Identification of RNA binding and processing targets of RBMX protein and their role in maintaining genome stability

Chileleko Siachisumo

Newcastle University Biosciences Institute

Abstract

RBMX is part of a family of RNA binding proteins that regulate pre-mRNA splicing. RBMX is conserved in all vertebrates and has a key role in splicing regulation. RBMX has a homologue called RBMY on the Y chromosome that is sometimes deleted in infertile men. RBMX has several extra-copies originating from retrotransposition (like ELAV in Drosophila) retroposed copies on autosomes include RBMXL2 that represses cryptic splice sites during male meiosis. RBMX has been implicated in cancer development and progression. When the presence of RBMX is altered in cells, it affects the activity of genes involved in DNA damage, suggesting that RBMX acts as a tumour suppressor. However, the precise role of RBMX in RNA processing, including its target RNAs, are unknown. To investigate RBMX-RNA binding I carried out an iCLIP analysis that suggests RBMX directly binds highly to exons and introns of protein-coding genes. RBMX binding was highest in genes involved in DNA damage and replication. To explore the role of RBMX in splicing, it was depleted from MDA-MB231 cells followed by RNAseq analysis. This showed that RBMX blocks selection of aberrant splice sites within some ultra-long exons that would interfere with genes needed for normal replication fork activity. Target exons include within the *ETAA1* (*Ewings Tumour Associated 1*) gene, where RBMX directly binds to exon 5 to enable full-length exon inclusion by blocking selection of an aberrant 3′ splice site. Branch point mapping of the *ETAA1* cryptic exon revealed a branch point closely adjacent to RBMX binding sites. These data suggest that RBMX may act as a steric block to selection of this cryptic ETAA1 branch point by the spliceosome. Preliminary experiments in MDA-MB-231 cells have been used to monitor genome integrity after RBMX depletion. Together our iCLIP RBMX binding data and RNA-seq splicing analysis reveal that RBMX is involved in the repression of long exons. Our data helps explain why RBMX is associated with gene expression networks in cancer, DNA replication and sensitivity to genotoxic drugs.

Covid-19 Impact Statement

In response to the Covid-19 pandemic the university was closed and access to the lab was restricted. This significantly impacted the work I have been able to include in this thesis. The pandemic occurred during my second and third year. I had put a significant amount of effort into establishing a research protocol (iCLIP) but not yet collected the data before the lockdown in March 2020. I was not then able to get my samples for sequencing until February 2021, even though before the lockdown I had optimised iCLIP and was collecting replicate samples for sequencing. Because of the closure I had to terminate my iCLIP experiments. It did take a while to get my cells growing and collect more data for other experiments (all this was working routinely before the shutdown). Once I was ready to get my sequencing done, there was a lag in joining the que for sequencing the genomics facility had a log back log. Without my sequencing data I was unable to progress on this part of my project as a big part of my thesis was to optimise and analyse the iCLIP dataset this set me back. During this time I learnt bioinformatics skills and was analysing external datasets, but was ultimately unable to answer questions central to my research.

Following the lifting of restrictions the space and time I could spend in the lab was limited due to social distancing we were operating on a limited capacity. This made it very difficult for me to make up for lost time. There are 13 members of our research group and due to social distancing restrictions we only had 8 daily spaces available in the lab. We had to work half days and only go in a few days a week to allow for others to do their own work too. This has led me to cut down the amount of work I could do on a daily and weekly basis and delayed my ability to get data for my project for example the DNA damage assays involving immunofluorescence staining and cell cycle analysis using the flow cytometry facilities all shown in chapter 5 of my thesis. I have been unable to complete several of the DNA damage experiments beyond the optimisation stage due to having limited time in the lab.

Declaration

I, Chileleko Siachisumo, 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.

…………………....

Chileleko Siachisumo

November 2022

Dedication

I would like to dedicate this thesis to my husband Daniel Robert Carr, thank you for your being enthusiastic and continuously encouraging me to aim for greatness. Thank you for your enduring patience, understanding and love.

Acknowledgements

I would like to sincerely thank everyone that has been involved in supporting me during this project. I would like to thank so many but I can only acknowledge a few named individuals. I would not have been able to successfully complete this work without the patience and guidance of all those who were involved in this project. I would like to express my gratitude to the Biotechnology and Biological Sciences Research Council (BBSRC) for funding this project.

Firstly, I would like to thank my supervisor and mentor Professor David Elliott for believing in me and allowing me to undertake this PhD. I could not imagine going through this journey with any other supervisor. You have been incredibly patient with me, your immense knowledge and experience have encouraged and supported me throughout my research.

Secondly, I would like to thank all the members of the Elliott and Munkley labs for their support and advice during my years of study. With particular mention to Mrs Caroline Dalgliesh who has been a very good friend, a patient listening ear for all my troubleshooting ideas and for her valuable advice. A special thank you to Dr Gerald Hysenaj whose work set the foundations of my project. Thank you to Dr Ingrid Ehrmann for all the help and advice, for giving me confidence in my ideas. Thank you to Dr Alice Coomer for being a constant pillar of support, you have been with me since the first day of our projects and I am glad we were on this journey together. Thank you to Mr Saad Aldalaqan for being a supportive friend and encouraging me to take care of my mental health. Thank you to Mr Rafiqul Hussain for his expertise on library purification when I was troubleshooting the iCLIP experiment, your advice was invaluable. Thank you to Dr Sara Luzzi for being there at the beginning of my bioinformatics learning journey and being a good sounding board for all my ideas.

Thirdly, I would like to thank my progression panel Professor Mary Herbert and Professor Joris Veltman for their advice and guidance throughout my project. I would like to thank Professor Mary Herbert, your encouragement and hope has given me the motivation to continue to pursue my scientific career.

I would like to thank the many collaborators: Dr Katherine James, Dr Kathleen Cheung, Mr Matthew Gazzara, Professor Yoseph Barash, Dr Graham R Smith, Dr Simon

Table of Contents

List of Figures and Tables

Chapter 4

Chapter 5

List of Abbreviations

ASO	antisense oligonucleotide
BP	branchpoint
ASE	alternatively spliced exon
BPA	branchpoint adenine
BSA	bovine serum albumin
cDNA	complementary DNA
CLIP	UV cross-linking immunoprecipitation
CTD	carboxyl terminal domain
DAPI	4′,6-diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DSB	double strand breaks
DTT	Dithiothreitol
eCLIP-seq	enhanced CLIP sequencing
EDTA	Ethylenediaminetetraacetic acid
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
FBS	Fetal Bovine Serum
FDR	false discovery rate
GFP	Green Fluorescent Protein
GO	gene ontology
GSEA	gene set enrichment analysis
HEK293 cells	human embryonic kidney cells
HITS-CLIP-seq	High-throughput sequencing CLIP sequencing
hnRNP	heterogeneous nuclear ribonucleoproteins
HR	Homologous recombination
HRP	horseradish peroxidase
iCLIP	Individual-nucleotide resolution CLIP
IF	immunofluorescence
IGV	Integrative Genomics Viewer
IP	immunoprecipitation
IR	infrared
irCLIP-seq	infrared CLIP sequencing

ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
lncRNA	long intergenic non-coding RNA
MCF7	Michigan Cancer Foundation-7 human breast cancer cell line
MDA-MB-231	MD Anderson human invasive breast cancer cell line
mRNA	messenger RNA
PAR-CLIP-seq	photoactivatable ribonucleoside-enhanced CLIP sequencing
PBS	phosphate-buffered saline
PCA	principal component analysis
PCR	polymerase chain reaction
pre-mRNA	precursor mRNA
PSI	percentage splicing inclusion
RBMX	RNA Binding Motif Protein X-Linked
RBMXL2	RBMX Like 2
RBP	RNA binding protein
RGG domain	glycine-arginine-rich (GAR) domain
RIP-seq	RNA Immunoprecipitation sequencing
RNA	ribonucleic acid
RNA-seq	RNA-Sequencing
RPMI	Roswell Park Memorial Institute 1640 medium
RRM	RNA recognition motif
RS domain	arginine/serine domain
RT-PCR	reverse transcriptase PCR
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
siRNA	small interfering RNA
SR protein	serine/arginine-rich protein
TBE	Tris-Borate-EDTA
TBS	Tris Buffered Saline
TBST	Tris-Buffered Saline and Tween-20
TE	Tris/EDTA
Tra2β	Transformer-2 protein homolog beta
U2OS	homo sapiens bone osteosarcoma cell
UCSC	University of California, Santa Cruz

UMI	Unique molecular identifier
UTR	Untranslated region
UV	Ultraviolet

Chapter 1. Introduction

1.1   Gene expression and pre-mRNA processing

RNA polymerase II (RNAPolII) initiates transcription of protein coding genes at promoter sites. The binding of RNA polymerase II is controlled by the binding of regulatory proteins at cis-regulatory elements (including promoters, enhancers and silencers). The binding of regulatory proteins changes chromatin structure to either promote or inhibit transcription by RNA polymerase II.

In eukaryotes DNA is transcribed to messenger RNA (mRNA) which is transported out of the nucleus and is translated by the ribosome. There are many intervening steps during this process of transcription to translation of mRNA. The events of pre-mRNA processing take place co-transcriptionally these include: 5′ capping with 7-methyl guanosine; pre-mRNA splicing; and 3′ end formation by cleavage and polyadenylation of mRNA (Elliott & Ladomery, 2016).

As transcription occurs an extra guanosine is added to the first nucleotide at the 5′ end of the nascent pre-mRNA. This guanosine forms a 5′-5′ linkage and is methylated at the 7th nitrogen (N) by a methyl transferase enzyme to form the 5′ 7-methyl guanosine cap (Gonatopoulos-Pournatzis & Cowling, 2014). Human genes are split into introns and exons. Pre-mRNA splicing is the removal of introns and joining together of exons in the mature mRNA transcript this discussed in more detail in section 1.2 of this thesis. The C terminal domain (CTD) of RNA polymerase II is responsible for the recruitment of Cleavage and polyadenylation specificity factor (CPSF) and Cleavage stimulatory factor (CstF) complexes responsible for the cleavage of the 3′ of the mRNA transcripts. Cleavage is followed by polyadenylation (polyA) by polyadenylate polymerase which adds a chain of monophosphate adenosines to the mRNA transcripts (many mRNAs are alternatively polyadenylated) (Lutz, 2008)(Tian et al., 2005). Both 5′ capping and 3′ polyadenylation play an important role in preventing mRNA degradation, promoting nuclear export and promoting translation of mRNA transcripts. Once pre-mRNA is processed mature mRNA it is transported into the cytoplasm where it is bound by ribosomes to initiate translation into protein.

The stability of mRNA transcripts is important for regulating gene expression. mRNA is degraded by ribonucleases or complexes with ribonuclease activity (Elliott & Ladomery, 2016). The essential processes involved in mRNA degradation are de-capping and de-adenylation by de-capping proteins and de-adenylases respectively and cleavage mediated by endonucleases.

1.2   Pre-mRNA splicing

Most human genes are split into exons separated by intronic sequences. Genes are transcribed into pre-mRNA, introns are removed and exons are joined together. Pre-mRNA splicing takes place in the nucleus and is catalysed by the spliceosome. Splicing was first described in adenoviruses in 1977 (Chow et al., 1977). The splicing machinery recognises exon-intron junctions by exon definition (Black, 1995). The selection of splice sites is mediated by RNA-binding proteins at splicing regulatory elements. RNA-binding proteins can act as splicing activators or repressors and influence the choice of adjacent exons being added into the mature transcript.

1.2.1   Exon/intron definition

RNA splicing factors are recruited to nascent pre-mRNAs, including heterogeneous nuclear ribonucleoproteins (hnRNPs) (Suzanne, 2008) and Serine and arginine-rich (SR) proteins (Jeong, 2017). These hnRNP proteins bind in intronic regions to repress assembly of the spliceosome machinery and SR proteins bind within exons to activate the spliceosome machinery.   The spliceosome is a macromolecular enzymatic complex that catalyses the removal of introns via a 2-step transesterification process (Figure 1-1).

Splicing relies on correct identification of exon-intron boundaries by cis sequences recognised by splicing machinery on the precursor transcripts. The cis elements are sequences at the exon-intron boundaries including the 5′ donor and 3′ acceptor sites, as well as the branch site and the polypyrimidine track upstream of the 3′ splice site (Figure 1-1).

Trans-acting elements including RNA binding proteins and components of the spliceosome machinery recognise the cis-splicing elements. Incorrect exon-intron definition by the splicing machinery can lead to the improper inclusion of exons and removal of introns causing alterations to the open reading frame. The definition of exons is well defined for shorter exons.
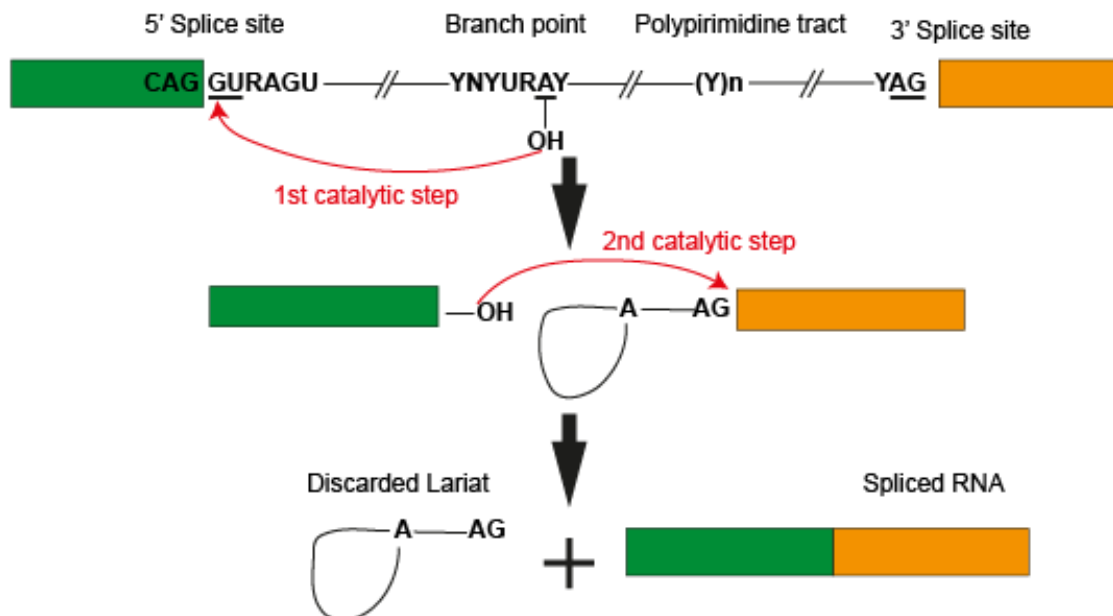
Figure 1-1 The splicing transesterification reaction. The cis-elements that guide intron recognition are; 5' splice site, branch point, polypyrimidine tract and 3' splice site. (R=purine, Y=pyrimidine, N= any nucleotide). There are two catalytic transesterification steps involved in the removal of introns directed by the major spliceosome (Section 1.2.3). In the first catalytic step the branch point adenosine 2'-OH carries out a nucleophilic attack against the 5' splice site of the intron. This forms a 2'-5' phosphodiester bond and a lariat intron. In the second catalytic step the 5' splice site -OH bond attacks the 3' splice site of the intron. This forms a 5'-3' phosphodiester bond releasing the lariat for degradation.

1.2.2 Exon definition

Human exons evolved to be short to enable molecular interactions across the exon between factors bound to the 3′ and 5′ splice sites. This is called exon definition. The median size of all human exons is only 120bp with 70% of all exons shorter than 200bp (Mokry et al., 2010). How long exons are recognised is less well understood. There are many mechanisms to recognize small exons but the mechanism for longer exons may be different. Long exons escape exon definition and may be more likely to

have more cryptic splice sites (splice sites not normally used) and could be particularly susceptible to mis-regulated RNA processing.

1.2.3 Spliceosome Assembly

The splicing of exons is an essential step in the control of gene expression, it is carried out by the spliceosome. the canonical spliceosome, is known as the major spliceosome (Turunen et al., 2013).

The major splicing pathway is also referred to as the U2 dependent splicing pathway. Trans-acting factors interact on the pre-mRNA substrate to form the spliceosome. The spliceosome sequentially forms the H, A, B and C complexes during the splicing cycle (Figure 1-2). There are 5 key Uridine-rich small nuclear ribonucleoproteins (snRNPs) called U1, U2, U4, U5, and U6. U4, U5 and U6 form the U4/U6.U5 tri-snRNP (Turunen et al., 2013). The assembly of the spliceosome begins with the interaction of U1 snRNP to the 5′ splice site of the intron. The branch point is recognised by SF1 (splicing factor 1) which is later displaced in an ATP-dependent manner by U2 snRNA. The interaction of the U2 snRNP component forms the pre spliceosome A complex. The polypyrimidine tract and 3′ splice site are recognised U2AF65 and U2AF35 respectively. The pre catalytic spliceosome (B complex) is formed by the interaction of the U4/U6.U5 tri-snRNP. RNA-protein interactions are rearranged leading to the dissociation of U1 and U4, resulting in the formation of the catalytically active spliceosome (B* complex). Prp2 ATPase catalyses the first transesterification reaction leading to formation of the first lariat structure in the C complex, followed by the second transesterification step which is catalysed by Prp22 and Prp16 ATPases. The conformational changes during splicing are facilitated by evolutionarily conserved DExH/D-box RNA ATPases /helicases and the GTPase Snu114. The principles of splicing and formation of the spliceosome are reviewed further by (Wahl et al., 2009), (Turunen et al., 2013).

