J. C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D. J., & Eyras, E. (2018). SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biology, 19*(1), 40. doi:10.1186/s13059-018-1417-1

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., . . . Darnell, R. B. (2006). An RNA map predicting Nova-dependent splicing regulation. *Nature, 444*(7119), 580-586. doi:10.1038/nature05304

Urbanski, L. M., Leclair, N., & Anczukow, O. (2018). Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *Wiley Interdiscip Rev RNA, 9*(4), e1476. doi:10.1002/wrna.1476

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

van Putten, M., Tanganyika-de Winter, C., Bosgra, S., & Aartsma-Rus, A. (2019). Nonclinical Exon Skipping Studies with 2'-O-Methyl Phosphorothioate Antisense Oligonucleotides in mdx and mdx-utrn-/- Mice Inspired by Clinical Trial Results. *Nucleic Acid Ther, 29*(2), 92-103. doi:10.1089/nat.2018.0759

Vaquero-Garcia, J., Barrera, A., Gazzara, M. R., Gonzalez-Vallinas, J., Lahens, N. F., Hogenesch, J. B., . . . Barash, Y. (2016). A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife, 5*, e11752.

Vaquero-Garcia, J., Barrera, A., Gazzara, M. R., González-Vallinas, J., Lahens, N. F., Hogenesch, J. B., . . . Barash, Y. (2016). A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife, 5*, e11752. doi:10.7554/eLife.11752

Venables, J. P., Bourgeois, C. F., Dalgliesh, C., Kister, L., Stevenin, J., & 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), 2289-2303.

Venables, J. P., Elliott, D. J., Makarova, O. V., Makarov, E. M., Cooke, H. J., & Eperon, I. C. (2000). RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing. *Hum Mol Genet, 9*(5), 685-694. doi:10.1093/hmg/9.5.685

Venkat, S., Tisdale, A. A., Schwarz, J. R., Alahmari, A. A., Maurer, H. C., Olive, K. P., . . . Feigin, M. E. (2020). Alternative polyadenylation drives oncogenic gene expression in pancreatic ductal adenocarcinoma. *Genome Res, 30*(3), 347-360. doi:10.1101/gr.257550.119

Vilar, E., & Gruber, S. B. (2010). Microsatellite instability in colorectal cancer-the stable evidence. *Nat Rev Clin Oncol, 7*(3), 153-162. doi:10.1038/nrclinonc.2009.237

Vinciguerra, P., & Stutz, F. (2004). mRNA export: an assembly line from genes to nuclear pores. *Curr Opin Cell Biol, 16*(3), 285-292. doi:10.1016/j.ceb.2004.03.013

Wahl, M. C., Will, C. L., & Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell, 136*(4), 701-718. doi:10.1016/j.cell.2009.02.009

Wang, J., Shivakumar, S., Barker, K., Tang, Y., Wallstrom, G., Park, J. G., . . . Qiu, J. (2016). Comparative Study of Autoantibody Responses between Lung Adenocarcinoma and Benign Pulmonary Nodules. *J Thorac Oncol, 11*(3), 334-345. doi:10.1016/j.jtho.2015.11.011

Wang, Y., Li, X., & Hu, H. (2014). H3K4me2 reliably defines transcription factor binding regions in different cells. *Genomics, 103*(2), 222-228. doi:https://doi.org/10.1016/j.ygeno.2014.02.002

Wang, Y., Wang, J., Gao, L., Stamm, S., & Andreadis, A. (2011). An SRp75/hnRNPG complex interacting with hnRNPE2 regulates the 5' splice site of tau exon 10, whose misregulation causes frontotemporal dementia. *Gene, 485*(2), 130-138. doi:10.1016/j.gene.2011.06.020

Wassarman, K. M., & Steitz, J. A. (1993). Association with terminal exons in pre-mRNAs: a new role for the U1 snRNP? *Genes & development, 7*(4), 647-659. doi:10.1101/gad.7.4.647

Watermann, D. O., Tang, Y., zur Hausen, A., Jäger, M., Stamm, S., & 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), 4774-4780. doi:10.1158/0008-5472.can-04-3294

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., . . . Kuster, B. (2014). Mass-spectrometry-based draft of the human proteome. *Nature, 509*(7502), 582-587. doi:10.1038/nature13319

Will, C. L., & Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor perspectives in biology, 3*(7), a003707. doi:10.1101/cshperspect.a003707

Wu, J., Huen, M. S. Y., Lu, L.-Y., Ye, L., Dou, Y., Ljungman, M., . . . Yu, X. (2009). Histone ubiquitination associates with BRCA1-dependent DNA damage response. *Molecular and Cellular Biology, 29*(3), 849-860. doi:10.1128/MCB.01302-08

Wu, J. Y., & Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell, 75*(6), 1061-1070. doi:10.1016/0092-8674(93)90316-i

Wu, X., Shell, S. M., Liu, Y., & Zou, Y. (2007). ATR-dependent checkpoint modulates XPA nuclear import in response to UV irradiation. *Oncogene, 26*(5), 757-764. doi:10.1038/sj.onc.1209828

Yoo, B. H., Bochkareva, E., Bochkarev, A., Mou, T.-C., & Gray, D. M. (2004). 2'-O-methyl-modified phosphorothioate antisense oligonucleotides have reduced non-specific effects in vitro. *Nucleic acids research, 32*(6), 2008-2016. doi:10.1093/nar/gkh516

Zhang, D., Qu, L., Zhou, B., Wang, G., & Zhou, G. (2018). Genomic variations in the counterpart normal controls of lung squamous cell carcinomas. *Frontiers of Medicine, 12*(3), 280-288. doi:10.1007/s11684-017-0580-1

Zhang, J., Sun, X., Qian, Y., & Maquat, L. E. (1998). Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. *Rna, 4*(7), 801-815. doi:10.1017/s1355838298971849

Zhao, H., & Piwnica-Worms, H. (2001). ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1. *Mol Cell Biol, 21*(13), 4129-4139. doi:10.1128/mcb.21.13.4129-4139.2001

Zhong, X. Y., Ding, J. H., Adams, J. A., Ghosh, G., & 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), 482-495. doi:10.1101/gad.1752109

Zhou, Z., & Fu, X. D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma, 122*(3), 191-207. doi:10.1007/s00412-013-0407-z

Zhou, Z., Wang, L., Ge, F., Gong, P., Wang, H., Wang, F., . . . Liu, L. (2018). Pold3 is required for genomic stability and telomere integrity in embryonic stem cells and meiosis. *Nucleic Acids Res, 46*(7), 3468-3486. doi:10.1093/nar/gky098

Zhu, J., Mayeda, A., & Krainer, A. R. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell, 8*(6), 1351-1361. doi:10.1016/s1097-2765(01)00409-9

Zou, W., Ma, X., Hua, W., Chen, B., Huang, Y., Wang, D., & Cai, G. (2016). BRIP1 inhibits the tumorigenic properties of cervical cancer by regulating RhoA GTPase activity. *Oncol Lett, 11*(1), 551-558. doi:10.3892/ol.2015.3963