A different class of introns are detected by the minor spliceosome. The minor spliceosome splices a 0.5% of introns. The removal of introns by the minor spliceosome is catalysed by U12.
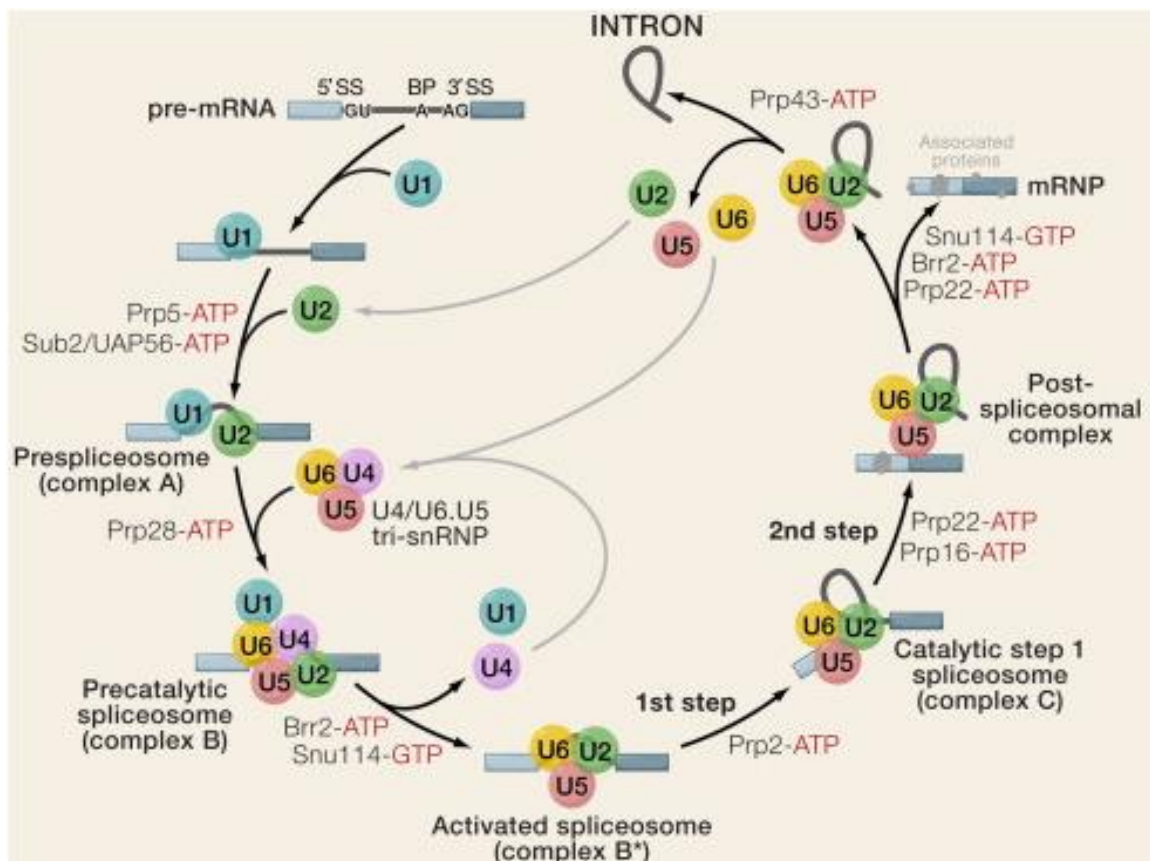
Figure 1-2 Assembly of the major spliceosome. *The interactions of the spliceosomal snRNPs at various stages of the major spliceosome assembly (complexes E, A, B*, and C). This image is taken from* (Wahl et al., 2009)*.*

1.3   Alternative Splicing

The transcripts from most human genes are alternatively spliced. It has been shown by genome-wide analysis of mRNA isoforms that 92–94% of the human genes have alternatively spliced isoforms (Wang et al., 2008).This is a mechanism by which pre-mRNA transcripts are processed to produce mature-mRNAs with different exon combinations. Alternative pre-mRNA splicing is an efficient way of generating proteome diversity and regulating cell homeostasis from a limited number of genes. Alternative splicing generates more than one protein from a single gene expanding the proteome. Alternative splicing is important in development and differentiation (Kalsotra & Cooper, 2011). An impressive example of alternative splicing is *Dscam* which encodes a cell surface protein in *Drosophila melanogaster* that has 38,016 alternatively spliced isoforms generated by mutual exclusion alternative splicing (Hemani & Soller, 2012). As a result of alternative splicing, transcripts produced from a single gene can have different properties such as transcriptional efficiency, mRNA stability (alternative mRNAs can be targeted for nonsense mediated decay) and cellular localisation of

proteins (Black, 2003) (Matlin et al., 2005). The different forms of alternative splicing include constitutive splicing, exon skipping, alternative 5′ and 3′ splice sites, mutually exclusive exons, intron retention, and alternative promoters and poly-A sites (Wang et al., 2015) (Black, 2003). These different kinds of splicing are shown in Figure 1-3. Mutations in exons and introns can affect splicing by disrupting splice site sequences, affecting binding of the splicing machinery, creating new splice sites and also activating cryptic splice sites.
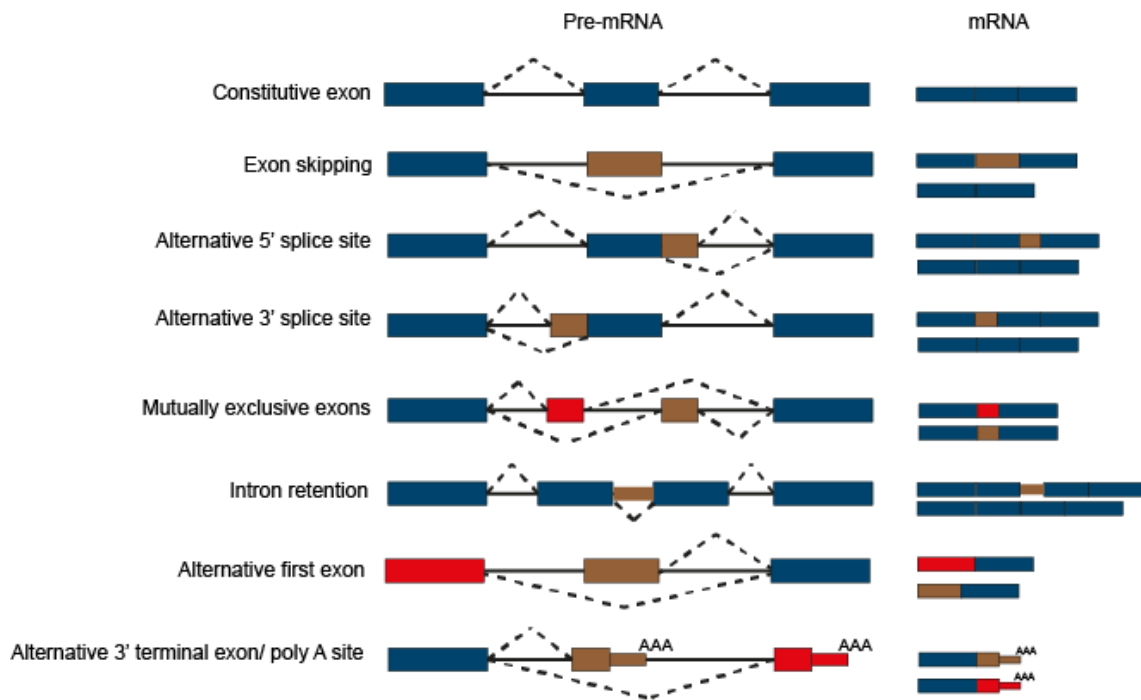
Figure 1-3. Alternative splicing events. Constitutive exons in blue, alternative regions in brown and red and introns as horizontal lines. Events represented include; Constitutive exons – all exons are included into the mRNA transcript, Exon skipping – exons are sometimes skipped and excluded from the mRNA transcript, Alternative 3' and 5' splice site – alternative splice sites within the exons compete for recognition with primary splice sites. Mutually exclusive exons – the process by which one exon from a group exons is included in the mRNA transcript, Intron retention – an intron is included in the mRNA, Alternative first exon – change the 5' untranslated region (5'UTR) in the mRNA, Alternative 3' terminal exon – an alternative poly A sites included in the mRNA. This image is adapted from (Cieply & Carstens, 2015).

1.3.1 Factors that affect alternative splicing

Alternative splicing is highly regulated by cis-acting elements on the pre-mRNA shown in Figure 1-1, and trans-acting RNA-binding proteins involved in the selection of splice sites shown in Figure 1-4. The expression of RNA-binding proteins involved in alternative splicing can vary between tissues, resulting in tissue specific alternative

splicing isoforms. For example RBMXL2 an RNA binding protein has a testis specific expression pattern (Ehrmann et al., 2019). Alternative splicing can also be affected by chromatin structure, RNA structure and transcription. Disruption to the process of alternative splicing can lead to disease. For example changes in splice variants are a hallmark in cancer. Alternative splicing changes are observed in many human cancers and they have been linked to cancer initiation, progression and response to treatment (Skotheim & Nees, 2007). SR protein are involved in splicing activation. Human Transformer2alpha and Transformer2beta (Tra2α and Tra2β) are paralog SR-like proteins that are ubiquitously expressed. They are splicing factors involved in the activation of alternative splicing of exons. Tra2α and Tra2β proteins have a central RRM flanked by 2 RS domains in the N and C terminals. Tra2β is thought to have a key role in mammalian spermatogenesis (Venables et al., 2000). Tra2β has been shown to be involved in the splicing regulation of genes associated with chromosome biology (Best et al., 2014). As mentioned previously Tra2β has been shown to interact with some RBMX family of proteins. As part of this thesis, it was important to know how RBMX and Tra2β work to regulate splicing of target exons.

1.3.2 Cis-acting factors

Exonic/intronic splicing enhancer (ESE, ISE) and exonic/intronic splicing silencers (ESS/ISS) sequences affect splice site selection and function by recruiting RNA-binding proteins (Figure 1-4). These elements work together to promote spliceosome assembly and inhibit recognition of weak splice site (cryptic splices sites). Splicing enhancer elements are particularly important for constitutive splicing whereas the splicing silencer elements are hugely important in alternative splicing (Wang & Burge, 2008).
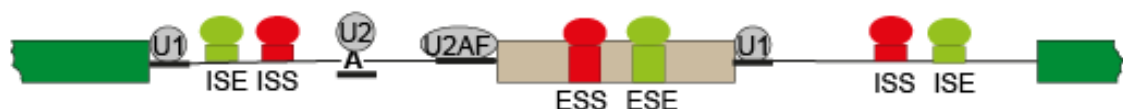
Figure 1-4 Mechanisms of alternative splicing regulation. *Schematic of the splicing regulatory elements important for exon definition. The core splice sites bound to the snRNPs are underlined in bold; the 5' splice site is bound by U1, the branch point is bound by U2 and the polypyrimidine tract and 3' splice site bound by U2AF 65 and 35 respectively. The splicing enhancer and silencers are shown in green and red bound to splicing regulatory proteins. ISE = intron splicing enhancer, ISS = intron splicing silencer, ESE = exon splicing enhancer, ESS = exons splicing silencer. The ISE and*

ESE promote exon splicing. ISS and ESS promote exon skipping. This image is adapted from (Elliott and Ladomery; Cieply and Carstens, 2015). Exons are shown as boxes and introns are shown as lines

1.3.3 Trans-acting factors

Trans-acting splicing factors include RNA-binding proteins that recognise and bind to cis-acting elements called splicing enhancer and silencers sequences that are important for defining splice sites. Serine and arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) represent 2 large families of RNA binding proteins involved in constitutive and alternative splicing (Turunen et al., 2013). Both families of RNA-binding proteins have an RNA recognition motif (RRM) for RNA-protein interactions; and domains for protein-protein interaction (Jeong, 2017). The binding of RNA-binding proteins to cis-acting elements in shown in Figure 1-4. SR proteins are largely known to bind exon splicing enhancers (ESE); and hnRNPs are known as splicing repressors and are largely known to bind to intron splicing silencers (ISS) on mRNA. This thesis will focus on heterogeneous nuclear ribonucleoprotein G (hnRNPG), also more commonly known as RBMX (RNA Binding Motif Protein X-Linked) and the SR protein Tra2β (Transformer 2β).

1.3.4 Cryptic RNA processing sites

Some splice sites are weak and are normally ignored by the spliceosome. Nonetheless, under certain conditions such as the absence or presence of certain RNA binding proteins, these weak splice sites are used, where they are referred to as cryptic splice sites. Cryptic splice sites can involve the inclusion of non-annotated exons. Cryptic splice sites may be found in introns and most commonly larger exons (Sibley et al., 2016). The use of non-annotated splice sites has been discovered by RNA-seq experiments and together with targeted genome editing the function of specific splice sites can be investigated (Sibley et al., 2016). The binding sites of specific RBPs involved in cryptic splicing can be mapped with individual crosslinking immunoprecipitation (iCLIP) experiments (König et al., 2012). RBPs such as RBMXL2 and hnRNP L can repress the use of certain exons by cryptic splice sites to protect the transcriptome (Ehrmann et al., 2019)(McClory et al., 2018). Computer programmes such as cryptic splice site finder can reliably identify cryptic splice sites (Kapustin et al., 2011). Use of cryptic splice sites may lead to disease when not well regulated. As

well as cryptic splice sites, there are also cryptic polyA sites that generate new mRNA 3' ends.

1.3.5 Epigenetic regulation of Alternative splicing

Epigenetic regulation such as chromatic structure, histone modifications and DNA methylation are known to affect alternative splicing (Kim & Kim, 2012). Certain histone modifications are enriched in internal exons compared to introns. H3K36me3, H3K79me1, H2BK5me1, H3K27me1, H3K27me2, and H3K27me3 modifications are enriched in exons compared to introns (Andersson et al., 2009; Li et al., 2018) these modifications are associated with actively transcribed regions. Histone modifications such as H3K36me3 have a role in transcription via splicing since these markings were found in weakly expressed and alternatively spliced exons (Wilhelm et al., 2011; Kolasinska-Zwierz et al., 2009). Signatures of histone modifications have been identified that, when modulated, result in splice site switching. By interactions of splicing regulators with chromatin-binding proteins, these histone marks may have an impact on mRNA splicing (Luco et al., 2010). It has been shown that mis-spliced exons can be distinguished by a startling rise in chromatin accessibility right upstream of the aberrant splice sites and a decline in nucleosome occupancy directly over the exon (Simon et al., 2014). Mechanisms of epigenetic regulation discussed above and many more are associated with alternative splicing. Epigenetic regulation of splicing changes based on cell type and diseases. Examples of how histone modification can affect alternative splicing is shown in Figure 1-5. In diseases such as cancer changes in epigenetic modifications effects aberrant gene expression by alternative splicing (Gimeno-Valiente et al., 2022).
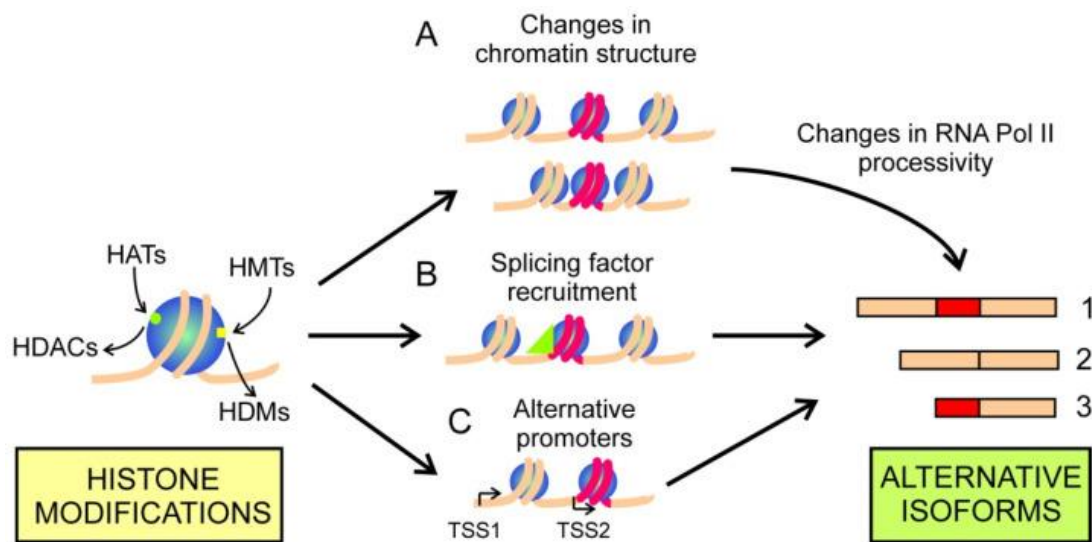
Figure 1-5 Examples of how histones can affect the production of alternatively spliced exons. *(A) Changes in the chromatin structure can affect the activity of RNA polII by affecting its speed resulting in either inclusion of the red exon, in isoform 1 or skipping of the red exon, in isoform 2. (B) Histone modifications may recruit trans-acting splicing factors to either enhance the splicing of alternative exon red, in isoform 1 or repress the inclusion of exon red, in isoform 2. (C) Histone modifications may affect the use of alternative promoters giving either isoform 1 or isoform 3. This image is taken from* (Gimeno-Valiente et al., 2022)*.*

1.4 The RBMX family of RNA-binding proteins

RNA Binding Motif protein, X-linked (*RBMX*) also referred to as *hnRNPG* (heterogeneous nuclear ribonucleoprotein G) is a 43kDa nuclear protein involved in mRNA splicing. *RBMX* is found on the X chromosome and is expressed ubiquitously in all vertebrates. The *RBMX* family of genes encode for proteins that have a well conserved RNA recognition motif (RRM). The RBMX family of proteins are involved in pre-mRNA splicing (Heinrich et al., 2009).

RBMY is a paralogue of *RBMX* located on the Y chromosome and is only expressed in the testes. The *RBMX* and *RBMY* genes were located on the original homologous pair of proto-sex chromosomes (Lingenfelter et al., 2001) (Figure 1-6). RBMY has been implicated in spermatogenesis it is located in the AZFb region and male infertility-causing deletions in the AZFb region are associated with azoospermia (Ma et al., 1993) (Black, 1995).

RBMX and *RBMY* genes have a structure that includes both introns and exons in their sequence. *RBMX* has multiple copies on the human genome these copies are referred to as *RBMX*-like genes. *RBMX-like* genes are intronless in their structure. This suggests that *RBMX-like* genes arose by retrotransposition of *RBMX* cDNAs onto other chromosomes (now human Chr 1, 4, 6, 9, 11, 20, and X). These *RBMXL* retrogenes correspond with retrogenes in primates and great apes suggesting that they evolved before the divergence of humans. The most similar of the *RBMX-like* genes are *RBMXL1* and *RBMXL2* (also known as *hnRNPG-T*). These genes have 95% and 73% sequence homology to *RBMX* respectively (Figure 1-6). With the exception of *RBMXL1* and *RBMXL2* the rest of the RBMX copies are pseudogenes and have no known functions. *RBMXL1* on chromosome 1 is ubiquitously expressed and has a full open reading frame, with few mutations in conserved regions (Lingenfelter et al., 2001). Some *RBMX* retrotransposons may be non-functional, while some of them have developed a tissue specific expression such as *RBMXL2*, which has a testes specific function. During the pachytene stage of meiosis the X and Y chromosomes are sequestered in the XY body so the chromosomes are inactive and the genes cannot be expressed (McCarrey & Thomas, 1987). *RBMXL2* may have a spermatogenesis specific function to replace *RBMX* while the X chromosome is inactive, or it may have developed a new meiosis specific function of its own (Elliott, 2004).
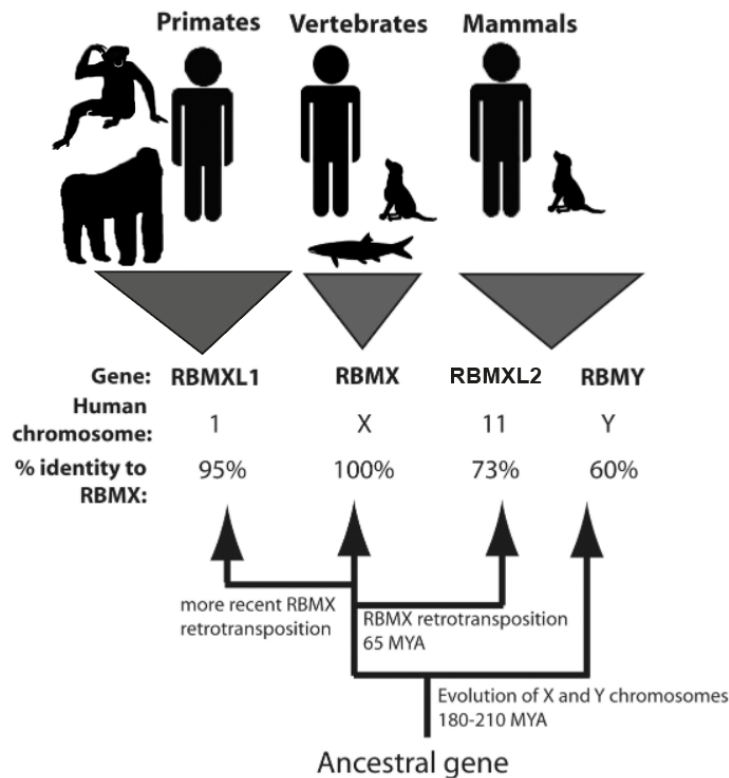
Figure 1-6 Cladogram of the RBMX family of proteins. *The RBMX paralog RBMY has 60% identity to RBMX. The RBMXL1 and RBMXL2 (hnRNPGT) retrogenes have 95% and 73% sequence homology to RBMX respectively. This image was taken from* (Elliott et al., 2019)

1.4.1 RBMX protein structure

RBMX has two distinct RNA binding domains, a highly conserved N-terminal RNA recognition motif (RRM) and a C-terminal RNA binding domain (RBD) (Kanhoush et al., 2010b), shown in Figure 1-7. The RRM is a common domain in RBPs and facilitates the binding to RNA. RBMX has been shown to bind specific RNAs using its RRM in vivo (Heinrich et al., 2009). The carboxyl-terminal referred to as the RGG domain is rich in arginines and glycines, and has also been shown to bind RNA. RBMX missing the C-terminal RBD retains its ability to bind RNA (Heinrich et al., 2009). Both RNA binding domains may play a critical role in binding RNA. The presence of more than one RNA binding domain suggests that RBMX may be involved in binding multiple RNA target sites at the same time. RBMX has been shown to have a role in splicing control, transcription and genome integrity (Elliott et al., 2019). RBMX's precise role in splicing regulation remains unresolved (including what its targets are, and how it regulates them).
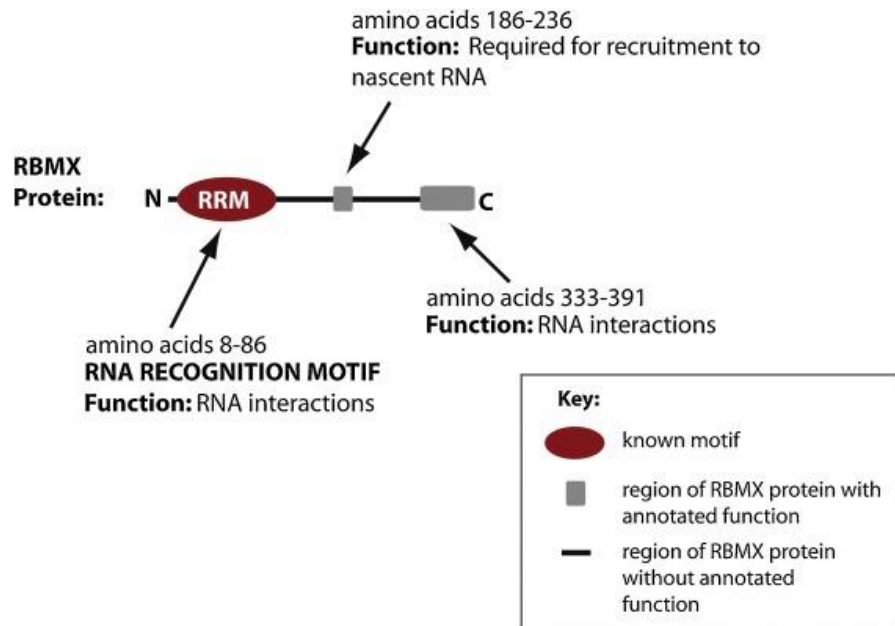
Figure 1-7 Modular structure of the full-length RBMX protein. *RBMX protein has a highly conserved RRM in the N-terminus for RNA recognition. There is a C-term RNA binding domain (RBD) for RNA interactions. This image is taken from a review by* (Elliott et al., 2019)

1.4.2 RBMX as a splicing factor

RBMX has been reported to control splicing transcriptome wide co-transcriptionally on nascent RNA by its association with the phosphorylated carboxyl terminal domain (CTD) of RNA polymerase II (Zhou et al., 2019). RBMX has been shown to promote splicing by reading N6-methyladenosine (m6A) modification patterns within pre-mRNAs to promote exon inclusion (Liu et al., 2017), (Zhou et al., 2019). The m6A modifications are deposited by the N6 methyltransferase complex with is made up of METTL3 and METTL14 methyltransferase-like proteins.

As mentioned before RBMX has an RRM domain that is highly conserved and important for RNA binding. Not much is known about RBMX target RNAs or whether it works by directly binding to targets at an RNA level of whether it works by binding to other RNA binding proteins.

RBMX also directly interacts with another splicing regulator called Tra2β, which is a member of the SR protein-family. It has been reported that sometimes RBMX and

Tra2β have the opposite effect on target exon splicing. Depending on the pre-mRNA target and cellular context both RBMX and Tra2β proteins can either stimulate or suppress exon inclusion. (Nasim et al., 2003). It has also been reported that RBMX can sequester Tra2β preventing its access to RNA targets thereby preventing its activity on splicing (Dreumont et al., 2010), (Liu et al., 2009). The other members of the RBMX family of proteins including RBMXL2 have been shown to interact with Tra2β as well (Venables et al., 2000).

1.4.3 Physiological Function of RBMX

RBMX family proteins may connect the fields of nuclear RNA processing, disease and sex chromosome biology (Elliott et al., 2019). RBMX is expressed ubiquitously in both humans and mice. RBMX expression is not elevated in any tissues. According to the International Mouse Phenotype Consortium RBMX knockout mice have changes in body size/weight, metabolism, renal and neurological systems. It has been shown that reduced expression of SRSF3 and its target RBMX showed significantly better prognosis in patients with head and neck cancer, suggesting they can be used as potential biomarkers to predict favourable overall survival among head and neck cancer patients (Guo et al., 2020). A frameshift in the *RBMX* gene has also been associated with acquired drug resistance in Melanoma Cells, due to an enhanced activity of the PI3K/AKT signalling pathway (Hartman et al., 2020).

RBMX has been shown to regulate gene transcription and may inhibit or activate gene transcription in different cellular environments (Elliott et al., 2019). For example during cellular reprogramming RBMX is enriched within heterochromatin and is associated with H3K9me3, a transcriptionally repressed chromatin marker, to stop or slow down gene transcription (Becker et al., 2017). Furthermore it has been shown that RBMX supresses HIV-1 infection by maintaining the repressive H3K9me3 modification blocking the recruitment of positive transcription factors thus stopping elongation. This contributed to the maintenance of viral latency. (Ma et al., 2020). On the other hand RBMX activates transcription of the sterol regulatory element-binding protein 1c (SREBP-1c) gene by binding to its promoter, in response to a high-fructose diet (Becker et al., 2017), (Takemoto et al., 2007).

However RBMX has also been shown to be involved in other mechanisms outside of gene expression, such as binding directly to replication forks in a non RRM dependent manner. (Matsunaga et al., 2012) have demonstrated that RBMX associates with chromatin and has an essential role in maintaining cohesion of sister chromatids. Depletion of RBMX leads to the delocalisation of the shugoshin complex and outer kinetochore proteins resulting in premature sister chromatid separation (Matsunaga et al., 2012).

RBMX has been shown to protect cells against DNA damage and act as a tumour suppressor in U2OS cells (Adamson et al., 2012). RBMX has also been shown to be important for resistance to DNA damage. RBMX accumulates at DNA lesions in a PARP1-dependent manner and promotes homologous recombination (HR) by facilitation proper expression of BRCA2 in a splicing dependent manner (Adamson et al., 2012).

Mutations of *RBMX* are have been implicated in splicing control during nervous system development within Shashi X – linked intellectual disability syndrome, which includes intellectual disability and facial dysmorphology (Shashi et al., 2015) and in Spinal Muscular Atrophy (SMA) (Sivaramakrishnan et al., 2017). RBMX connections with cancer and intellectual disability could be controlled via RNA processing. Further investigation into the phenotypic consequences of RBMX over-expression/knock-down may indicate the role of RBMX as a tumour suppressor in breast cancer, whilst identification of direct RNA targets may also reveal key pathways regulated by RBMX direct binding via RNA processing.

1.4.4 hnRNP splicing protein RBMXL2

The homology of the *RBMXL2* sequence to *RBMX* is 69% at the nucleotide level and 73% at the amino acid level, whereas other *RBMXLs* have 92–96% homology at the nucleotide sequence level to RBMX. This means that RBMXL2 likely is more ancient than the other *RBMXL* genes, and in fact *RBMXL2* is conserved in all placental mammals. The difference in nucleotide sequence between *RBMX* and *RBMXL2* is possibly because it represents an ancient retrotransposition that has diverged in its sequence. *RBMXL2* has a testes specific expression, which suggests that it has

evolved a testes specific function. RBMX is expressed ubiquitously in both males and females but is subject to X-inactivation (Lingenfelter et al., 2001). RBMXL2 could provide compensatory expression when the X chromosome becomes inactivated during spermatogenesis (Elliott et al., 2000). A conditional knockout of *RBMXL2* in mice leads to sterile male mice (Ehrmann et al., 2019). RBMX and RBMY are thought to regulate pre-mRNA splicing by their interaction with other proteins (Venables et al., 2000). RBMXL2, RBMY and RBMX have been reported to interact with the same SR protein that is a known splicing activator Tra2β. RBMXL2 has a suggested role as a splicing inhibitor. It has been shown that RBMXL2 is required for the prevention of cryptic splicing in mouse testes (Ehrmann et al., 2019). If RBMX is binding directly to RNA targets to regulate splicing it could be binding at branch points or at other splicing sequences to prevent cryptic splice sites form being used similar to RBMXL2.

1.5 Techniques to investigate RNA targets of RNA binding proteins

Splicing regulation is a very complex process. Many techniques to study splicing have been developed over the years. These include spicing assays using minigenes. However some high-throughput techniques such as RNA-sequencing and UV cross-linking and immunoprecipitation (CLIP) have been developed which have facilitated the study of splicing and RNA binding proteins involved in splicing.

1.5.1 Minigenes

The regulation of splicing can be studied *in vivo* using minigenes (JMardon et al., 1987). The construction of minigenes involves the cloning of alternatively spliced exons and their flanking intronic regions that contain the cis-acting sequence elements from genomic DNA. These genomic sequences are then cloned into an expression plasmid that has a eukaryotic promoter to drive transcription of the exon enabling analysis of mRNA splicing. The splice sites in the minigene can be studied by causing mutations that affect the recognition of cis-elements such as mutations or introduction of new sequences (Anna & Monika, 2018).

1.5.2 RNA-sequencing

The development of RNA-sequencing (RNA-seq) techniques has permitted transcriptome-wide investigation of alternative splicing. RNA-seq is a highly sensitive tool that was designed to allow deep sequencing of complete transcriptomes. RNA-seq has been used to simultaneously study differential gene expression and differential alternatively spliced isoforms of mRNAs (Stark et al., 2019). The first steps of RNA-seq involve the collection of RNA, which is then reverse transcribed into cDNA. Sequencing adaptors are added to the cDNA fragments which are then sequenced on a high-throughput platform such as Illumina. Computational steps are used to map the sequencing reads to the genome in order to generate an expression profile for each gene. The standard type of RNA-seq technologies is short-read sequencing which commonly generates fragmented mRNA reads between 100bp and 300bp. Emerging technologies are able to perform long-read sequencing which can generate reads up to 1000bp that can represent full mRNA transcripts. RNA-seq has been used to study the alternative splicing, alternative polyadenylation and gene expression profiles of human genes in 15 diverse tissues (Wang et al., 2008).

1.5.3 UV cross-linking and immunoprecipitation (CLIP)

UV cross-linking and immunoprecipitation (CLIP) is performed to investigate the global direct binding targets of RNA binding proteins. Versions of CLIP have been previously used on several RNA-binding proteins to generate transcriptome-wide binding maps (Yeo et al., 2009; Ule et al., 2005, 2003; Licatalosi et al., 2008). CLIP relies on immunoprecipitation of UV cross-linked protein-RNA fragments. The RNA fragments are purified reverse transcribed and sent for sequencing. Unlike other versions of CLIP that relied on clusters to identify protein binding individual UV cross-linking and immunoprecipitation (iCLIP) identifies RNA-binding at a single-nucleotide resolution. This is achieved by taking advantage of the fact that reverse transcription of RNA terminates at the covalently attached amino acid position (König et al., 2010).

A very large data set is generated from iCLIP experiments. There are several ways of approaching the data analysis using various bioinformatics tools such as those outlined by (Busch et al., 2020). iCLIP data has also been used to identify crosslinks of endogenous, untagged spliceosomal factors on pre-mRNAs. iCLIP of spliceosomal factors (Briese et al., 2019) has been used to map the crosslinking profiles of RBP that

are known to participate in splicing control. (Briese et al., 2019) have performed an iCLIP on SmB/B spliceosomal protein to map branch points (BPs) transcriptome-wide and spliceosomal interactions with pre-mRNAs.

1.6 Background to the thesis

The lab has previously published work on RBMXL2 a paralogue of RBMX. This published work showed that RBMXL2 is functionally important in meiosis. RBMXL2 is expressed during meiosis and just after meiosis (Ehrmann et al., 2008). RBMXL2 protein may provide a like for like replacement for RBMX during meiosis, while RBMX and RBMY are inactivated within the heterochromatic XY body (Wang, 2004). Work by (Ehrmann et al., 2019) showed that RBMXL2 protein was important for meiosis, and protecting the meiotic transcriptome from the use of cryptic splice sites in male mice. The presence of RBMXL2 is needed to prevent male infertility in male mice. Given the high protein similarity between RBMX and RBMXL2 we wanted to investigate if RBMX carried out a similar role to RBMXL2 in protecting the transcriptome from the use of cryptic splice sites.

The work in this thesis is a continuation of the work started by Dr Gerald Hysenaj to identify RBMX splicing targets. Dr Hysenaj performed RNAseq in MDA-MB-231 cells treated with negative control siRNA and siRNA targeting *RBMX*. MDA-MB-231 cells are a human epithelial cell line. The cells are derived from the breast tissue (mammary gland) of a 51 year old white female. In order to investigate the expression of different isoforms, Dr Hysenaj carried out splicing analysis using MAJIQ on single replicates and identified some splicing targets. The tools to identify splicing targets are much stronger when analysed using multiple biologically independent RNAseq replicates. In this thesis I have further analysed the RNAseq data using SUPPA2 with replicated control and the RBMX knockdown samples provided by Dr Hysenaj.

1.7   Research aims and objectives

The hypothesis of this thesis is that RBMX regulates RNA processing by binding directly to RNA sequences.

1.7.1   Aim 1. Identification of direct targets of RBMX using iCLIP

An early objective was to identify RBBMX-RNA interactions. In order to investigate RBMX-RNA interactions I used the individual cross-linking and immunoprecipitation (iCLIP) technique developed by (König et al., 2010). I used iCLIP to map the transcriptome wide binding sites of RBMX in HEK293 flip-in RBMX expression cell line. The iCLIP method was used to crosslink RBMX-RNA interactions in vivo. The RBMX-RNA complexes were immunoprecipitated using a specific antibody, the cross-linked RBMX-RNA complexes were purified and the RNA, reverse transcribed to create a cDNA library. I was able to map novel RNA targets of RBMX on a transcription wide scale. Higher-resolution positional data produced by iCLIP revealed further information on the mechanisms of splicing regulation of RBMX protein.

1.7.2   Aim 2. Transcriptome analysis of mRNA seq data

RBMX is important for maintaining genome integrity and we hypothesised that RBMX would regulate RNA processing of genes with roles in maintaining genome integrity. Therefore my aim was to identify these RNA processing events and determine if RBMX directly binds to these targets using the iCLIP data.