Appendix

Table 1. Tra2α/β regulated targets identified by MISO

MISO identified alternative events regulated by Tra2 proteins: primers, genomic locations and PCR validation					
Gene	**HG19 Event coordiantes**	**Forward primer**	**Reverse primer**	**Internal/alt primer**	**Validated dPSI**
AKIP1	chr11:8934000:8934080:+	CGAACGGATAAGAGAAGAGGAG	TCGACCATGGGCAGCATAG		**-23**
AKT1S1	chr19:50376174:50376559:-	TTACTGAGGAAGAAATTGTTGACC	TTTGATATTCTCCATTGTTGGAA		**-34**
AP1AR	chr4:113184144:113184242:+	CCATGGTGGTCTTGCGGG	GTCCGCCATTTCTTCTGCAT		**-58**
APOPT1	chr14:104037960:104038157:+	AATCGGTACAAGATGGCGGA	CTCGCCATGTAAAGCCTGTC		**-37**
ATL2	chr2:38570410:38570654:-	CTTTGCAGGAGTTAGCCGAC	TCCTTGCGTCTGTCCTTCTT		**-45**
BCLAF1	chr6:136588167:136588313:-	CAGACTTGGCCCAGTGTTTC	TAAACCCTGGAAGTGGTGCA		**-30**
BIRC5	chr17:76212745:76212862:+	CAGATTCAGAAGCGTCAACAGT	TGCTCTTGTTTTCGATGCCT	TGGGAGAATCTGACTTGCCA	**-31**
CCPG1	chr15:55657386:55657507:-	GACGAACCAGTGAGCTAAGC	TGGAAGAGCACATAAACCGAG		**-37**
CDKN3	chr14:54866612:54866694:+	GGACTCCCCTCACAAGTTCT	GCTTGGGTCTTACTTGCTGG	ATTGATGGAGTTGGCAGCAC	**-40**
COX10	chr17:13980052:13980373:+	CCTGCAACTTCGCCACTTAA	TGACTTCTTCTAGGGATGGCA		**-43**
DEPDC1	chr1:68960113:68960378:-	TTGAGAGAGACACGACACGG	ACTTCCCTCTCCAGTCATGC	CAGACACAAGCTGGAAAGAGT	**-37**
FAM53C	chr5:137677498:137677555:+	AGGACAAGACCATCGAGCTC	CGGTAATAAAGTCCCGCTGC		**-27**
FBXW11	chr5:171341347:171341409:-	GGATTTCCTTTTGCCTTCGGT	GAGTTTGAGACGAACCTGGG	ATGAGGAGAAGAGCACAGGG	**-36**
HNRNPC	chr14:21679565:21679722:-	CCTGCAAAAGTGGAAGCGAA	TCCTCAGTCTTCGTTTCCCC		**-34**
IP6K2	chr3:48732127:48732257:-	GGCCATTAAGAAGGAGCTGA	AAGAGTCATCCTCGCCATTG		**-45**
MAPKAP1	chr9:128434679:128434922:-	CTTCTGGAGCCCTTTGTCCA	ATCCTCCCAAAGTCTTCCCC	CCAACAAAGACCCCACCATG	**-36**
MBD4	chr3:129155286:129156151:-	CTCCCCTCAGGAACGTGT	TTGCATTTGATCTGGTTTTGTC		**-76**
MFF	chr2:228193394:228193505:+	CAAGAACCCATCGCTTCTGC	ATCTGTGTTCGTGGGATGGT	AGTCACCTTTCCTTTGGGCT	**-32**
MPP5	chr14:67745735|67745967:67746254:+	CACGAACACTCTTCCGCTAC	GAGAAGGAAATGCTGCCCTG		**-36**
MRPS33	chr7:140710417|140710460:140710219:-	ACCGGAGAAGAGAAACTAAAAGG	CCTCTCTCCTACGCCTCATG		**-33**
NUSAP1	chr15:41643186|41643231:41643327:+	GCAACCAGGAAGAAGCTGAG	AGATCCTGGCTTTCCTGCTT		**-22**
PARPBP	chr12:102542013:102542241:+	GTGCAGACTCCATGCTCTTG	ACGGTTGTATTTACTTGTTGGTG		**-31**
PCNP	chr3:101298634|101298686:101298848:+	GGGAGACGAGAAGCCTGAAA	TGTAGGCTTTGTTGGGAGGT		**-31**

PPP2R3C	chr14:35585816:35585943:-	GAACTTGTGGCAGCCTGAAG	CCCCTGCTCACAACACATTT	TTGTTCCTCCCTACTGCGAG	**-40**
RBBP4	chr1:33145241|33146127:33146254:+	GACGCCCAACACCTGTCC	AACAGTTCTCTGCTTTTTCTTTG		**-59**
RNPS1	chr16:2314574:2314761:-	TGGTGGTCATACTGCCAAGA	GCTTTGTTGAATCAAGCCCC	ATGGAGAGTAACAGGCCTGC	**-26**
SGOL1	chr3:20215741:20216547:-	TGAACTTCCAGGACAAGGAGA	CTGCTGGGCTTGCTTTATTC		**-38**
SNHG6	chr8:67834559:67834627:-	CGAAGAGCCGTTAGTCATGC	ATACATGCCGCGTGATCCTA	AAGAACATTCATCTACAGCAACC	**-35**
SUV39H2	chr10:14938845:14938976|14939516:+	GAAGTGCCCGTTACTGCTTC	ACTGTCCTCGTCTTTCAGCT		**-44**
TCF19	chr6:31126674|31126920:31127484:+	TGCTTGAAGAGGAGGCTACT	TGCAAACTCAGGAACACGTC		**-24**
UTP23	chr8:117782431:117782605:+	GCTTTCCATGGTTGAAGAGGG	TACCACCAGAGCAGAGTTGG	CTGGCTAGACTGGTCACTGT	**-49**

Table 2. List of 31 MISO identified targets showing affected protein domains, type of alternative event and co-identification of the event by MAJIQ

Gene	Identified by MAJIQ	PTC/Frameshift	Pfam predicted effect	Type of event	Validated dPSI
AKIP1	**Yes**	-	**N/A**	Cassette exon	**-23**
AKT1S1	**Yes**	-	**N/A**	Alternative 5' splice sites	**-34**
AP1AR	**Yes**	-	**N/A**	Cassette exon	**-58**
APOPT1	**Yes**	-	**N/A**	Cassette exon	**-37**
ATL2	**Yes**	-	**GBP**	Cassette exon	**-45**
BCLAF1	**Yes**	-	**N/A**	Cassette exon	**-30**
BIRC5	**Yes**	**Yes**	**BIR**	Cassette exon	**-31**
CCPG1	**Yes**	-	**N/A**	Alternative 3' splice sites	**-37**
CDKN3	**Yes**	**Yes**	**CDKN3**	Cassette exon	**-40**
COX10	**Yes**	**Yes**	**UbiA**	Cassette exon	**-43**
DEPDC1	**Yes**	**Yes**	**DEP**	Cassette exon	**-37**
FAM53C	**Yes**	-	**N/A**	Mutually exclusive exons	**-27**
FBXW11	**Yes**	-	**N/A**	Cassette exon	**-36**
HNRNPC	**Yes**	-	**N/A**	Cassette exon	**-34**
IP6K2	**Yes**	-	**N/A**	Cassette exon	**-45**
MAPKAP1	**Yes**	-	**SIN1**	Cassette exon	**-36**
MBD4	**Yes**	-	**MBD**	Cassette exon	**-76**
MFF	**Yes**	-	**DUF800**	Cassette exon	**-32**
MPP5	**Yes**	-	**N/A**	Alternative 3' splice sites	**-36**
MRPS33	**Yes**	-	**N/A**	Alternative 3' splice sites	**-33**
NUSAP1	-	-	**N/A**	Alternative 3' splice sites	**-22**
PARPBP	**Yes**	-	**N/A**	Cassette exon	**-31**
PCNP	-	-	**N/A**	Alternative 3' splice sites	**-31**
PPP2R3C	**Yes**	-	**N/A**	Cassette exon	**-40**
RBBP4	**Yes**	-	**N/A**	Alternative 3' splice sites	**-59**
RNPS1	**Yes**	-	**N/A**	Cassette exon	**-26**