Prior to the start of my project Dr Hysenaj performed RNA-seq analysis on negative control siRNA and siRNA RBMX depleted cells in MDA-MB-231 breast cancer cells. An aim of my project was to identify target RNA processing events regulated by RBMX using SUPPA2 and visual inspection of junction reads on IGV (Integrative Genomics Viewer). The RNA-seq BAM files were loaded onto IGV software for visualisation and analysis of expression and splicing. Using RNA-seq data from RBMX knockdown and control in MDA-MB-231 breast cancer cells 611 differential RNA processing events were identified.

The RBMX iCLIP data taken together with the RBMX RNA-seq data would give us an understanding of how RBMX regulates splicing. The objective was to identify target exons directly bound by RBMX and functionally responsive to RBMX depletion.

I set out to validate and characterize splicing of an RBMX target exon using minigenes. This was done under different conditions with other splicing factors such as Tra2β and RBMXL2. The minigene experiments were used to test the role of cis-acting elements on splicing regulation, if they increase splicing efficiency or cause the activation of the alternative cryptic splicing sites.

I also investigated if RBMX binding at branchpoints was required for splicing regulation of RBMX splicing and binding target ETAA1.

1.7.3 Aim 3. Biological effect of RBMX knockdown on genome integrity

RBMX is involved in maintaining genome integrity. I thus set out to investigate the role of RBMX in maintaining genome integrity in MDA-MB-231 cells. I hypothesized that depletion of RBMX would cause increased DNA damage and replication defects causing the cells to halt during the cell cycle. I also hypothesized that RBMX depletion would sensitise cells to treatment of genotoxic drugs.

Chapter 2 Materials and Methods

2.1 Cell culture

MDA-MB-231, NCI-H520, MCF7 cells were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium (Life Technologies). HEK293 cells were cultured in Dulbecco′s Modified Eagle′s Medium (DMEM) high glucose pyruvate medium (Life Technologies), supplemented with 10 % fetal bovine serum (FBS) (Life Technologies), 1 % penicillin/streptomycin (Life Technologies). Cells were incubated at 5% CO_2 and 37°C grown in 6 well plates for siRNA KD and minigene assays.

2.2 SiRNA and ASO transfection

Cells were seeded at 100,000 cells per well in 6-well plates and forward transfected with Lipofectamine® RNAiMAX transfection reagent (Invitrogen) according to manufacturer′s instructions. Briefly, a mixture of 150 µl of opti-MEM medium and 9 µl of Lipofectamine® RNAiMAX Reagent was added to 30 pmol of siRNA/ASO and incubated at room temperature for 5 min prior to addition. The cells were harvested 72hrs after transfection. The siRNAs used to knock down RBMX and ASO for ETAA1 were ordered from Integrated DNA Technologies (IDT). See ***Table 2-1*** for siRNA and ASO sequences.

Table 2-1 RBMX siRNA and ETAA1 antisense oligonucleotide sequences used to create the knockdown and block RBMX binding respectively.

RBMX siRNA sequences	
hs.Ri.RBMX. 13.1-SEQ1	5′- rArUrC rArArG rArGrG rArUrA rUrArG rCrGrA rUrArG rArGA T -3′
hs.Ri.RBMX. 13.1-SEQ2	5′- rArUrC rUrCrU rArUrC rGrCrU rArUrA rUrCrC rUrCrU rUrGrA rUrGrG -3′
ETAA1 antisense oligonucleotides (ASOs)	
ETAA1 Exon 5 antisense oligonucleoti des (ASOs)	5′- mC*mU*mA*mA*mA*mA*mG*mU*mU*mU*mC*mA*mA*mA*mU*mG*mU*mA*mA*mA*mU*mG -3′

2.3   Western immunoblotting

Whole cell proteins were used. Cell pellets were resuspended in 2x sodium dodecyl sulfate (SDS) loading dye. The cell were sonicated and heated to 95°C for 5 minutes. Whole cell proteins were used for phosphorylated protein detection, where protein was collected using PhosphoSafe™ Extraction Reagent (MerckMillipore) using manufacturer's instructions. The lysates were diluted in 6x sodium dodecyl sulfate (SDS) loading dye. The protein samples were electrophoresed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), then transferred from the gel to a nitrocellulose membrane, incubated in 5% Bovine Serum Albumin Fraction V (Merk-Roche) blocking buffer for the phosphorylated proteins and 5% milk for all other membranes. The membranes were stained with primary antibodies (at concentrations indicated below) diluted in blocking buffer overnight at 4°C. After incubation, the membranes were washed three times with TBS-T. Secondary antibodies conjugated with horseradish peroxidase were used to detect primary antibody binding. Antibody binding was detected using the Clarity Western ECL Substrate (GE Healthcare Systems), and developed using medical X-ray film blue film in an X-ray film processor developer machine. The phosphorylated protein membranes were stained with Licor secondary antibodies as indicated.

The following primary antibodies were used: anti-RBMX (Cell Signalling, D7C2V) 1:1000; anti-SGO2 (Bethyl Laboratories, A301-262A, A301-261A) 1:1000; anti-ETAA1 (Sigma, HPA035048) 1:1000; anti-Tubulin (Abcam, ab18251) 1:2000, anti-beta actin antibody (Proteintech, 60008-1) 1:1000; anti-FLAG antibody (Sigma, F1804) 1:1000; anti-phospho Chk1 (S345) (Cell Signalling,133D3) 1:500; anti-Chk1 total (Cell Signalling, 2G1D5) 1:500; anti- phospho RPA32 (Bethyl Laboratories,A300-245A) 1:1000; anti-RPA32 total (Bethyl Laboratories, A-300-244AT) 1:1000; anti-RAD51 (Protein Tech, 14961-1AP) 1:1000; anti-H2AX total (cell signalling, 2595) 1:1000; anti-phospho H2AX (S139) (Merck Millipore, 05-636) 1:1000.

The following secondary antibodies were used: Anti-mouse HRP (Jackson ImmunoResearch Labs 115-035-003) 1:2000, Anti-rabbit HRP (Jackson ImmunoResearch Labs 111-035-003) 1:2000; IRDye® 800CW Goat anti-Mouse (Licor) 1:10,000; and IRDye® 680RD Goat anti-Rabbit (Licor) 1:10,000.

2.4 Flow cytometry

Cells were washed in PBS and fixed in 70% ethanol. The cells were permeabilised using 0.1% Triton-X in PBS, stained with 1:100 cy5 coupled MPM2 (Merckmillipore) treated with RNAse A (0.2 mg/ml) (Thermo Scientific) and stained with propidium iodide (50 µg/ml) (Invitrogen) 20 minutes before analysis. Samples of untreated or treated MDA-MB-231 cell cultures were analyzed for DNA content/cell cycle by flow cytometry on a BD LSRFortessa™ cell analyzer. Cell cycle distribution was calculated after appropriate gating of cell populations in FL-2-Area vs FL-2-Width plot of PI fluorescence as shown in Figure 2-1. Assays were carried out in triplicates, and the results were representative of three independent experiments.
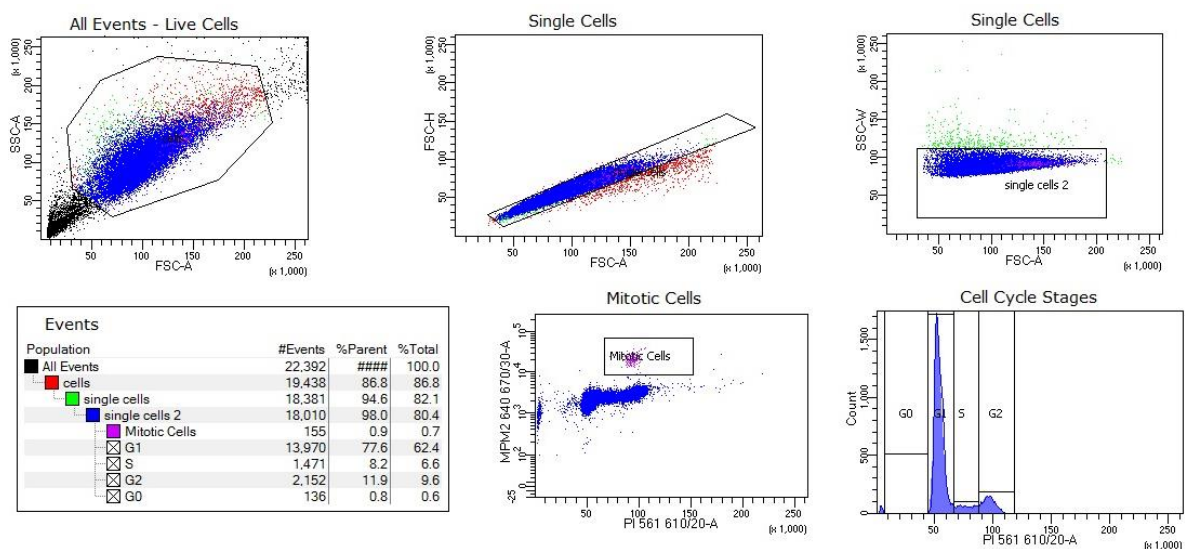
Figure 2-1 Flow cytometry cell cycle analysis gating method used for the selection of live cells, single cells and mitotic cells in MDA-MB-231 cells.

2.5 Minigene splicing assay

The ETAA1 exon 5 alternatively spliced exon and flanking intron sequences were PCR-amplified from human genomic DNA using the cloning primers shown in Table 2-2. The PCR products were digested with *Mfe*I (Promega) restriction enzyme and cloned into the pXJ41 vector using the *Mfe*1 site midway through the 757 nucleotide β-globin intron (Bourgeois et al., 1999). Insertion of the correct amplicons was confirmed by sanger sequencing (Source Bioscience).The splicing analysis was carried out in HEK293 cells. Lipofectamine 2000 (Invitrogen) was used for the transfection of plasmids following the standard manufactures protocol. RNA was extracted with TRIzol (Invitrogen), RNA concentration was quantified by NanoDrop UV-

Vis spectrophotometer, and analysed using a one-step RT–PCR (PCR with reverse transcription) kit from Qiagen, both using the standard protocol. RT–PCR experiments used 100 ng of RNA in a 5-µl reaction, using primers within the β-globin exons of pXJ41. The primers used for the RT-PCR are shown in Table 2-2. Reactions were analysed and quantified by Qiaxcel capillary electrophoresis system (Qiagen). PSI was calculated using the following formula:

$$\frac{Concentraction\ of\ exon\ included\ (ng/ul)}{Concentration\ of\ Included + excluded\ \ exons\ (ng/ul)}\ X\ 100 = PSI(\%)$$

Table 2-2 Primer sequences used for amplification of ETAA1 and flanking intronic regions from genomic DNA into the pXJ41 minigene. *MfeI restriction sites and 9 adenosine (A) sequences were added to primers for amplification of ETAA1 from genomic DNA (upper table). Primer sequences used for splicing analysis of minigenes (lower table). Primers were designed using Primer 3 Plus and the UCSC In-Silico PCR tool was used to predict the PCR products.*

Minigene cloning primer sequences	
ETAA1 short minigene Forward	AAAAAAAAACAATTG GAACATGGAGCCAAACTAACTC
ETAA1 short minigene Reverse	AAAAAAAAACAATTG TGATAGAATGGAGACTTGGGGA
ETAA1 medium minigene Forward	AAAAAAAAACAATTGGAACATGGAGCCAAACTAAC TC
ETAA1 medium minigene Reverse	AAAAAAAAACAATTGTGATAGAATGGAGACTTGGG GA
ETAA1 long minigene Forward	AAAAAAAAACAATTGAGTTAAGACTTTTCAGCTTTT CTGA
ETAA1 long minigene Reverse	AAAAAAAAACAATTGAGTGCTGGGAAAGAATTCAA TGT
Minigene RT-PCR primer sequences	
pXJRTF	GCTCCGGATCGATCCTGAGAACT
pXJB	GCTGCAATAAACAAGTTCTGCT
ETAA1 internal primer	CTTGGCTTAACTGTCCGCTAC

2.6 Splicing analysis of endogenous target exons

RNA was extracted from cells using TRIzol (Life Technologies) following manufacturer's instructions and RNA concentration was quantified by NanoDrop UV-Vis spectrophotometer. In each case cDNA was synthesized from 500ng total RNA in 10µl reactions using Superscript VILO cDNA synthesis kit (Invitrogen), following manufacturer's instructions. Primers to analyse the splicing control were designed using Primer 3 Plus and the UCSC In-Silico PCR tool was used to predict the size of expected PCR products (Full list of primers is seen in Table 2-3 below). 2µl of cDNA was used as template in the end point PCR using GoTaq® G2 DNA Polymerase (Promega) in the standard 25ul reaction following the manufacturer's instructions. PCR products were examined using the Qiaxcel capillary electrophoresis system (Qiagen) and PSI was calculated using the following formula:

$$\frac{Concentraction\ of\ exon\ included\ (ng/ul)}{Concentration\ of\ Included + excluded\ exons\ (ng/ul)}\ X\ 100 = PSI(\%)$$

Table 2-3 Full list of PCR primer sequences used for analysis of RBMX regulated exons.

PCR primer sequences			
Gene	Forward primer	Reverse primer	Internal primer
ETAA1	GCTGGACATGTGGATTGGTG	GTGGGAGCTGCATTTACAGATG	GTGCTCCAAAAAGCCTCTGG
REV3L	TCACTGTGCAGAAATACCCAC	AGGCCACGTCTACAAGTTCA	ACATGGGAAGAAAGGGCACT

2.7 Branch point mapping

Using the RNA from the long minigene transfections with *RBMX* and *RBMXΔRRM*, total RNA was extraction using TRIZOL reagent (Life Technologies), RNA concentration was quantified by NanoDrop UV-Vis spectrophotometer and treated with DNase I (Thermo Scientific). 1µg of purified RNA was reverse transcribed with a reverse transcription cDNA synthesis kit, SuperScript™ III Reverse Transcriptase (Thermo Fisher) using ETAA1 DBR R1 RT-PCR primer (**Table 2-4**) and treated with Rnase H. 1µl of the cDNA was used for PCR amplification reactions using using

GoTaq® G2 DNA Polymerase (Promega) in the standard 25ul reaction following the manufacturer's instructions. PCR amplification was carried out using 2 different primer sets ETAA1 DBR R2 and ETAA1 DBR F1, ETAA1 DBR R2 and ETAA1 DBR F2 with 18 cycles of amplification at an annealing temperature of 56°C. PCR products were sub cloned into the pGEM-T-Easy vector (Promega) using manufactures instructions. Briefly PCR amplicons were ligated into pGEM-T-Easy vector, ligation reaction was transformed into DH5α competent E. coli, the transformation mix was grown on LB Agar Ampicillin-100, X-Gal, IPTG plate overnight. Plasmid DNA was extracted from randomly selected white E. coli colonies using the QIAprep Spin Miniprep Kit (QIAGEN). The expected insert size was confirmed using gel electrophoresis. Insert-containing plasmids were Sanger sequenced (Source BioScience) and the sequences were checked with bio edit and aligned to the genome with UCSC.

Table 2-4 Full list of PCR primer sequences used for branchpoint analysis of ETAA1 exon 5.

Branchpoint analysis primers	
ETAA1 DBR R1	AAGTTCTTCTTCTTGACTTTGTGTT
ETAA1 DBR R2	GCTCTTGAATCACATCTAGCTCT
ETAA1 DBR F1	AGCCAAACTAACTCAGCAACA
ETAA1 DBR F2	AGCATTTGAATCCAGGCAGC

2.8 Over expression of RBMX protein using stable cell lines

Flip-In HEK293 cells over expressing FLAG-RBMX (Kheirollahi Kouhestani & Kouhestani, 2013) cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) high glucose pyruvate medium (Life Technologies), supplemented with 10 % fetal bovine serum (FBS) (Life Technologies). Cells were incubated at 5% CO_2 and 37°C grown in 10cm plates for iCLIP and 6 well plates for analysis of endogenous splicing targets and treated with 1μg/ml tetracycline (Sigma-Aldrich) to express the FLAG-RBMX. Cells made with an empty pcDNA5 vector were used as negative control.

2.9 Individual nucleotide resolution CLIP (iCLIP)

Direct targets of RBMX in HEK293 cells were identified using the following iCLIP protocol developed by the Ule Lab (unpublished). This protocol takes aspects from the

(Huppertz et al., 2014) iCLIP protocol, with some changes introduced bringing features from the eCLIP (Van Nostrand et al., 2016) and irCLIP method (Brian J. Zarnegar et al., 2016).

2.9.1 UV Cross-linking

Flip-In HEK293 cells were grown in 10cm dishes to approximately 80% confluence. The media was removed, cells were placed in 4°C cold 1xPBS (Gibco) and irradiated once with 400 mJ/cm2 in a Stratalinker 2400 at 254 nm. Cells were harvested by scraping, using cell lifters into 2 ml microtubes, centrifuged at 514 g at 4°C for 1 min to pellet cells, and then remove supernatant. Pellets were snap frozen on dry ice and stored at -80°C until use.

2.9.2 Immunoprecipitation

2.9.2.1 Bead preparation

100 µl of protein G Dynabeads (Invitrogen) were washed 2x with lysis buffer (Table 2-6) (all washes throughout the protocol are performed with 900 µl). Beads were resuspended in 100 µl lysis buffer (**Table 2-6**) with 5 µg anti-flag antibody (Sigma F1804) per experiment, incubated at room temperature for 60 minutes, washed 3x with 900 µl lysis buffer and left in 400 µl lysis buffer (**Table 2-6**) until the lysate was ready.

2.9.2.2 In-lysate partial RNA digestion and centrifugation

The cell pellet with the UV cross-linked cells was resuspend in 1 ml lysis buffer (**Table 2-6**) supplemented with protease inhibitors, transferred to a new 1.5 mL tube and sonicated using a Bioruptor for 10 cycles with alternating 30 secs on/ off at low intensity. The lysate concentration was measured with Bradford assay and all samples diluted to 1 mg/ml, or to the concentration of the least concentrated sample (whichever is less). Cell lysates at 1mg/ml protein concentration were treated with 2 µl Turbo DNase (Ambion, AM2238) and with either 0.8 units/ml lysate RNase I (Thermo Scientific, EN0602) or 2.5 units/ml RNase I (Thermo Scientific, EN0602) for the low RNase and high RNase conditions respectively. The cell lysates were incubated at 37°C shaking at 1100 rpm the low RNase was incubated for 3 min and high RNase was incubated for 5 min and transferred to ice for >3 min. Lysates were centrifuged at 15000 rpm at 4°C for 10 minutes, and the supernatant transferred to a new 1.5ml tube.

Lysates were transferred to Proteus Clarification Mini Spin Column (Generon; GEN-MSF500), and spun at 4°C at 13.000 rpm for 1 min (500 µl at a time)

2.9.2.3 Immunoprecipitation

Cell extracts were added to the beads from step 2.9.2.2. The beads/lysate mix was rotated overnight at 4°C, in the cold room, washed 2x with high-salt wash (**Table 2-6**) (the second wash was rotated for at least 1min in the cold room). Beads were washed 1x with PNK buffer (**Table 2-6**), resuspend in 1X PNK wash buffer (Table 2-6) and moved to a new tube.

2.9.2.4 3' end RNA dephosphorylation

Beads were resuspended in 40µl dephosphorylation mix (8 µl 5x PNK pH 6.5 buffer, 1 µl PNK (NEB M0201L), 0.5 µl FastAP alkaline phosphatase (ThermoFisher, EF0654), 0.5 µl RNasin, 30 µl water), and incubated for 40 min at 37°C in thermomixer at 1100rpm.

2.9.2.5 3' adaptor ligation

Beads were washed 1x with 1x Ligation Buffer (DTT free) (containing 6.3µl water, 3µl 10X ligation buffer (no DTT) , 0.8µl 100% DMSO, 2.5µl T4 RNA ligase I – high concentration (M0437M NEB), 0.4µl RNasin, 0.5µl PNK (NEB M0201L), 2.5µl pre-adenylated adaptor L3-???-App (see Table 2-5 for sequence) and 9µl 50% PEG8000). Beads were resuspended in 25µl of the ligation reaction, incubated for 75min at RT flicking the tube every 10 minutes. The beads were washed 2x with high-salt wash (Table 2-6) (rotate wash for at least 1 min in cold room), 1x with 1 ml PNK buffer and resuspended in 1ml of PNK buffer.

2.9.2.6 Adaptor Removal

The beads were moved to a new tube, placed on a magnet to remove the buffer and 20µl of removal mix (containing 12.5µl Nuclease-free H2O, 2µl NEB Buffer 2, 0.5µl 5' Deadenylase (NEB M0331S), 0.5µl RecJf endonuclease (NEB M0264S), 0.5µl RNasin and 4µl 50% PEG8000) were added. The beads were incubated for 1hr at 30°C, then for 30mins at 37°C whilst shaking at 1100rpm. The beads were washed 2x with high-

salt wash (Table 2-6) (rotate wash for at least 1 min in cold room) and 1x with 1 ml PNK buffer.

2.9.3 SDS-PAGE and nitrocellulose transfer

The PNK buffer was removed and the beads were resuspended in 20µl of preheated 1x NuPAGE loading buffer (prepared by mixing 4x stock; Invitrogen, with water and reducing agent DTT; final conc. 100mM (from 1M stock = 10%) at 70°C. The samples 1x NuPAGE loading buffer incubated shaking at 70°C for 1 min, then immediately placed on the magnet to collect the supernatant to load on the gel.

The supernatant was loaded on a 4-12% NuPAGE Bis-Tris gel (Invitrogen) with 6 µl of a pre-stained protein size marker according to the manufacturer's instructions. The gel was run with 0.5L 1 x MOPS running buffer (Invitrogen) for 65 min at 180 V. The dye front was cut off the gel and discarded (the dye front interferes with downstream imaging). After size separation the protein-RNA-L3complexes were transferred to from the gel to a Protan BA85 Nitrocellulose Membrane (Whatman) using the Novex wet transfer apparatus according to the manufacturer's instructions (Invitrogen) for 2 hrs at 30 V. The membrane was rinsed in PBS buffer the wrapped in a clear document sleeve and imaged on the Odyssey LI-COR CLx imager scanning in both the 700nm and 800nm channels.

2.9.4 RNA isolation

A print out template of the infrared gel was used to cut out the low-RNase samples from the membrane. The membrane slices were put in a 1.5ml (Eppendorf Test tube DNA LoBind 1.5mL) microfuge tube with 10µl proteinase K (Roche, 03115828001) per 190 µl PK+SDS buffer (Table 2-6), incubated shaking at 1100 rpm for 60 min at 50°C. The sample plus 200 µl Phenol:Chloroform:Isoamyl Alcohol (Sigma P3803) were added to a separate tube mixed and added to pre-spun 2 ml Phase Lock Gel Heavy tube (713-2536, VWR), incubated for 5 min at 30°C shaking at 1100 rpm and spun for 5 min at 13000 rpm at room temperature. 800ul of Chloroform was added to the top phase of the Phase Lock Gel Heavy tubes, inverted to mix spun for 5 min at 13000 rpm at 4°C. The aqueous layer was transferred into a new tube and mixed with 0.75 µl

glycoblue (Ambion, 9510), 20 µl 5 M sodium chloride and 500µl 100% ethanol precipitated overnight at -20°C.

2.9.5 Reverse transcription

The precipitated RNA was resuspend in 5.5 µl and reverse transcribed in 1 µl primer irCLIP_ddRT_## (1 pmol/µl) (see Table 2-5 for sequence or RT primers with individual barcodes for each reference) and 0.5 µl dNTP mix (10 mM). The RNA samples were incubated at 65°C for 5 min and cooled to 25°C hold until the RT-PCR mix (containing 2 µl 5x SSIV buffer (Invitrogen), 0.5 µl 0.1 M DTT, 0.25 µl RNasin, 0.25 µl Superscript IV) using the thermal program: 25°C for 5 min, 50°C for 5 min, 55°C for 5 min and 4°C hold. The newly transcribed cDNA was treated with 1.25ul of 1M NaOH, incubate at 85°C for 15 minutes for alkaline hydrolysis of the RNA and treated with 1.25ul of 1M HCl to neutralise the pH.

2.9.6 cDNA purification via AMPure XP beads capture

The cDNA was added to 37.5µl (3x) Agencourt AMPure XP beads, 21µl (1.66x) Isopropanol, extensively mixed and incubated for 5 minutes at room temp. The beads were placed on a magnet for 2-3 minutes. The liquid was removed and washed twice with 200ul 85% Ethanol for 30 seconds. The ethanol was removed and the beads left to dry for 1 minute at room temperature. The cDNA was eluted into 9µl of water, incubated for 3 minutes and collected to a new tube.

2.9.7 cDNA elution and circularisation

The purified cDNA was circularised in the following mix (1.5 µl 10x CircLigase Buffer II (Epicentre), 0.75 µl CircLigase II (Epicentre), 0.75 µl 50 mM MnCl2 (provided in the CircLigase II kit), 3 µl 5M betaine (provided in CircLigase II kit)) incubated over night at 60ºC 2 hour. The circularised cDNA was purified using AMPure XP beads as above (step 2.9.6) scaling up the reaction to 45ul of AMPure XP beads, 25ul of isopropanol, and eluted in 10ul of water.

2.9.8 PCR amplification

The circularised cDNA was PCR amplified in the following PCR mix: 4 µl cDNA, 15 µl water, 1 µl primer mix P5Solexa/P3Solexa 10 µM each (Table 2-5) and 20 µl Platinum HS master mix. PCR was performed using the thermal programme 98°C for 40 sec,19 cycles of [98°C for 20 sec, 65°C for 30 sec, 72°C for 45 sec], 72°C for 3 min and hold at 25°C.

2.9.8.1 Purification with AMPure

The amplified cDNA was purified using AMPure XP at twice the volume of the cDNA library and mixed by vortexing for 30 sec. Beads were incubated for 5 minutes at RT, placed on the magnet stand and washed with 500 µl 70% Ethanol. I incubating for 30sec while the tube was still placed on the magnetic stand, took of the ethanol and left to dry for 1 min. I eluted the purified PCR product with a 40 µl TE Buffer and vortex for 30 secs.

2.9.8.2 Size selection of cDNA library by gel purification

The purified PCR product was size separated on a 6% TBE gel, run at 180V for 30 min, stain with SYBR Green I and visualised with blue light transilluminator. DNA in the range of 145-400 nt was excised from the gel. DNA was isolated from the gel by crushing the gel in 500µl of Crush-Soak Gel buffer (Table 2-6). I incubated at 65°C for 2 hours with thermomixer settings of 15s at 1000rpm, 45s rest. The liquid portion of the supernatant was added onto a Costar SpinX column (Corning Incorporated, 8161) into which two 1 cm glass pre-filters (Whatman 1823010) were placed. The sample was spun at 13000 rpm for 1 min, the liquid portion added to 1µl glycoblue, 50µl 3M sodium acetate, pH 5.5, 1.5 ml 100% ethanol and precipitated over night at -20°C.

2.9.8.3 High-throughput sequencing

Prior to sequencing the iCLIP protocol was monitored at two crucial steps, (1) infrared of protein-RNA complexes after membrane transfer and (2) the gel image of the PCR amplified cDNA library on the 6% TBE gel. In addition the samples were analysed on the tape station and submitted for high throughput sequencing on the Illumina NextSeq 500 at the Newcastle University Genomics Core Facility in collaboration with Rafiqul Hussain.

Table 2-5 Adaptor and primer sequences used for iCLIP all sequences. *All sequences were ordered from IDT, 3' infrared adaptor were ordered as HPLC purified DNA with 5' phosphate.*

iCLIP oligos and primers	
L3-IR-phos	/5Phos/AG ATC GGA AGA GCG GTT CAG AAA AAA AAA AAA /iAzideN/AA AAA AAA AAA A/3Bio/
irCLIP_ddRT_33 **RT primer**	/5Phos/ WWW CTGTG NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC
irCLIP_ddRT_34 **RT primer**	/5Phos/ WWW GAAAC NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC
irCLIP_ddRT_35 **RT primer**	/5Phos/ WWW GAGTT NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC
P3 Solexa	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCT GAACCGCTCTTCCGATCT
P5 Solexa	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACG ACGCTCTTCCGATCT

Table 2-6 List of buffer components used in the iCLIP experiment

(A)

Lysis Buffer
50 mM Tris-HCl, pH 7.4
100 mM NaCl
1% Igepal CA-630 (Sigma I8896)
0.1% SDS
0.5% sodium deoxycholate

(B)

High-salt Wash
50 mM Tris-HCl, pH 7.4
1 M NaCl
1 mM EDTA
1% Igepal CA-630 (Sigma I8896)
0.1% SDS
0.5% sodium deoxycholate

(C)

PNK Wash Buffer
20 mM Tris-HCl, pH 7.4
10 mM MgCl2
0.2% Tween-20

(D)

4x Ligation Buffer
200 mM Tris-HCl, pH 7.4
40 mM MgCl2
4 mM dithiothreitol

(E)

10x Ligation Buffer (DTT free)
500mM Tris-HCL 7.5
100mM MgCl2

(F)

PK +SDS Buffer
10 mM Tris-HCl, pH 7.4
100 mM NaCl
1 mM EDTA
0.2% SDS

(G)

Crush-Soak Gel Buffer
500mM NaCl
1mM EDTA
0.05% SDS

2.10 Bioinformatics analysis of iCLIP sequencing data

The iCLIP data was analysed using the iMAPs pipeline to remove PCR-duplicates and aligned to the human genome sequence (version hg38/GRCh37) using STAR. I created a merged file of the crosslink sites from the three replicates. The merged file contains crosslink-sites present in all three replicates. The iCLIP data gives protein binding to a single nucleotide level, so in order to merge the files, 5 nucleotides were added to each site and a single file was using bedtools intersect (Quinlan & Hall, 2010), https://github.com/arq5x/bedtools2 to get regions that were close together. iCLIP binding sites were annotated using the RCAS annotation package which is a suite for RNA datasets using R (Uyar et al., 2017), https://bioconductor.org/packages/RCAS/. Galaxy hub (Afgan et al., 2018), https://usegalaxy.org) toolshed was used to make sample correlation and PCA analysis on the three replicate samples.

2.11 Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed by Dr Kathleen Cheung (Bioinformatics Support Unit, Newcastle University) using the Molecular Signatures Database gene sets (Liberzon et al., 2011). The analysis was done in triplicate sets of RNA-seq data of RBMX KD v control provided by previous work done in the Elliott Lab. GSEA ranks the differential expression of all the genes in the data set. The maximum gene set size was set at 500 and the minimum gene set size was set at 15 this resulted in filtering out 221/674 gene sets. The remaining 453 gene sets were used in the analysis. A Nominal p value of ≤0.05 was considered significant in the GSEA analyses.

2.12 Splicing analysis using SUPPA2

The changes in splicing were analysed using SUPPA2 (Trincado et al., 2017) performed by Dr Kathleen Cheung. The splicing changes were visualised using the UCSC Genome Browser (http://genome.ucsc.edu/). The RNA-seq reads were aligned in context of the hg38 reference human genome on Interactive Genomics Viewer (IGV) to visualise splice junctions (Robinson et al., 2011). The splice junctions were visualised using sashimi plots that display the coverage for each alignment track as bar graphs this together with the junction tracks can be used to highlight alternative splicing (Katz et al., 2014).

2.13 Gene Ontology (GO) enrichment analysis and interaction network

Gene Ontology Enrichment Analyses represented in dot-plots for iCLIP and chord-diagram for RNA processing events were performed with GOstats v.2.54.0. The chord diagram was produced using the GOplot (v1.0.2) by Dr Katherine James. The dot-plots were produced using ggplot2 v.3.3.2 on R v.4.0.2 by myself. All analyses were done with a p-value cut-off of 0.05.

2.14 Analysis of long human exons

Annotations of all human exons was performed by Dr Sara Luzzi using Ensembl Genes 101 (http://www.ensembl.org/biomart/). Density plot was produced using ggplot2 v.3.3.2 on R v.4.0.2.

2.15 Immunofluorescence microscopy

siRNA of MDA-MB-231 cells was carried on in 6 well plates on cover slips. Cells were fixed in 4% (w/v) paraformaldehyde (PFA) for 20 minutes at RT. Fixed cells were washed 3 times in PBS, permeabilised in 1% Triton X100, washed 3 times in PBS and once in PBST (0.1% Tween20). Coverslips were blocked for one hour in PBS with 10% horse serum in a humid chamber, then incubated with either anti-γ-H2AX (Milipore JBW301) or anti-phospho-RPA32 (Bethyl Laboratories A300-245A) primary antibodies diluted in blocking buffer (1:200) for 2 hours at room temperature or 4°C overnight, washed three times for 5 minutes in TBST and incubated with secondary antibodies in blocking buffer (1:200) in the dark. The Secondary antibodies used were Alexa Fluor® 594 donkey anti-mouse IgG (Abcam ab150108) and Alexa Fluor® 488 donkey anti-rabbit IgG (Abcam 150073). Blots were washed three times for 5 minutes in TBST and mounted onto slides in VECTASHIELD with 4,6-diamidino-2-phenylindole (DAPI). Cells were imaged directly using fluorescence microscopy Zeiss AxioImager (System 3). Analysis of images was carried out using the ZEN blue software. Subsequent image handling was carried out in Adobe Illustrator 2021.

2.16 Comet Assay

The comet assay was perfomed using the Abcam kit according to maufactures instructions. Briefly MDA-MB-231 cells transfected with siRNA were harvested, $1x10^5$ cells were mixed with cold PBS. Cells in PBS were mixed with low melting comet agarose (1/10) and layered on the glass slides pre-coated with low melting comet agarose. The slides were lysed in 1x lysis buffer (pH10.0) (Abcam) for 60 min at 4°C, immersed in Alkaline solution (300 mM NaOH, pH>13, 1 mM EDTA) for 30 min at 4°C in the dark and then electrophoresed in Alkaline Electrophoresis Solution (300 mM NaOH, pH>13, 1 mM EDTA) at 300mA, 1volt/cm for 20 min. The slide was then washed in pre-chilled DI H_2O for 2 min, fixed in 70% ethanol for 5 min and stained with 1x Vista Green DNA Dye (1/10000 in TE Buffer (10 mM Tris, pH 7.5, 1 mM EDTA)) for 15 min and visualized under fluorescence microscopy Zeiss AxioImager (System 3). Analysis of images was carried out using the ZEN blue software. Subsequent image handling was carried out in Adobe Illustrator 2021.

2.17 Quantification and statistical analysis

Quantification and statistical analyses were done using GraphPad Prism v 9.31 (http://www.graphpad.com/).

2.18 Licor Imaging

All IRDye® western blots and iCLIP SDS-PAGE blots were imaged wet on the Licor Odessy FC imaging system using 680 nm and 780 nm channels. Protein detection for the western blots was performed using Image Studio Ver 5.2.

Chapter 3. Global analysis of RBMX RNA-binding protein targets by iCLIP

3.1   Introduction

In the work described in this chapter, I aimed to identify RBMX binding sites on RNA. This was achieved with individual crosslinking and immunoprecipitation (iCLIP) which identifies sequence motifs bound by RBPs (König et al., 2010).

Once DNA is transcribed into RNA these RNA transcripts are bound by proteins that facilitate processing. Crosslinking of RBP and RNA has been used together with deep sequencing techniques to identify global protein-RNA interactions. Over the past several years, CLIP has proved to be a powerful tool in understanding the principles of post-transcriptional regulation. Different versions of CLIP have been used to identify RBP binding sites with great specificity. CLIP relies on identifying the direct protein-RNA interactions by immunoprecipitation and stringent gel purification. The first version of CLIP relied on Sanger sequencing to identify RNA binding sites of Nova in the mouse brain (Ule et al., 2003). Since the initial study, many versions of CLIP have been employed using different experimental conditions (Ule et al., 2005), (Granneman et al., 2009), (Licatalosi et al., 2008), (König et al., 2010).