SGOL1	**-**		**N/A**	Cassette exon	**-38**
SNHG6	**Yes**		**N/A**	Mutually exclusive exons	**-35**
SUV39H2	**Yes**	**-**	**SET**	Alternative 5’ splice sites	**-44**
TCF19	**-**		**N/A**	Alternative 3’ splice sites	**-24**
UTP23	**-**	**-**	**Fcf1**	Cassette exon	**-49**

Table 3. Tra2α/β regulated targets identified by MAJIQ

Gene	HG19 Event coordinates	Forward primer	Reverse primer	Internal primer	Validated dPSI	MAJIQ predicted dPSI	Pfam predicted effect	Identified by MISO
AKIP1	chr11:8,932,990-8,936,477	CCACCTAGAGAA ACAGCCGG	CACATGTGCAGC TTCGACTC		-8	**-20**	N/A	**Yes**
ANTXR2	chr4:80,822,303-80,828,621	AGGGAAGCAGG AAGACCAAG	ATCAAAGCAGGC GCACTTTG	Gttgctttttcctcttc tacacattc	-11	**-35**	N/A	
ARHGAP11A	chr15:32,921,796-32,925,309	ATGCTAAGCGTA CATTGCCAG	TTCGCCTTGGCAC ACCTG		-4	**-21**	N/A	
ASNSD1	chr2:190,530,083-190,532,671	AGTATATAGGCA ACAACAGAACAGC	ACGTGAGCAGAA AATAAACACTGG	GAAGAAATTCGA CCCAGTTCCATC	-39	**-55**	MBD	
ATP6V0A1	chr17:40,620,028-40,629,760	ACACCGGTGAAA ACCCAGAG	GGAGTGCCTCTT CCCATCTC		-16	**-34**	Cep57_MT_bd	
BRIP1	chr17:59,924,462-59,934,592	CAACTATCCAAGC ACACCACC	TTCTGTGGCGAAA AGGAGTTTATC		-84	**-43**	N/A	
C19orf25	chr19:1,461,142-1,478,915	AGATGGGCTCCAA GGCAAAG	CCTCTGCCTCAGC ACGTTG	CTTGAACTGGG GAGTCGGAG	-5	**-35**	N/A	
CASP3	chr4:185,559,555-185,570,663	TCCTAGCGGATG GGTGCTAT	TGAGTTTTCAGTG TTCTCCATGGA	AGTTTCGTGAGT GCTCGCAG	NS	**-25**	N/A	
CASP8	chr2:202,137,361-202,139,676	CCTGCTGGATAT TTTCATAGAGATGG	CTGTGATTCACTA TCCTGTTCTCTTG		-25	**-27**	N/A	
CBWD1	chr9:175,698-186,011	GGAGGAGGAAA AGTCTGGCC	ACCGCTACTCTTT TACTATGTTGC		-14	**-23**	CBFD_NFYB_HMF	
CCBL2	chr1:89,435,092-89,458,459	CAATCCCTGTCC CGGCTC	TCAAGTCCTTCA ATCCGTTTTGC		NS	**-32**	N/A	
CDADC1	chr13:49,829,975-49,842,195	AAGAGGGAAAGC ATGGACCC	TCATTTTCAAACA AGCAGAACATGG		-17	**-31**	N/A	
CENPQ	chr6:49,431,091-49,438,741	TTTACGAATGTTT GGGCGCG	CTTCAGAAGACAG ATCTTTCAGATGAT		-75	**-54**	N/A	
CHCHD4	chr3:14,157,926-14,166,370	CCAGAGGAGAGG GAGGTCAC	TGCTCCTCGTATGG ATCGTTG		13	**25**	N/A	
CHEK1*	chr11:125,525,120-125,546,150	GAGCAGCCAGAA GATTTGGC	CTAGCCAGGTGTG GTGGTG	CTTCACAGACTGA TAAATCTACTGTCC	-12	**-23**	NUC129	
CIRBP	chr19:1,271,980-1,274,594	GAGTGGTGGCTA CAGTGACC	ACGATCTTGAGCA GGAAGGG	AGCACTTACAGG GGACAAGC	-10	**-23**	Neuregulin	
COPB2	chr3:139,074,442-139,077,585	ACTCCGGTTATTG TGGCCTC	TGGAAAGCATTAC AGTCAATCATCC	GGCACACACTTTT ATAGGACTGTG	-42	**-52**	N/A	
CSDE1	chr1:115,282,374-115,300,648	TATTCCTACGTA CCGGGCCG	TGCTGCTGAAGT ACCATTAGGG		-11	**-30**	N/A	

CSF1	chr1:110,467,769-110,473,614	TGACAGACAGGTGGAACTGC	ATTCAGTCAAGGGTCTGCGG	CAGGGTAGAGATGGTCTGGC	-5	**-24**	MRG	
DDX5	chr17:62,495,734-62,498,187	TGGAAAAGCTCCTATTCTGATTG	AGGTATTAAATCCACCCCTTTTG		-13	**-28**	N/A	
DNAAF2	chr14:50,091,892-50,101,948	TCCAGAGAGCCATGGACATTG	CAGAATCAGTTGTAGGGGTGTTAC		-13	**-23**	PID	
DPH5	chr1:101,479,205-101,491,343	TTAGACGCTGCAGTCGAGTG	AGGAATTCCCAGCTTTGTTGC		-8	**-18**	BAH domain	
DPY19L1	chr7:34,978,888-34,981,496	CTCTTCTTGACACCACACATGTG	GGATGATTCACAATGGGCCG		-24	**-43**	Arf	
DROSHA	chr5:31,515,095-31,526,527	GGAGAGACACAGGCATCGAG	TCTCCTCGGGCTCTTTTTCC		8	**21**	N/A	
FAM118B	chr11:126,104,890-126,120,628	GATGGCTTCTACAGGGAGCC	AGGGCTCCATTTTCCATCAGG		-41	**-36**	cobW	
FAM122B	chrX:133,923,253-133,927,923	CCAACCTTACACACTTAGAACTCG	CATGTGCAGTTTCTCTGTTCACC		16	**38**	N/A	
FAM35A	chr10:88,856,691-88,917,912	ATGTCTGTTGCTGACCCCTG	TGTAGTACTGTCTCGCCAATCC		-38	**-31**	N/A	
FAM45A	chr10:120,863,598-120,867,676	GGCGGACACTCAGCTGATG	GGAAGAATCTGGAACTTCAATTGTTG	GTACGGCCTATGGTGTGTCC	7	**21**	UCH	
FARSB	chr2:223,507,570-223,521,056	GTGAGTTCGACACACCATGC	ACCAATCCTTCCAGACACAGG		28	**30**	N/A	
FNDC3B	chr3:171,851,261-171,945,053	TTCTCGTTCAAGTTAATCCAGGTG	TCAATCGCAGGCTCTCTTCAC	ACCTGGAGGCACATGAATGG	NS	**-38**	N/A	
GBP3	chr1:89,477,430-89,479,962	TTTCTCCCTGGACTTGGAAGC	CCAGGACTGCGTTCTCCATG		-7	**-26**	N/A	
GINS1	chr20:25,388,363-25,405,963	ACTACCATGTTCCTCGAACGAG	GGGGTGGCTATTGACAGGAC		-9	**-22**	PseudoU_synth_1	
GLS	chr2:191,796,271-191,818,352	GATCTTGTTTCTCTGTGTAATTTCCAT	TTATGGCTCCAGCACCCAAG	ACACATCTCCAGTATATGCAGCA	-9	**-21**	N/A	
HMGXB4	chr22:35,653,445-35,659,219	CGCGCTTTTCTCTCCTTCTC	AGAATTCCTGACCTGAGCAGC		10	**30**	zf-AN1	
IFI44	chr1:79,115,481-79,119,964	GTGCCTGCCTAGCTATCCAG	TGACCCCTTTTATCTTTCTTTCATCC		-23	**-32**	CHGN	
IKBIP	chr12:99,007,183-99,028,191	AGCAGTCAGAAAAATTTGCAAAGG	TGCCATGGAAGTTGTACTTTGG	GCATCCTTTGAGTGAGCACC	3	**20**	N/A	
INO80C	chr18:33,048,297-33,077,955	GAACAGCAAGAAGAGGCCGG	CCTTCCTCAGGGCCAGGTAG		-15	**-25**	MAP7	
JOSD2	chr19:51,009,635-51,013,705	CCTCAACAACGTTCTGCAGC	GCAGCTTGGAGTCCAGGTTG		6	**22**	WD40	

KIAA1715	chr2:176,788,620-176,802,584	GAGTTTAGTTTTG AGAAGAGGCAGG	GTCAGATGCTTTTT CAATCACTTGG		9	**21**	N/A	
LPAR1	chr9:113,735,761-113,798,481	CCTGGAGGGAAG TTGCCG	TGCTGTAGGTGTCA GTCCTG	AAGAGAGGCCAC ATTTCCGG	NS	**-23**	N/A	
LYPLA1	chr8:54,965,214-54,974,884	GGCTTTCACCAG ATTCACAGG	AAGGAAGCCCGAA GTGGAAG		-7	**-21**	N/A	
MADD	chr11:47,307,984-47,311,044	ATCGGAACCACA GTACCAGC	AATTGCTGGATCGA CCCTGG		NS	**-23**	N/A	
MAPKAPK5-AS1	chr12:112,278,137-112,280,706	TCTGTCAATGGT TCTTCATTCACG	ATCCATCTTCTCAGC CAGGC	GCAGTAGCAGC TCTGTTTACG	-27	**-21**	Ribosomal_L22	
MED16	chr19:875,244-880,148	ATGCAGCTATCG TGGACGTC	TGTGGTAGTCGCAC ACGC		6	**26**	N/A	
MKKS	chr20:10,389,272-10,401,406	GCAGCCCTGTTT TCCAAATACC	TCCCAGCTACAAGAA GCCAG	CTTCAGCTCATC CCAGGCAG	-25	**-54**	N/A	
MR1	chr1:181,003,067-181,018,833	TTGGGAGAAGGC CTTGTGTG	TTCTGCCGAGTGACA CTGTC		NS	**32**	N/A	
MRPS33	chr7:140,705,854-140,714,479	AGCTTGGAGCCG GGTAGAG	AGAGTCCAAGAAATC GGAGCG		-8	**-20**	N/A	**Yes**
MSH2	chr2:47,707,835-47,740,373	TTGGGATTCATGT TGCAGAGC	AGGGCATTTGTTTCA CCTTGG	Aacacagagataga ttctttgccatc	-11	**-20**	N/A	
MTA1	chr14:105,911,755-105,916,521	AAGTGGTGTGCTT CTACCGG	GAGAGGAACAGCTC CCGATG		6	**21**	N/A	
MYEF2	chr15:48,444,071-48,446,273	AGGCATTGGGATG GGACTTG	TCGGCCAACTCCTCC AAATC		-8	**-29**	Cyt-b5	
NAA35	chr9:88,590,029-88,592,361	GATGTGGAGGATG ACATGCAAAG	GCACTGTCAGTAAC ACACGAG		-28	**-32**	N/A	
NCK1	chr3:136,664,425-136,665,137	ACCTAGTGTGCCA GATTCTGC	TACCAAGGATTGCC AGCAAAC		-9	**-33**	N/A	
NRG1	chr8:32,617,718-32,622,548	CCGTAGAAAACAG TAGGCACAG	TCTTCTTCCATGAAG GGGCTG		8	**25**	N/A	
NT5C3A	chr7:33,053,742-33,055,538	GGAATCGGCGATG TACTAGAGG	AGAGTTGGCTACTT CTAATGATTCATC	Cgcctgtagttccagc tacc	NS	**32**	BTB	
NTPCR*	chr1:233,092,093-233,119,628	CCTGGAAAACGTG AATGCCG	CAGGCTCCTCTTGAA CTGGG		-13	**-27**	MMR_HSR1	
OARD1	chr6:41,032,517-41,040,059	AGTGAGGATTGTC GCATGGG	CGGTCCTATGGCCCA GTAGA		-7	**-14**		
ORC3	chr6:88,304,071-88,313,246	TGCTTTGTTTTTAA GCCAAACTCC	TTGACCGCCCAAGTC TCTTG	TCTTTTCAATTAGA GCCCTGGGG	NS	**21**	Mod_r	
OSBPL6	chr2:179,203,718-179,214,116	TGCACATTGCCAG TCAAACC	TCCAGAGGGTCAGTG AGGTG		15	**25**	N/A	