In my PhD I used iCLIP, a method of CLIP able to identify the binding RBPs at a single-nucleotide resolution. RNA-protein are cross-linked and immunoprecipitated, after immunoprecipitation, protein-RNA complexes are cut from the SDS-PAGE membrane and collected in a size specific manner. Proteinase K is used to digest the protein from the immunoprecipitated RNA, leaving a small peptide at the crosslink site (Figure 3-1, step 7). This peptide is crucial for the reverse transcription step, as these peptides block transit of reverse transcriptase, so these cross linked peptides are where the cDNA fragments are commonly truncated to identify the binding site with nucleotide resolution (König et al., 2010). Previously these cDNAs were lost, however in iCLIP the point of crosslinking in preserved by circularising the cDNA fragments. The iCLIP method has been adapted to plants (Meyer et al., 2017), archaea exosome of S.solfataricus (Bathke et al., 2020) and for spliceosomal branch point mapping (Briese et al., 2019).

iCLIP relies on experimental and computational steps to identify global protein-RNA binding sites. Briefly iCLIP uses ultraviolet (UV) crosslinking to form a covalent bond between RNA and proteins that are in direct contact (Figure 3-1, step 1). The cross-linked protein-RNA fragments are then immunoprecipitated with an antibody specific to the RBP of interest (Figure 3-1, step 3). The immunoprecipitated complexes are separated by SDS-PAGE and visualised using a LI-COR machine (Figure 3-1, step 6). The protein-RNA complexes are cut out of the membrane in a size specific manner, purified and reverse transcribed. The reverse transcription is carried out using a primer with a 5′ barcode or unique molecular identifier (UMI) (Figure 3-1, step 8), the truncated cDNAs are amplified, sequenced and mapped to the genome.

In this chapter, I describe experiments where I optimised and used iCLIP to identify the global targets of RBMX. I identified the RBMX bound genes and performed gene ontology (GO) analysis to identify what set of genes are over represented. I also identified the genomic regions enriched with RBMX crosslinks.
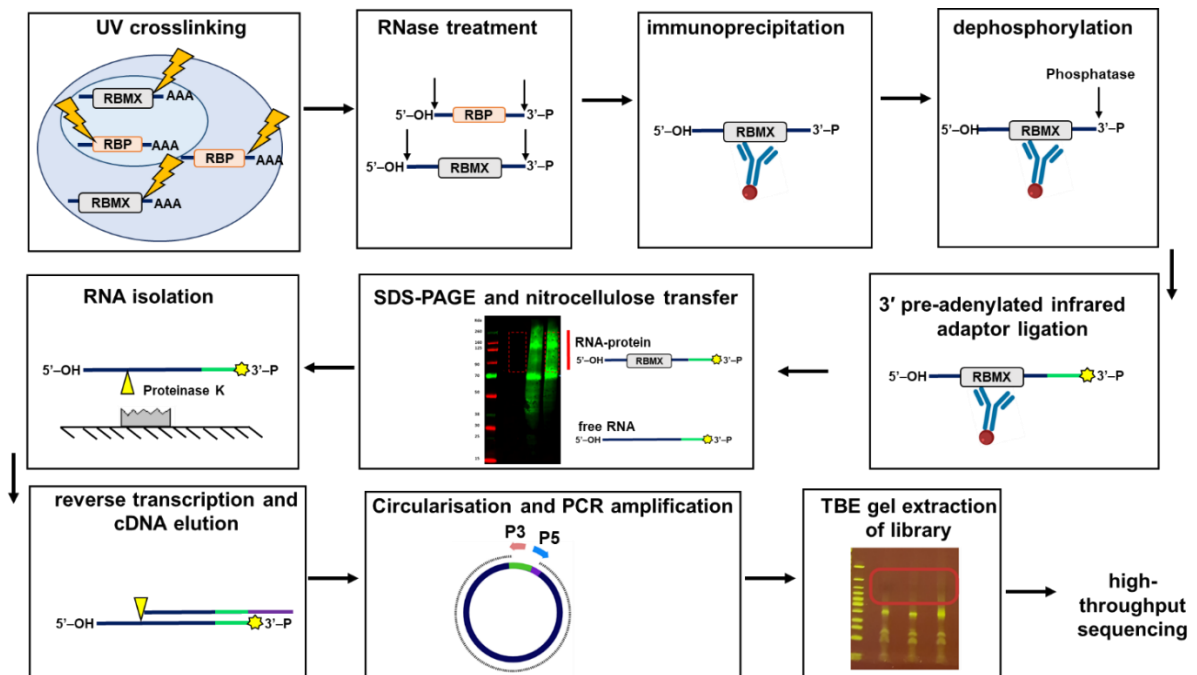
Figure 3-1 *Schematic representation of the iCLIP protocol. (1) Cells were UV irradiated to crosslink proteins and RNA that were in direct contact by causing an irreversible covalent crosslinking. (2) Cells were lysed, sonicated and RNA partially digested with RNase. (3) Using a highly specific antibody the protein and RNA complexes were immunoprecipitated. (4-5) Following the IP RNA was dephosphorylated, and a pre-adenylated L3-IR adaptor 3' RNA linker was added to the RNA. (6) SDS-PAGE was used to separate the protein-RNA complexes by size followed nitrocellulose transfer. (7) The protein-RNA region of interest was separated from the membrane using a printed template of the membrane, then the protein was digested with proteinase K. (8) The RNA was purified and reverse transcribed into cDNA using a primer with a barcode sequence. (9) The cDNA was purified via AMPure XP beads capture, circularised using a ssDNA circligase enzyme (the cDNA was circularised so that the site of transcription termination was annealed to the barcode sequence within the DNA linker) and PCR amplified using solexa primers designed for pair end sequencing. (10) The cDNA was purified via AMPure XP beads capture, the eluded cDNA was then separated on an agarose gel and DNA in the range of 150-400nt excised and purified. (12) The purified cDNA was sent for next generation sequencing and reads were aligned to the genome.*

3.2   Chapter aims

1. To identify global targets of RBMX using individual nucleotide-resolution CLIP (iCLIP) in HEK293 cells that can be induced to overexpress RBMX

2. To investigate whether RNA binding and splicing could explain the global function of RBMX in cells

3.3 Results

In order to identify the global RNA targets of RBMX, I begun to optimise the iCLIP experiment for RBMX in MDA-MB-231 cells according to the protocol published in (Konig et al., 2011). This was the same cell line as our RNA-seq data that will be discussed in Chapter 4. I encountered difficulties with loss of RNA in the multiple steps of the protocol. The commercial anti-RBMX antibody was also very dilute and quite expensive. I therefore switched to use a HEK293 flip-in RBMX expression cell line created by (Kheirollahi Kouhestani & Kouhestani, 2013) in addition with a modified iCLIP protocol (Lee et al., 2021). This engineered cell line expresses a FLAG tagged RBMX protein.

The Ule lab recently developed an improved iCLIP protocol that incorporated the use of an IR dye, thus avoiding radioactive labelling of the immunoprecipitated RNA with ^{32}P radioactivity as used in the original iCLIP experiments. The crucial biochemical steps such as proteinase K digestion, RT PCR and circularisation are kept the same as in the original iCLIP. Therefore, the results remain the same. The modified iCLIP protocol uses an L3 adaptor conjugated to an infrared dye 800 via an internal Azide modification, enabling the visualization of the adaptor and ligation products using a LICOR imaging machine. The modified protocol also used different elution conditions from IP beads, circularization conditions (betaine) and 2x AMPure beads for cDNA size purification increasing the amount of sample retained during the protocol.

3.3.1 Optimisation of the RBMX iCLIP experiment

The iCLIP was carried out in an RBMX over expressing cell line. This was largely due to the lack of a high quality RBMX antibody available commercially. I first tested whether the RBMX-FLAG Flp-In HEK293 cell line was expressing a similar amount of RBMX as compared to expression of endogenous RBMX by western blot (Figure 3-2). RBMX-FLAG Flp-In HEK293 cells were grown in culture. RBMX-FLAG Flp-In HEK293 cells were plated in 10cm dishes and left to settle for 24 hours. RBMX-FLAG expression was activated by treating the cells with 1mg/ml of tetracycline and cells were grown for an additional 24 hours. Western blots with the commercial anti-RBMX antibody could distinguish the endogenous RBMX protein from the slightly larger RBMX-FLAG protein. These Western blots showed that the amount of RBMX-FLAG in the HEK293 cell line was comparable to the endogenous RBMX expression (Figure 3-2). The RBMX-FLAG protein is thought to be functional because the FLAG protein

is at the C-terminus which is far from the RNA recognition motif we are interested in investigating.   In addition an RBMXL2-FLAG cell line designed in the lab is fully functional and has been shown to rescue RBMX KD in cells. Furthermore, when the RBMX-FLAG cell line was used to observe RBMX splicing targets I observed a splice site switch in *ETAA1* from the cryptic 3′ to the normal splice site (Figure 4-11).

Since the tetracycline-inducible RBMX protein expressed by these RBMX-FLAG Flp-In HEK293 cells has an N-terminal flag tag, a FLAG specific antibody was used for the detection. As a control, Flp-In HEK293 cell cells made with an empty pcDNA plasmid, and so not expressing the RBMX-FLAG, were also used in the western blot and throughout the iCLIP experiments. GAPDH was used as a loading control, and even loading was seen in both the control Flp-In and the RBMX-FLAG Flp-In HEK293 cell lines. Exactly as predicted, no RBMX-FLAG signal was detected in the control cell line (Figure 3-2).

24hr tetracycline induced expression of RBMX
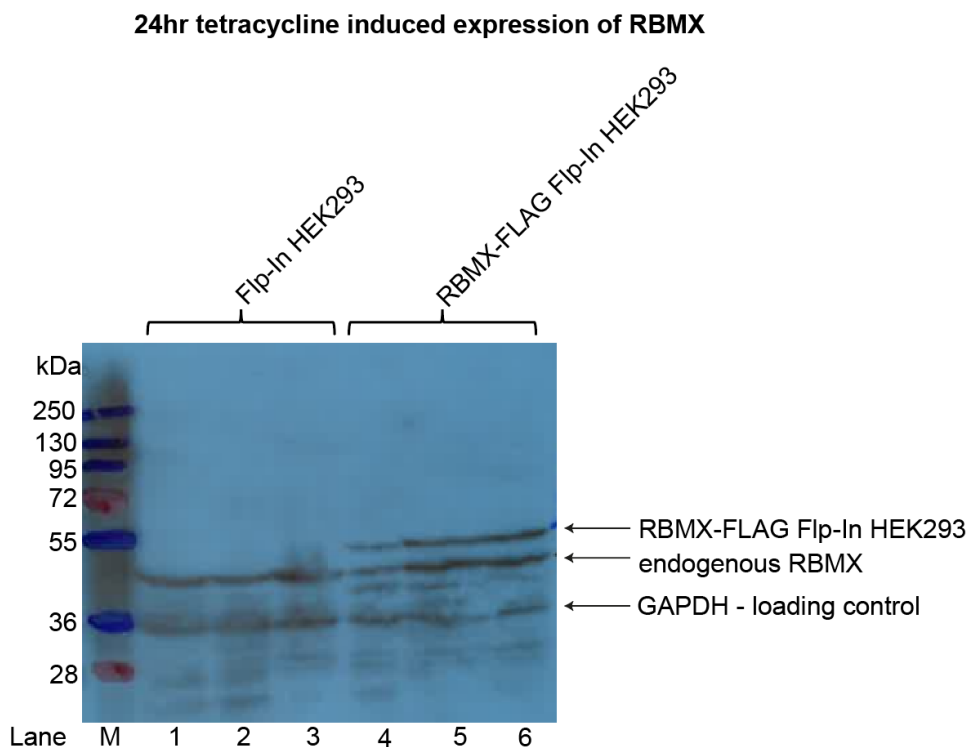
Figure 3-2 Tetracycline induced expression of RBMX-FLAG in the RBMX-FLAG Flp-In HEK293 cell line. Control Flp-In HEK293 cell line cells do not express FLAG-tagged RBMX, but do express endogenous RBMX and GAPDH. Lane M has protein size marker, lanes 1-3 were the control Flp-In HEK293 cell line, and lanes 4-6 were the RBMX-FLAG Flp-In HEK293 cell line.

41

The next step to optimise was the immunoprecipitation of RNA-protein complexes, monitored by polyacrylamide gel electrophoresis and detection of IR-labelled complexes. This is step 3 in Figure 3-1. The iCLIP protocol requires optimisation at a number crucial steps, to ensure specific immunoprecipitation of RBMX-FLAG and library preparation. The FLAG antibody is highly specific. This was tested by using a control IgG antibody for an IP on the UV cross-linked iCLIP samples. No signal was seen in the IgG antibody IP samples, and an IR signal was only seen in the RBMX-FLAG expressing samples when a FLAG antibody was used for the IP (Figure 3-3).
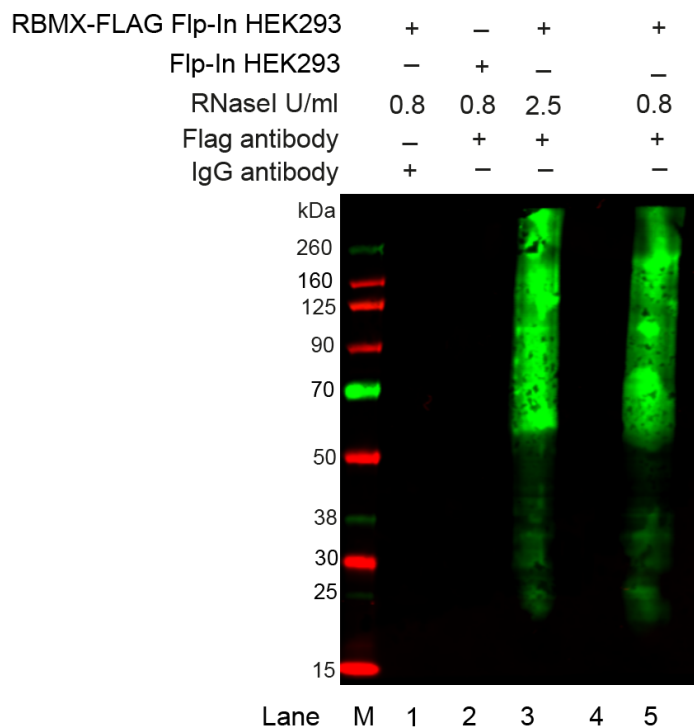
Figure 3-3 Optimisation of the immunoprecipitations step. *This shows a sample at the SDS-PAGE and nitrocellulose membrane stage of the iCLIP procedure shown in Figure Figure 3-1 step 6. UV cross-linked samples were used for IP using IgG 9 (lane 1) and anti-FLAG antibodies (lanes 2, 3 and 5). I then used SDS-PAGE followed by Western blotting, and a LiCor machine to detect IR-labelled immunoprecipitated samples. The RNA-protein complexes run as a smear above the size of RBMX-FLAG protein (50KDa).*

I next optimised the RNase concentration used for the digestion of immunoprecipitated RNA sequences. During step 2 of Figure 3-1 I treated cross-linked cell lysates with

increasing amounts of RNase I. I determined that 0.8U/ml was the ideal RNaseI dose (Figure 3-4) following advice from the Ule lab.
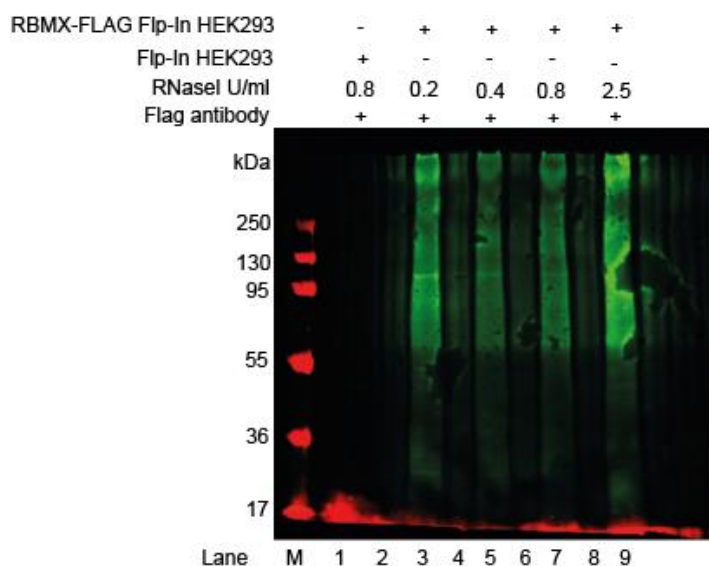
Figure 3-4 Optimisation of RNaseI digestion. LICOR analysis of Western blot after immunoprecipitated samples were analysed by SDS-PAGE. In this experiment, cross-linked HEK293 cell lysates were treated with decreasing concentrations of RNaseI before immunoprecipitation in lanes 3, 5 and 7. A sample in lane 1 used a lysate from the control Flp-In HEK293 cell line *that does not express FLAG-RBMX as a control for antibody specificity.*

The next stage to optimise was the number of PCR cycles to use (in Step 9 in Figure 3-1). PCR cycle number is important for maintaining complexity of the final iCLIP library. Fewer cycles are better, but there must be enough cycles to generate PCR products for cloning. The number of PCR cycles for amplification of the final cDNA library was optimised to 19 cycles. This is the number of cycles that gave the highest specific amplification with little non-specific amplification.

3.3.2   iCLIP analysis of RBMX binding in HEK293 cells

Having optimised the RNase concentration and PCR cycles, the full iCLIP experimental procedure was carried out. The RBMX-FLAG Flp-In HEK293 cells were treated with tetracycline for 24 hours to induce RBMX protein expression before UV cross-linking. Then after the iCLIP procedure, libraries prepared in three biological replicate experiments were sent for sequencing on an Illumina NextSeq 500.