PARP4	chr13:25,067,734-25,072,367	TGGTGGAGCTTCAGTGTTCG	GAAGCATGTGTTCCAGGTGG		-13	**-33**	N/A	
PNPLA8	chr7:108,154,588-108,166,762	TTGTCTTCTCCTCCAAGGTCTAC	GTCCCTTGGGAGCAGAAGTG	GCAGAGCTTTGGGTCAGTTG	-50	**-55**	DUF800	
POLD3	chr11:74,347,242-74,380,162	CTTACCTGGATGGGGAAGGC	ATGGACACCTGTCTGTTGGC	GCAGAAGGCTCAGCTCTTTTTC	-12	**-21**	N/A	
POLR3C	chr1:145,608,404-145,608,659	AGAAAGCCCTGTGTGTCCTC	CTCCATGGTCTCTGTGAGCC		-6	**-22**	Myb_DNA-binding	
PPP2R3C	chr14:35,579,731-35,591,307	TGGAGCAGGTAGGAAACAGC	GTAGAAAGACAGCTCTTGATTCCTC		-19	**-24**	N/A	**Yes**
PRKRA	chr2:179,309,149-179,315,958	ACATAGAGCTGCAGAGGCTG	TGGTTGCTTGGAAGGGTCAG		6	**23**	N/A	
QKI	chr6:163,899,812-163,983,101	CTTGAAGCAGAAACCGGATG	CCATTCAGAATCGCAAGCTC		-16	**-27**	N/A	
RBBP5	chr1:205,065,810-205,066,523	TGGACCCTATTGCTGCCTTC	CTTCTTAGGTGGCTGGGACC		-16	**-24**	N/A	
RCOR3	chr1:211,474,803-211,489,727	GTTGGACCACAGAGGAGCAG	GGACGAAGAAGTGGTGGAGG		-9	**-20**	N/A	
RGS10	chr10:121,275,021-121,302,220	CACCAGAGCCTCAAGAGCAC	TCTGGAACATCAGAGGGTGC		NS	**-18**	N/A	
RIC8B	chr12:107,254,046-107,283,090	ATGGGAATGCTGCAGGACTG	GGATGGGAATACAATACCTGGAAAG	AGTCTGTGTCAGAGCTGGTC	-10	**-24**	N/A	
RNPS1	chr16:2,313,898-2,318,413	CTTCCCGCTCCTTGACTCTG	GGTACTGCTGCTACCACTGG		-15	**-26**	BTB	**Yes**
SAT1	chrX:23,801,917-23,803,546	ACCCCTTTTACCACTGCCTG	TGCCAATCCACGGGTCATAG		11	**24**	N/A	
SLC35A1	chr6:88,182,643-88,211,028	GAACCATGGCTGCCCCGA	CACCACTTTTGTAGCTTGGGC		-18	**-23**	RNA_pol_L	
SPPL2B	chr19:2,351,432-2,353,565	TGGCCTCCTTGTGACATTCG	CTCGGACGTGCGTGATTTTG		NS	**-28**	GBP	
SS18	chr18:23,615,055-23,618,623	CACCACAGCAGTACTCAGGC	CTGCTGCTGTCCTGGGTAAC		16	**27**	N/A	
STARD4	chr5:110,841,914-110,848,288	ACTGTTGAGAGCGGTGTGAG	CTTCTGAGGGTTTTCTCCAAACAG		-14	**-27**	N/A	
STAU2	chr8:74,529,527-74,659,172	GCGGGGTCAGTTCTCTGTAG	AACTGCTGTATCCCTCACGG		-17	**-35**	N/A	
TBP	chr6:170,863,390-170,871,321	AGTGACATTATCAACGCGCG	GCTGCCTTTGTTGCTCTTCC		-23	**-39**	BolA	
TBRG1	chr11:124,495,567-124,496,946	GAGCTGGGCCTATTTCAGGG	TCTGGTCTGGGCACTTCATG		-9	**-20**	N/A	

TMEM126B*	chr11:85,339,629-85,346,131	CCACCGGCAAGTCACATGAG	CCTTGAAGCAGCGTCTGAAC		-25	**-55**	N/A	
TMEM159	chr16:21,169,698-21,172,599	GCTCTGGCTCCCTCTCGG	AGCTCACTTCAATCTCTTTAGGGATG		-21	**-25**	MoaE	
TOM1	chr22:35,713,870-35,719,170	GGGCCCTCAACATGGAGATC	CACTCTCCACGAAGTCCTGG		-15	**-27**	IL1	
TRIOBP	chr22:38,155,161-38,161,824	GTGCACTGAGATCCCAGGAG	AAGGGTCCAAGTGCAGAGTG	CACTCCTTCTCCTGCTGCAG	-7	**-22**	N/A	
TRMU	chr22:46,733,676-46,742,441	AACATGGGGTCTGTACTGCC	CGGGCTTCTTAACGTGCTTC		-10	**-37**	N/A	
TTC13	chr1:231,076,186-231,090,149	GTCTTGATTGGCAGTGGTCTG	GGGTTCCCCGATGTCTGTAC		3	**21**	Utp12	
TTC4	chr1:55,182,272-55,186,913	TATGTCGAGAGCGCCATCAG	TGAGGTGGCAGGGTTTTAGC		-22	**-46**	PH	
UBAC2	chr13:99,890,681-99,966,474	CCTTCACGCAGTCAAGAACG	TTGTGATGGACAACGGACCC		-15	**-23**	HTH_9	
UIMC1	chr5:176,385,094-176,402,481	TGGGTTGCAGAAAACGAAGAC	AGCAGGTACAGAGTTTCCCTC	AAAAGCAGAACCCCTCCCAG	-61	**-49**	SET	
UMAD1	chr7:7,712,940-7,841,374	AGCAGCAGCAGCAATGTTTC	TTGGTTGGCCTCTATGTCCG		16	**40**	N/A	
VPS37A	chr8:17,137,884-17,144,120	GCAAAGGCAGCATGAACTTAGTG	CTGTGCATTGCTATCGCCTG		-13	**-37**	N/A	
WDR20	chr14:102,661,275-102,691,184	CAGCCACAGCAGAAAGTGTC	GTTGAACACCCGCAGAAACC	CACTCCCTCGAGTCCAAGTC	NS	**-25**	N/A	
WDR48	chr3:39,093,489-39,116,441	GCCTGGCCATGAATCAACTG	TACATCTCTGCTGGCCAAGG		-28	**-46**	N/A	
XPA	chr9:100,437,191-100,456,041	TGACACAGGAGGAGGCTTCA	TCGTCTCCCTTTTCCACACG		-62	**-63**	UbiA	
ZFP82	chr19:36,874,022-36,896,574	GAGCCTTGGAAAGTTGTGAGG	TGAGACCCTTAAGGCCATGG	AGATAGGACTAGTGGGCTTTCTATTC	-7	**-11**	tRNA_Me_trans	
ZNF655	chr7:99,158,130-99,174,076	CTCCAGAGCAGTGATGGAGG	TCCTGAAGTCTTTCTAATCTGTCCTC	AAAGGAGTTTCCGAGCAGGG	20	**30**	zf-C2H2	

Table 4. List of targets previously identified by DexSeq and validated by Dr Best.

Gene	HG19 Exon coordinates	dPSI	Identified by MAJIQ	Identified by MISO	Event type	PTC/ Frameshift	Domain affected
ANKRD1	chr10:92678888-92679025	**-73**	Yes	-	Cassette exon		N/A
CDCA7L	chr7:21951235-21951372	**-26**	-	Yes	Cassette exon		N/A
CHEK1	chr11:125497502-125497725	**-55**	Yes	Yes	Cassette exon	Yes	Tyrosine Kinase
CSDE1	chr1:115292442-115292828	**-5**	Yes	-	Cassette exon		N/A
GLYR1	chr16:4881980-4882222	**-41**	Yes	Yes	Cassette exon		N/A
HDLBP	chr2:242194789-242194995	**-27**	Yes	Yes	Cassette exon	-	KH_1
KIAA0586	chr14:58926561-58926788	**-37**	Yes	-	Cassette exon		N/A
MSL3	chrX:11779619-11779701	**-41**	Yes	-	Cassette exon	-	MRG
NAP1L1	chr12:76462689-76462774	**-11**	Yes	Yes	Cassette exon		N/A
NUB1	chr7:151046159-151046326	**-41**	Yes	-	Cassette exon		N/A
PDCD6IP	chr3:33895371-33895505	**-4**	Yes	-	Cassette exon		N/A
PPP1R7	chr2:242092891-242093019	**-23**	Yes	-	Cassette exon		N/A
SMC4	chr3:160137146-160137331	**-66**	Yes	-	Cassette exon	-	SMC_N
TBC1D12	chr10:96256812-96256929	**-39**	Yes	-	Cassette exon		N/A
ZCCHC7	chr9:37126309-37126939	**-72**	Yes	-	Cassette exon		N/A
ZCCHC9	chr5:80600577-80600960	**-68**	Yes	Yes	Cassette exon		N/A

Table 5. List of *RBMX* regulated genes in MDA-MB-231 cells. List of target genes and respective alternative events regulated by RBMX (knockdown/overexpression) and Tra2 proteins

Gene	HG19 Coordinates:	Forward primer	Reverse primer	Internal primer	dPSI in RBMX kd	dPSI in Tra2 kd	dPSI in RBMX OE	PTC/ Frameshift	Type of event
ARFIP1	chr4:153,701,261 -153,750,878	CCATCTGGACTTGGT CTCTCAG	CAAGTTCTAACTT TTCCATTGCAGG		-0.12	|	-0.01	Yes	Casette Exon
C6orf89	chr6:36,853,630- 36,867,409	CTAGGGTACAAGGC CTCGC	GGGGGTCTCTGA GGTTCATTC		0.20	|	-0.09		Casette Exon
CD44	chr11:35,211,382 -35,236,461	GCTTCTCTACATCAC ATGAAGGC	**TCAGATCCATGA GTGGTATGGG**	ACCCGCTATG TCCAGAAAGG	0.30	|	-0.03		Alternative PolyA Site
ETAA1	chr2:67,629,926- 67,632,467	**GCTGGACATGTGGA TTGGTG**	GTGGGAGCTGCA TTTACAGATG	GTGCTCCAAA AAGCCTCTGG	-0.67	|	0.13		Exitron
FNIP2	chr4:159,754,680 -159,757,810	TGCAAGACAGCTTTG AGTACATC	ATGCCACTGTCC CTGTCTTC		-0.30	|	-0.13		Casette Exon
STX3	chr11:59,562,845 -59,573,354	TCATGCACACAGTGG ACCAC	GTACTTGAGAGC TGGTGGGG		0.16	|	0.01		Intron Rentention
HMGXB4	chr22:35,653,445 -35,659,219	CGCGCTTTTCTCTCCT TCTC	AGAATTCCTGAC CTGAGCAGC		0.11	49	-0.09		Casette Exons
ERI2	chr16:20,802,100 -20,810,695	ACGATTCTCGGAATA CTGCCC	GAGACGCCTCCA GGAAGAAC	ACTTGTAATT GATCCTGTTG TGCC	-0.33	-28	|		Casette Exon
FAM122B	chrX:133,923,253 -133,927,923	CCAACCTTACACACT TAGAACTCG	CATGTGCAGTTT CTCTGTTCACC		0.07	45	|	Yes	Casette Exon
FAM35A	chr10:88,856,691 -88,917,912	ATGTCTGTTGCTGAC CCCTG	TGTAGTACTGTCT CGCCAATCC		-0.66	-66	|		Casette Exon
FARSB	chr2:223,507,570 -223,521,056	GTGAGTTCGACACAC CATGC	ACCAATCCTTCCA GACACAGG		0.15	76	|		Casette Exon
FGFR1OP2	chr12:27,110,534 -27,116,388	AGAACATCAGTCGGC CTTGG	GTTCCTTGCAACC CTGTTGC		-0.21	-54	|		Alternative 3' Splice Sites
MYO18A	chr17:27,409,334 -27,413,819	GGAGGCTGCTAACC AGAGC	CATCATCAGAAG CTGCCTTGG		-0.25	-74	|		Casette Exon
OSBPL6	chr2:179,203,718 -179,214,116	TGCACATTGCCAGTC AAACC	TCCAGAGGGTCA GTGAGGTG		0.13	91	|		Intron Retention
PCTP	chr17:53,852,153 -53,900,451	TAACCCGGGTGGCCA AATTC	GGTAGTTCTGAC AGGCTCTTGC	cctgcacccatttc agtgtttc	-0.14	-19	|		Casette Exon
PI4KB	chr1:151,288,049 -151,300,191	CAGTGTCGGCAGTCC AACC	AAGTGGGCTCAG AAGTTGGC		0.15	-25	|	Yes	Casette Exon