I monitored the progress of the iCLIP protocol at multiple checkpoints. First, in these experiments I could see the position of the RBMX-FLAG infrared cross-linked protein at about 57KDa. Infrared-labelled RNA-protein complexes migrating above the position of this 57kDa band corresponded to the position of RBMX bound to RNA fragments. No such RNA-protein complexes were detected using lysates from the control Flp-In HEK293 cell line (Figure 3-5, A). Secondly, the cDNA library was checked on a TBE gel after library amplification (Figure 3-5, B). The cDNA library fragments between 150bp and 400bp with no over amplification were gel purified. Samples were then quantified on the bioanalyser and then sequenced by the genomics core facility at Newcastle University.
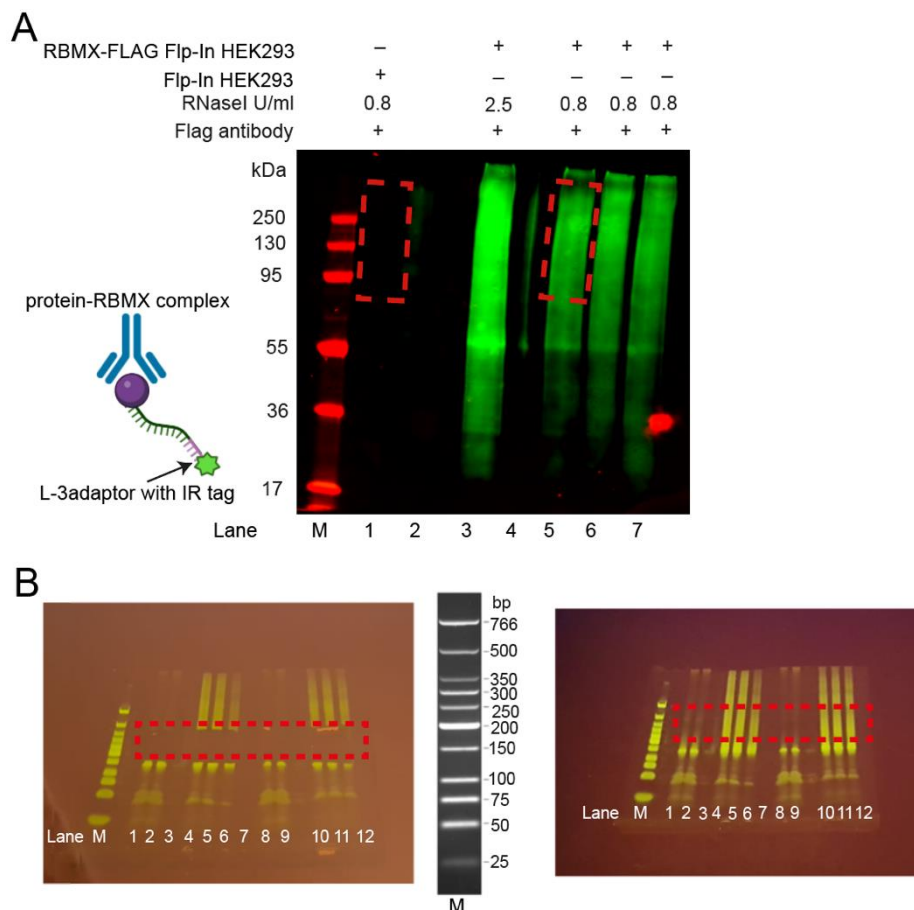
Figure 3-5 Checkpoints used to monitor the progress of the iCLIP experiment. *(A) SDS-PAGE and western blot IR signal of RNA-RBMX protein complexes detected by LI-COR. The red box around the lanes represents the region that was excised from the membrane for RNA-protein complex isolation. Lane M is the protein size marker, lane 1 Flp-In HEK293 control, lane 3 the RBMX-FLAG Flp-In HEK293 high RNase positive control, Lane 4-7 RBMX-FLAG Flp-In HEK293 low RNase condition. (B) TBE agarose gel of two replicates of the cDNA libraries. In the red dotted box is the region excised for gel purification. The ladder in the middle of the gels is the DNA size marker*

(M). Lane M is the DNA size marker, lane 1-2, 8-9 Flp-In HEK293 control cDNA library, Lane 4-6, 10-12 RBMX-FLAG Flp-In HEK293 low RNase condition cDNA library.

3.3.3 Analysis of genome-wide RNA targets for RBMX iCLIP

After the iCLIP procedure, sequencing data was obtained as Fastq files and submitted to the iMAPs web server (http://imaps.genialis.com). Unprocessed fastq files were demultiplexed and analysed using the primary analysis pipeline. iMAPs was used to check the quality of the sequencing reads, demultiplex the reads according to the UMI, trim adaptors and remove duplicates. Then iMAPS pipeline aligned the iCLIP libraries to the GRCh38/hg38 human reference genome. Once the reads were mapped to the genome by alignment, crosslink events were identified for each replicate. The icount group analysis pipeline on iMAPs (https://imaps.genialis.com) was used to merge the three replicates. Crosslinks within a maximum of 20bp of each other were merged into clusters and this file is the final bed file used for crosslink position identification. A total of 14,653 genes were identified as binding to RBMX. Principal component analysis (PCA) and correlation analysis was run using the Galaxy server (https://usegalaxy.org/), using the raw bam files (Figure 3-6). PCA showed that a high correlation was seen between all 3 replicates (Figure 3-6, A). In addition the replicates all clustered together in PCA 1 in the PCA and scree plots (Figure 3-6, B).
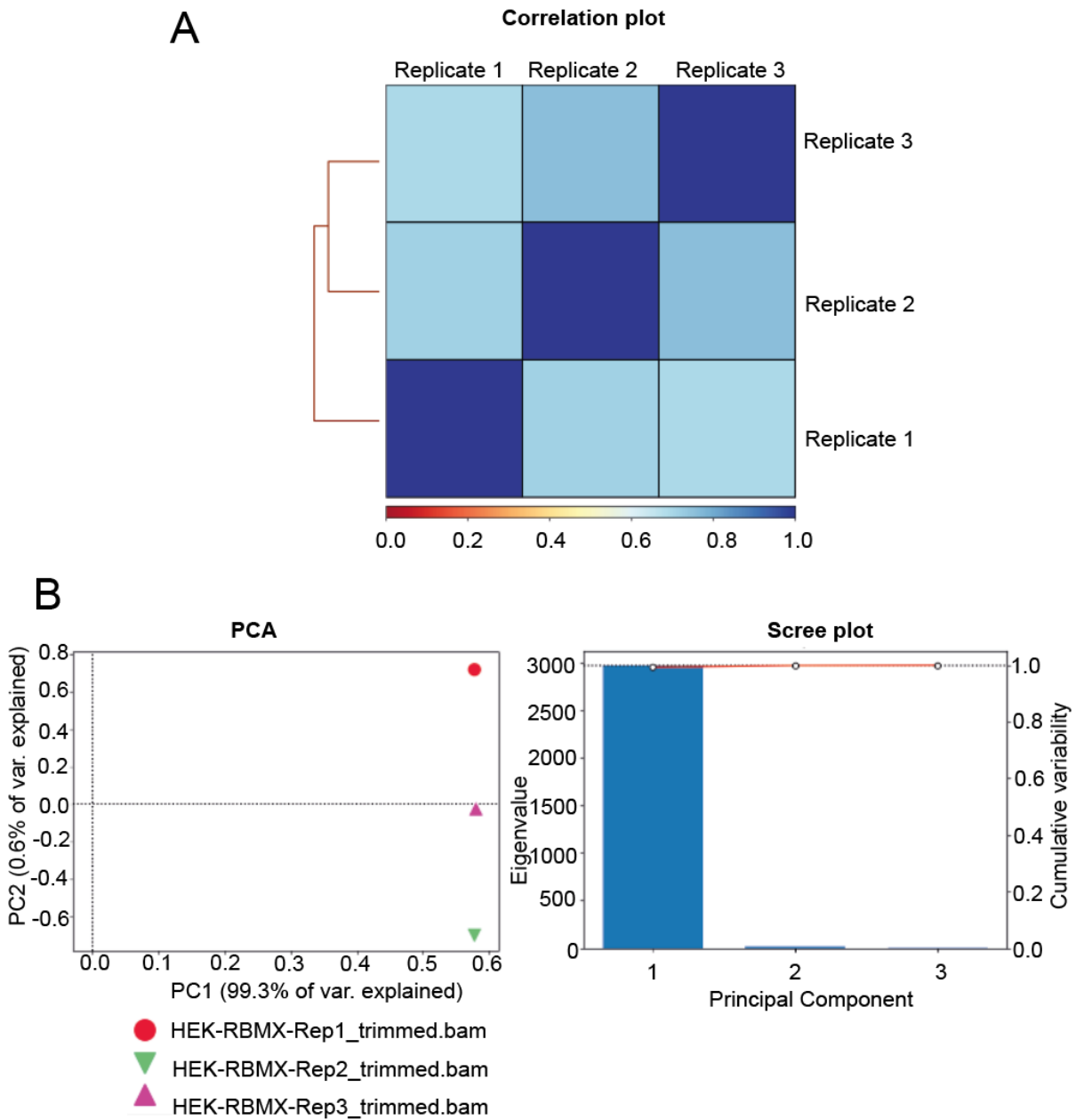
Figure 3-6 Analysis of the reproducibility between the raw bam files of each of the 3 biological replicate iCLIP analyses of the RBMX-FLAG Flp-In HEK293 cell line. *(A) Correlation matrix indicates a positive correlation between the replicate samples. (B) PCA plot indicates little variance in the replicates as they all fall under PC1 in the scree plot on the right. Plots produced on https://usegalaxy.org/.*

3.3.4 Analysis of known RNAs that have RBMX binding sites

Although global targets of RBMX are poorly understood, there are some reported examples of RBMX targets. One of these, *NORAD*, is a long non coding RNA (lncRNA) that is involved in maintaining genomic stability in mammalian cells by preventing aberrant mitosis. A previous study showed that RBMX was involved in the binding and regulation of NORAD and that it had the highest RBMX binding in the transcriptome

(Munschauer et al., 2018). A more recent study showed that RBMX was binding to NORAD but it was not necessary for its role in genomic stability (Elguindy et al., 2019). RBMX binding was detected within *NORAD* in the iCLIP data set. However, this RBMX binding is very specific to one region in the 5′ region of the *NORAD* RNA (Figure 3-7). The published binding of RBMX on *NORAD* shows binding in an 800nt region of NORAD and high binding in the 5′ region.

RBMX and its retrogene RBMXL1 have been shown to control chromatin state by controlling the transcription of *CBX5* (also known as heterochromatin protein 1 alpha (HP1-α)) in acute myeloid leukaemia in humans and in mice (Prieto et al., 2021). Using data from the iCLIP experiment in this study I found that RBMX binds to *CBX5* mRNA, and this binding is particularly high in the 3'UTR region (Figure 3-7). Another of the characterised targets of RBMX is *SMN2* exon 7. Using data from the iCLIP experiment I found that RBMX binding was identified 188bp upstream of *SMN2* exon 7, but not within exon 7 itself (Figure 3-7). We have identified RBMX binding sites within *SMN2* RBMX however it is not a major binding target of RBMX.
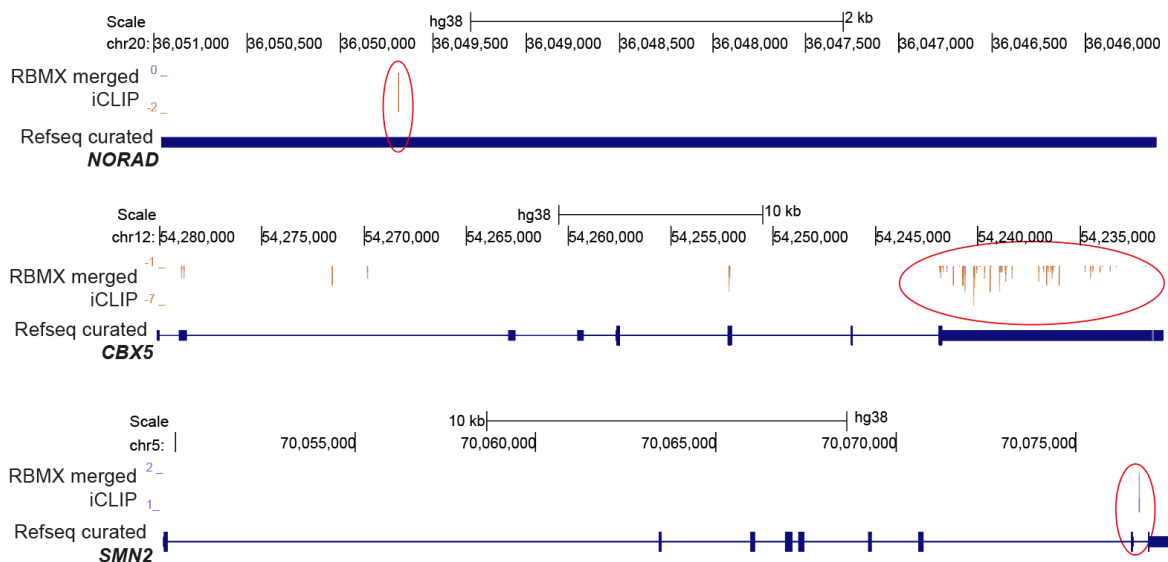
Figure 3-7 UCSC screenshots of RBMX crosslink sites identified within NORAD, CBX5 and SMN2. *Regions with the highest RBMX binding density are circled in red.*

47

3.3.5 Location of RBMX binding within different classes of genes

Having determined that the iCLIP libraries made by the 3 biological replicate experiments had a positive correlation with each other and little sample variation, descriptive downstream global analysis was performed using the merged file. The location of RBMX binding sites in transcript regions can help decipher its function in binding RNA. Thus, RBMX binding sites were overlaid with gene annotations to assign each binding site with a gene. Binding sites that did not fall within a gene were categorised as intergenic. To identify the RBMX bound genes I modified an R script from (Busch et al., 2020). The target spectrum of RBMX bound genes is shown in Figure 3-8. These analyses showed that RBMX binding sites were highest in protein-coding genes and lincRNA.

Total=14,653
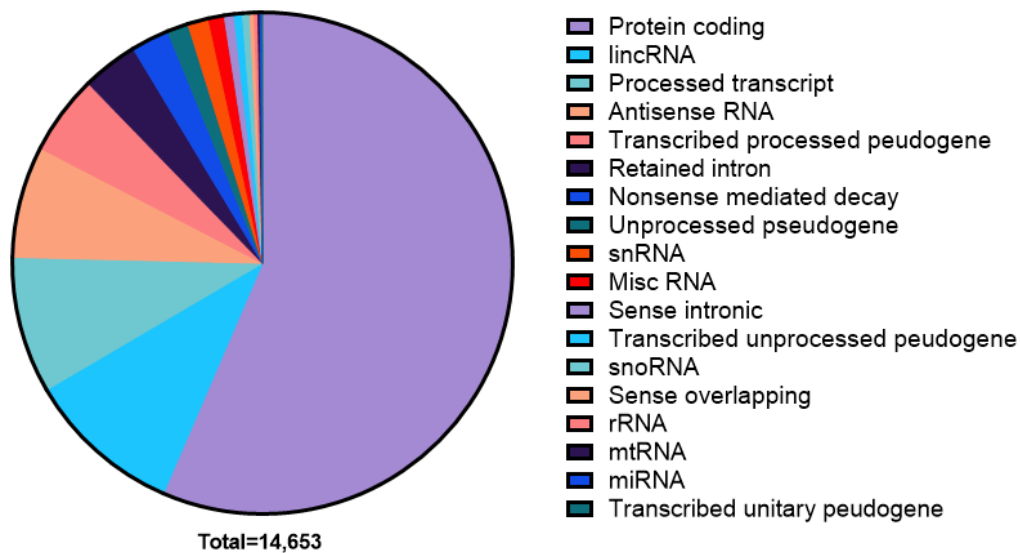
Figure 3-8 Bar chart showing RBMX binding within different classes of genes. RBMX binding is highest in protein coding and lincRNA genes.

3.3.6 Patterns of RBMX binding within the genome

Transcript regions can be divided into intron, exons and intergenic (the regions between genes). In protein coding genes exons are further divided into 5′ untranslated region (UTR), coding sequence (CDS) and 3′ UTR. The transcript regions that RBMX preferentially binds to were identified using the iMAPS pipeline. This binding profile is shown as a bar chart in Figure 3-9. A vast majority of RBMX binding sites fall in the intron and intergenic regions. One explanation for this could be that the intron and

intergenic regions are much longer and therefore form a big part of the genome. High RBMX binding in the intron region suggest a role of RBMX in pre-mRNA splicing. RBMX-crosslink sites are also highly detected within the CDS and 3' UTR of protein coding genes (Figure 3-9).