SS18	chr18:23,615,055-23,618,623	CACCACAGCAGTACTCAGGC	CTGCTGCTGTCCTGGGTAAC		0.23	50	|		Casette Exon
STK4	chr20:43,600,719-43,607,212	GCAGCTGAAAAAGTTGGATGAAG	GTCACATTGCTGCATTATAGAGATTTC	aagctttgttttctccctggtg	0.15	6	|		Alternative Promoter
ABLIM3	chr5:148,619,322-148,624,578	**AGCCTTTCCCCATTGGAGAC**	AGGTCTGGCACTTGAAGCAG	tatggtggtggtgtgatggc	0.44				Casette Exon
ASPH	chr8:62,496,502-62,546,286	**AGTAACTGCTCCCCCTGAGG**	CCTTCCCAAAGATCTTGATCTGC	TTTTGCTTTTTGTTCTGGATCATCTG	0.19				Casette Exon
ATRX	chrX:76,937,012-76,940,498	TGAAACTTCATTTTCAACCAAATGCTC	ATCAAGGGGATGGCAGCAG		-0.65				Alternative PolyA Site
CARS2	chr13:111,315,798-111,319,820	TGCCTTTCTAGGAGCTGCTG	CTGCTGTGTTTGGTGCCATC	ACTCATCTCTGGTTCTGCGC	-0.12				Alternative PolyA Site
FDPS	chr1:155,278,759-155,279,996	ACCGAGCAGGAATCCGTATC	TCTGGGAGAAGTGCTGAACG		0.10			Yes	Casette Exon
FMNL3	chr12:50,050,865-50,055,832	AAGGCCTGGATGTACTGGTG	GGACGTCATCCTTCTGGCTC		-0.25			Yes	Casette Exon
FNBP1	chr9:132,671,169-132,687,436	GGTTTGAGCCTCCTGGAGAC	CATCGACTTTCTGCTGCAGC		-0.45				Alternative PolyA Site
GALNT3	chr2:166,626,696-166,651,192	CTGCACCAAGCTTCCAGAAC	TCTACTGAGCCGCAACTGTG	CTTCTGGTTCCCCAAGGACC	-0.12				Casette Exon
KIAA1217	chr10:24,831,349-24,834,245	**TTCCATGAAGATGTACGGAAATCTG**	TTCTTGGCACTCATGGGGAC	CCTCTTTCCCACTTAGGAGGG	0.20			Yes	Casette Exon
KLHL7	chr7:23,164,599-23,180,563	**TGTTCAGTCTTTGCTGGATGC**	GCTCTCACAGTCAGAGTGTCC	CTGGACGTTCATTCATGTGCC	0.36				Casette Exon
LINC00467	chr1:211,570,109-211,599,625	**TGCTGCCAAGGGAAAAATTAAGAAG**	GTCAGGAGTTCGAGACCAGC	GTGACATATGTGGTTGTTTCCTCC	0.20				Alternative PolyA Site
LRIF1	chr1:111,492,473-111,506,701	**GCAACTCCAGCCCAACATTC**	TTGTCATTTCTTCCTTGGGCAC	TCCCTTTCAAAGCATCAGACATG	-0.31				Alternative PolyA Site
LRRC8E	chr19:7,953,390-7,960,626	AGCTCCGCTCCCAGAGAC	CAATCATGAGCATGGCCACG		0.08				Casette Exon
MROH1	chr8:145,255,331-145,275,576	**TGGACACCAGCAATGAGAGG**	TTGCTGTTGGTGTCCAGGAG	TGTAACAGCCGTGGTGACTC	0.18			Yes	Casette Exon
N4BP2L2	chr13:33,101,012-33,112,970	TGGGAGGTAAAGGTCAATGC	**TGTTCTGGTTCCAGTCATGG**	AATGGCTGCAGTATGCAAGA	0.24				Casette Exon
NCOA7	chr6:126,236,479-126,242,214	TAGCTCAGCGTGGCTACAAG	AGTAGCTGCATGTGCTCTTCC	TCAGTGGTCTCCCGATGTC	0.09			Yes	Casette Exon

NFX1	chr9:33,347,036-33,351,880	**AGTGAGGAGAAGTGTCCCCC**	TACAAGCAGTCACAGGGCAG	TCCACTTGTTGACTGATGGGC	0.18			Yes	Casette Exon
ODF2L	chr1:86,814,344-86,818,667	**GAGAAAATGTCATCTAGAGAGAGTGC**	ACTCTCATTTTGATTTTCCGACTTCTC	GGACATTACAGCAAAACCCACC	-0.42				Alternative PolyA Site
OGFOD3	chr17:80,350,091-80,356,195	CTCCTTCGACTACACCTCGC	CAACAGGAGCGGGTGGAC		0.24				Casette Exon
PIGQ	chr16:630,858-634,136	TGCTCGTGGACCTCATCAAC	CAGCAGTTCCCTCAGTCCTG		-0.11				Casette Exon
RAPH1	chr2:204,298,405-204,309,733	**AGTTTCTGACACCCAGCCAG**	CTTGACGAACATTGGGGCTG	AATAAGCTTGACACACACACTCG	0.20				Alternative PolyA Site
SLC10A7	chr4:147,175,127-147,179,989	GAAGATCGTGTTTGCAGGCC	CGCTCTTGTCAAATACATTTTAAGGC		-0.06				Alternative PolyA Site
SPTAN1	chr9:131,353,756-131,356,652	CGATCGTCAGGGTTTTGTGC	AACCTGCTCCAAGTCTGCTC		0.17				Alternative Promoter
SRGAP2	chr1:206,626,444-206,632,319	**AAGTTTGACTACGTGGGCCG**	TCAGGTGGCTCAGAAATGACC	CCCCTTCTCCCCCACAATTC	0.06				Casette Exon
TRAPPC10	chr21:45,478,984-45,483,666	**TGCAGTACGACGAACTGGAC**	AGAGCTTCAGTTCCTGCACG	GGGCATTTTATAAGTTCACAGCAATG	0.19				Alternative PolyA Site
TSPAN4	chr11:842,808-847,300	GTCAGAGCCAGGTTCAGCTG	**CCACAGGAAAGAGACCAGGC**	CTCAGGCTCTGTGCATCTCC	-0.13				Alternative Promoter
UBAP2L	chr1:154,241,233-154,243,986	**TCATTCCGGTACTCCTGCTG**	GAAGATGGGACCGGGAGAAG	GGAGGCCCAGTTGGTACAG	-0.14				Alternative PolyA Site
ZFYVE28	chr4:2,321,897-2,337,521	AGCTCGGACATGTCTTCCAC	ATTGATGACTACGAGCCGGC	ATGATCTTGGCCCCCTTGTC	0.47				Alternative Promoter
SMN2	chr5:70,241,891-70,248,837	ACCACCTCCCATATGTCCAG	TTTGAAGAAATGAGGCCAGTT		See text				Casette Exon

Other RBMX related sequences		
3' UTR primers (human)		
RBMX 3'UTR F	CCAAGTGGAGGTTCCTACAGAG	Product size (F + xR)
RBMX 3'UTR 1R	AGATCCACAGGAACCCCTGA	236
RBMX 3'UTR 2R	TGAAAAGCAATGCGAAAGTCA	558
RBMX 3'UTR 3R	TAAAAGCAAGCACCACACAGC	130
RBMX 3'UTR 4R	TTCTTTGAACTGGGATTTTGGTC	391
3'UTR primers (mouse)		
Rbmx 3'UTR Fm	GATGGCTATGGTCGGGATCG	Product size (F + xR)
Rbmx 3'UTR 1Rm	AGGAATCACGTGGAGGAGGG	301 (968)
Rbmx 3'UTR 2Rm	TGCCAGGGGTTGAATGCTAG	571
qPCR primers for mouse rbmx (from UCSC - primer pair name: Rbmx_uc009thh.2_4_1_1)		
Rbmx qPCR Fm	CAAGAGGATTCGCTTTTGTCA	Product size
Rbmx qPCR Rm	CCTTGATGGCTTTCCCATCT	99
qPCR Human RBMX (from UCSC - primer pair name: RBMX_uc004fae.2_4_1_1		
RBMX qPCR F	CCTCCAAGAAGTAGAGGCCC	Product size
RBMX qPCR R	TTCATGGAATATCCACCGTCA	116

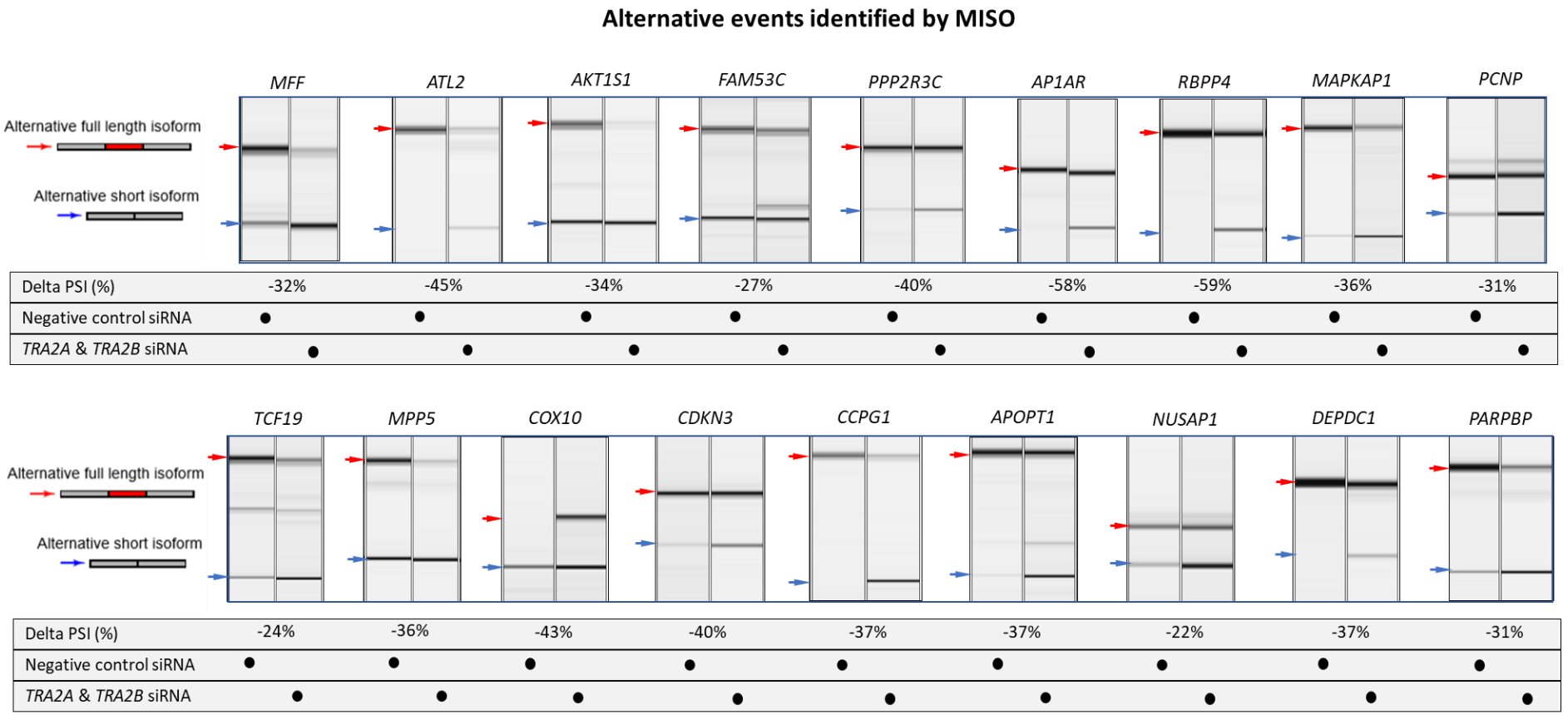

Figure Apx.1. Representative RT-PCR capillary gel electrophoresis images of MISO identified Tra2α/β regulated targets. All changes shown are statistically significant, calculated from three biological replicates. Student's *t*-test, for all events shown $p<0.05$

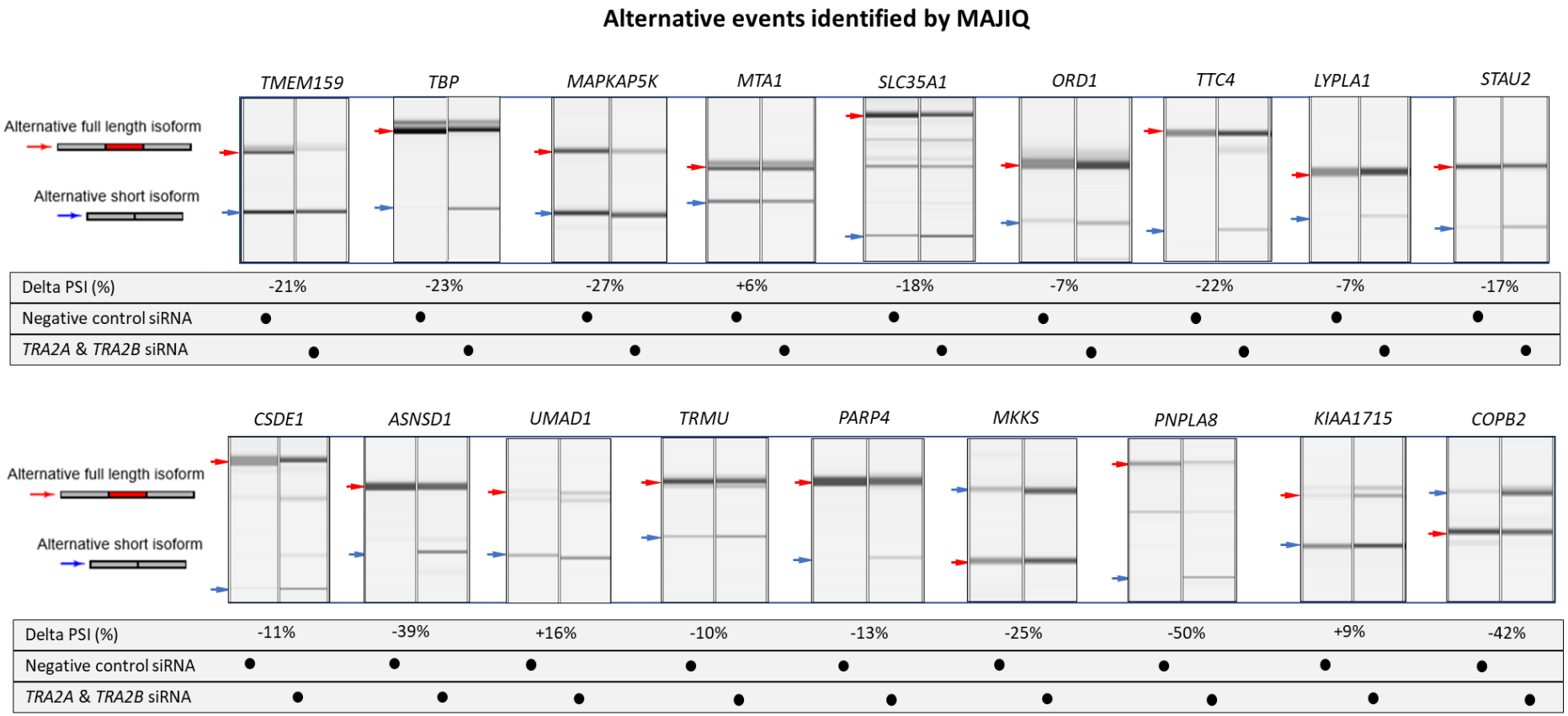

Figure Apx.2. Representative RT-PCR capillary gel electrophoresis images of MAJIQ identified Tra2α/β regulated targets. All changes shown are statistically significant, calculated from three biological replicates. Student's *t*-test, for all events shown $p<0.05$