Global distributions of RBMX iCLIP crosslinks within the genome
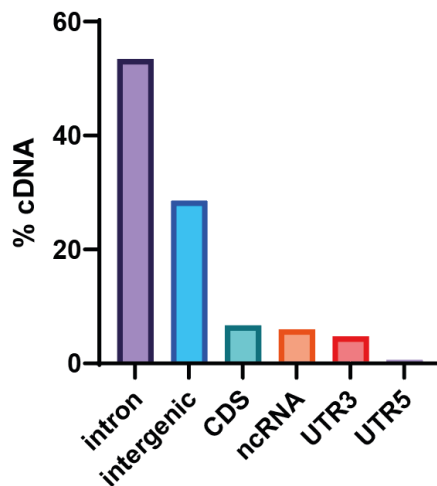
Figure 3-9 Bar chart representing the percentage of RBMX bound cDNA within the genome regions. *RBMX binding is high in intronic and intergenic regions.*

The full list of RBMX bound genes was used to identify enriched gene ontology (GO) terms. This Gene Ontology analysis (shown in Figure 3-10) revealed that many of the top 20 significantly enriched pathways were related to mitotic sister chromatid separation, DNA replication, DNA repair and mitotic cell division.
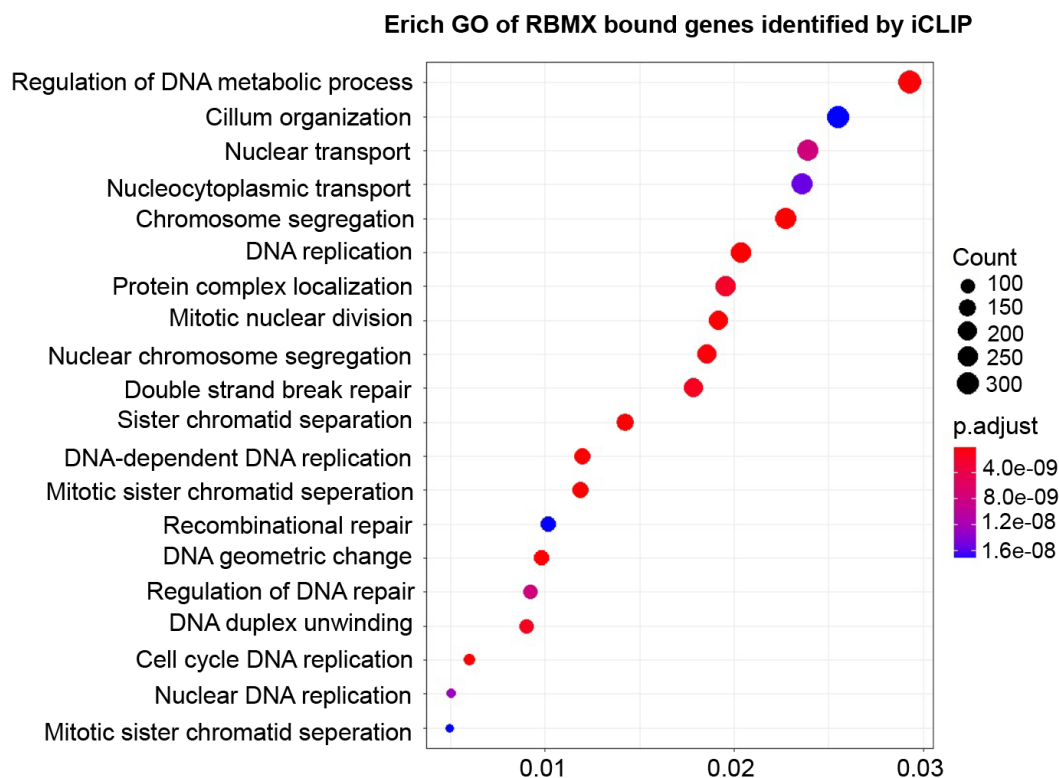
Figure 3-10 Dot plot representing gene ontology enrichment analysis of RBMX bound genes identified by iCLIP. *The plot was generated using GOstats v.2.54.0, and ggplot2 v.3.3.2 packages on R v.4.0.2. Count = number of genes.*

3.3.7 Identification of RBMX binding site sequences

Most RBPs recognise specific linear motif sequences of 6-8nt long. However, the RNA binding site of RBMX determined by NMR was only AAN (Moursy et al., 2014). The presence of the motifs in the RBMX bound RNAs was identified from the iCLIP data. With the help of Dr Ivaylo Yonchev (Sheffield University) enriched motif and metagene analysis was carried out. Consistent with expectations (Moursy et al., 2014), motif analysis identified that RBMX binding was enriched in AAA/AAG rich sequences (Figure 3-11).
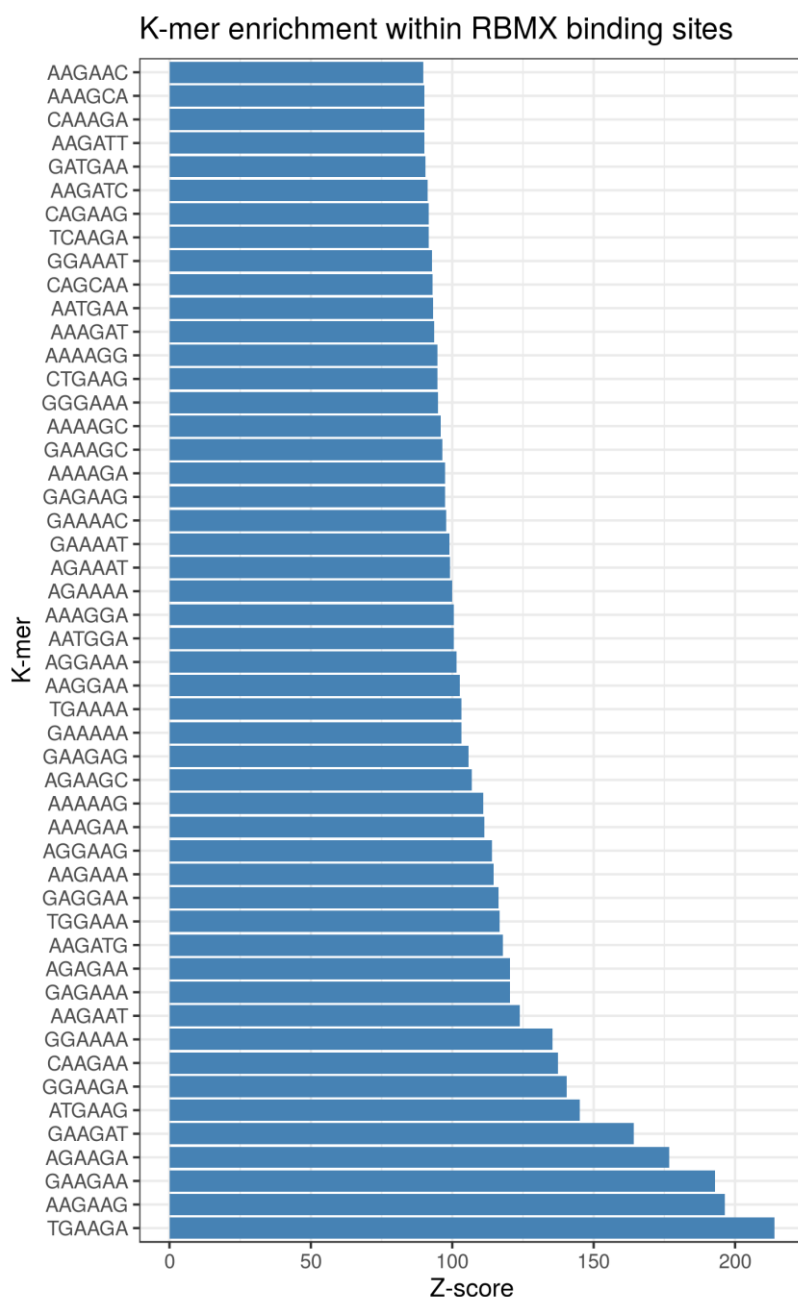
Figure 3-11 Top 50 K-mer enrichment within RBMX binding sites.

3.3.8 RBMX binding density

Metagene analysis, showing the RBMX binding sites profile was also carried out by Dr Ivaylo Yonchev (Sheffield University) (Figure 3-12). The geneset used for these metagenes is generated from the ensembl_noalt_85 hg38 annotation set. The metagenes were created by taking every protein-coding gene and removing the ones that have an annotation for a different gene overlapping on the same strand. Genes that had annotations for a different gene on the same strand within 1kb of the start or

end of the gene were removed. The full list of RBMX regulated genes bound by RBMX is shown in the appendix B.

In this analysis, each gene gives a pattern of enrichment and every gene has an equal contribution to the final metagene. The strongest signal of RBMX binding lies within the end of the full transcript. The 3' end of the transcript is enriched in RBMX binding sites and the 5' signal is weaker Figure 3-12.
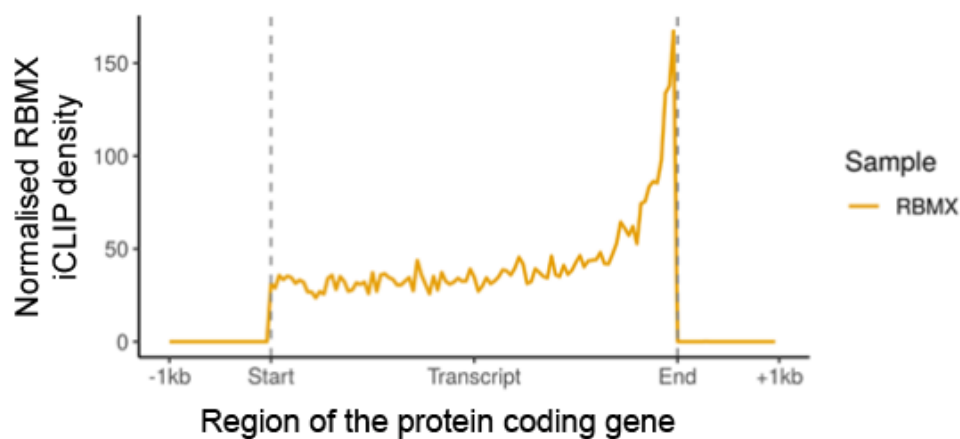
Figure 3-12 Metagene analysis of RBMX iCLIP from HEK293 cells shown across all transcript regions normalised per gene. *Binding density refers to the number of RBMXbindng sites within the region. The longest protein coding gene transcript has been used in the metagene analysis*

3.4 Discussion

In this chapter, I have used iCLIP to identify RBMX binding sites genome wide. RBMX was found to bind highly in protein-coding genes. Gene Ontology analysis revealed that RBMX binding is enriched in genes involved in DNA replication, DNA repair and the cell cycle. Using our iCLIP data, we were able to confirm the binding of RBMX to RNAs that have been shown to be regulated by RBMX in the literature. The iCLIP data was not normalized for gene expression in HEK293 cells. Future experiments would benefit from normalizing the iCLIP data with gene expression data to account for the abundance of transcripts expressed in HEK293 cells.

3.4.1 RBMX iCLIP in RBMX-FLAG Flp-In HEK293 cells

RBMX binds to RNA via its N-terminal RNA recognition motif. These interactions have been investigated for certain genes like *SMN2* exon 7 (Sivaramakrishnan et al., 2017). Even so, RBMX-RNA interactions have not been investigated using genomic profiling techniques. The major aim of this part of the project included doing global analysis of RBMX binding in HEK293 cells. Here we identified the RBMX binding sites across the genome. We used an iCLIP-seq method to identify RNA targets of RBMX genome wide in RBMX-FLAG Flp-In HEK293. The iCLIP method in this chapter utilized an IR linker added to the L'3 adaptor to label the RNA instead of 32P-γATP labelling. The iCLIP experiment was optimised for RBMX immunoprecipitation and cDNA amplification. Visualisation of the SDS-PAGE and transfer showed that the FLAG antibody used was specific, and so should isolate genuine RBMX-RNA interactions. The use of UMI (randomly generated barcodes) sequences during the iCLIP protocol allowed for identification of unique cDNAs in combination with computational analysis: PCR duplicates from over amplification can be removed after sequencing. The data generated by this iCLIP is highly consistent between samples, as all three biological replicates used have a high correlation.

3.4.2 RBMX binding within the genome

We used iCLIP to analyse the genome wide distribution of RBMX binding sites. We found that RBMX binds in different gene groups with the greatest enrichment in protein-coding genes and lincRNA. The patterns of RBMX binding within the genome showed that RBMX binding is high within intronic and intergenic regions. Binding within introns

on pre-mRNA indicates a role in pre-mRNA processing such as splicing. One of the well-known roles of RBMX is pre-mRNA splicing. RBPs binding in intronic regions is associated with exon repression. RBPs such as hnRNPC, PTB Fox and Nova have been found to bind within branchpoints and 3′ splice sites were they are involved with silencing exon inclusion. RBMX may be following the same pattern.

RBMX has been shown in the literature as a splicing repressor binding to intron splicing silencer sites within introns (Thonda et al., 2022). The iCLIP data presented here together with the RNA-seq data support the role of RBMX as splice site silencer binding within introns.

3.4.3 *Identifying RBMX binding motifs*

The kmer analysis to identify common RBMX binding sequences showed that RBMX binds highly in regions rich in AAA/AAG nucleotides, suggesting perhaps a role in binding at branch points or 3'ss of introns during exon definition. The motif identified in this iCLIP matches the known RBMX binding site identified by gel shift assays in SMN2 exon 7 (Moursy et al., 2014). Interestingly metagene analysis across transcript regions showed a high enrichment of RBMX binding in the 3'UTR of transcripts suggesting a role for RBMX in alternative 3′ UTR splicing. The binding of RBMX in 3'UTR may have a role in inhibiting or enhancing mRNA decay having a role in mRNA stability and gene expression.

3.4.4 *RBMX bound genes have a role in genome instability*

Gene Ontology analysis of the RBMX bound genes shows an enrichment of terms relating to DNA repair, DNA replication and cell cycle. Our iCLIP data aligns with the previously described role of RBMX in literature where it has been shown that RBMX is involved in the regulation of genome stability.

The iCLIP in this chapter allowed us to map the transcriptome-wide binding sites of RBMX-RNA interactions at single-nucleotide resolution. An important thing to consider for future experiments is to sequence the negative control condition to account for any background binding.

3.5 Chapter summary

This chapter enabled me to identify the RBMX-RNA targets and the RBMX binding sites genome wide using iCLIP in HEK293 cells. The next step is to identify alternately spliced exon regulated by RBMX. In the next chapter, I identify RBMX splicing targets and investigate if RBMX binds directly within or near the exons.

Chapter 4. Transcriptome-wide analysis of RBMX RNA processing targets using RNA-Seq

4.1 Introduction

RBMX is an RNA binding protein (RBP) that is ubiquitously expressed in all vertebrates (Elliott et al., 2019). In this chapter, I used previously available RNA-seq data provided by Dr Gerald Hysenaj (Hysenaj, 2020) following siRNA depletion of RBMX in MDA-MB231 cells. I have identified alternative RNA processing events between the control and RBMX knockdown samples.

In addition to having a role in splicing RBMX has been implicated in DNA damage and chromosome biology. RBMX is involved in the cohesion of sister chromatids in HeLa cells. Upon RBMX depletion an increase of premature separation of sister chromatids was seen. This activity was not dependent on the RNA Recognition Motif (Matsunaga et al., 2012). An siRNA screen showed that RBMX was a positive regulator of homologous recombination (HR) by facilitation of proper expression of BRCA2 (Adamson et al., 2012). When RBMX was depleted in MDA-MB231 cells, we hypothesised that like RBMXL2 there would be an increased use of cryptic RNA-processing events within genes important for DNA replication and DNA damage. To examine whether RBMX similarly regulates the splicing of genes involved in biological processes related to DNA damage and chromosome biology we use Gene Sent Enrichment Analysis (GSEA) in collaboration with Dr Kat Cheung (Newcastle University, Bioinformatics Support Unit) and Gene Ontology in collaboration with Dr Katherine James (Northumbria University) to test for enrichment of pathways in genes with RBMX regulated exons.

SUPPA2 analysis was done to identify differentially expressed mRNA isoforms upon RBMX depletion in MDA-MB-231 cells. To validate the alternative RNA processing events, RT-PCR experiments were carried out in MDA-MB231 (breast cancer), MCF7 (breast cancer) and NCI-H520 (lung cancer) cell lines following RBMX knockdown. In addition, RT-PCR was carried out to investigate if there was a splicing switch upon RBMX over expression in HEK293 cells over expressing RBMX. In this Chapter, I describe analyses that correlate the RNA processing sites identified by RNAseq together with the iCLIP data to identify RBMX binding. I hypothesised that RBMX

binding could affect RNA processing by sterically blocking splice sites, branch points or polyA sites close to the RNA processing events.

I aimed to determine whether RBMX binding sites were close or on RBMX regulated exons (Zhou et al., 2019) and which splicing events were conserved between the two cell lines. The iCLIP data was from HEK293 cells and the RNA-seq data was from MDA-MB-231 cells. We also used a published RNAseq data set from GEO under accession number GSE114311 after RBMX knockdown in HEK293 cells. Dr Sara Luzzi performed genome alignment of the RNAseq data from HEK293 cells to produce bam files for visual inspection on IGV. We found 48% of RBMX-regulated RNA processing events were conserved between HEK293 and MDA-MB-231 cells. Using these RNA processing events that were conserved between the two cell lines, we were able to identify RBMX binding sites within the regulated exons.

In order to investigate the splicing profile further minigenes were constructed. The use of mini-genes is a traditional way of studying alternative splicing (Cooper, 2005; Stoss et al., 1999). In these experiments exons or splice sites of interest are cloned into an exon trap vector to make a minigene. The minigene is then co-transfected into HEK293 cells either with expression vectors that express GFP or splicing regulators fused to GFP.

During pre-mRNA splicing the spliceosome recognises a number of key signals within the target RNA. These include the 5′ splice site, 3′ splice site, branch point and polypyrimidine tract. The minigene approach was also used to map the branch points used by the cryptic splice event. In this chapter I also describe RT PCR experiments where I mapped the branch point adenine upstream of the cryptic 3′ splice site in *ETAA1* exon 5 used for the splicing on the small exon (Vogel et al., 1997). Using the iCLIP data from chapter 3 I have been able to identify the RBMX crosslink sites in close proximity to the branch point and cryptic exon splice site.

4.2 Aims

The aims of the work described in this chapter were to:

1. Investigate transcriptome changes in HEK293 and MDA-MB231 cells after siRNA depletion of RBMX.

2. Identify RBMX binding sites in exons that are alternatively spliced when RBMX is depleted.

3. Validate splicing regulation of alternatively spliced exons identified from the RNA-seq data.

4. Investigate whether RBMXL2 and Tra2β proteins are involved with RBMX in splicing co-regulation of target exons.

5. Correlate the profile of RBMX binding sites and branchpoints used by cryptic splice sites in the *ETAA1* gene.

4.3 Results

4.3.1 *Analysis of differentially expressed genes using RNA-seq data from MDA-MB-231 cells*

Dr Kat Cheung performed DEseq analysis (Love et al., 2014) to identify differentially expressed genes using the RNAseq data obtained after RBMX knockdown in MDA-MB-231 . The MA plot in Figure 4-1 was made using DESeq values from this analysis. The genes identified as significantly differentially expressed are highlighted in red. In this dataset some genes involved in cell cycle and tumour suppression had reduced expression in RBMX knockdown cells compared to controls. These include *RAB17*, which is a a tumor suppressor gene encoding for a protein involved in inhibiting cell invasion (Wang et al., 2020); and *CCNB2* which encodes a cyclin protein involved in the activation of CDKs during the G2/M phase in the eukaryotic cell cycle (Wang et al., 2020). On the other hand we saw increased expression of *CCL3, CCL3L3, SERPINE1, SERPINB2* genes that encode for proteins involved in the inflammatory and immunoregulatory responses. The strongest expression change was detected in *RBMX*, which was depleted by siRNA in these experiments.
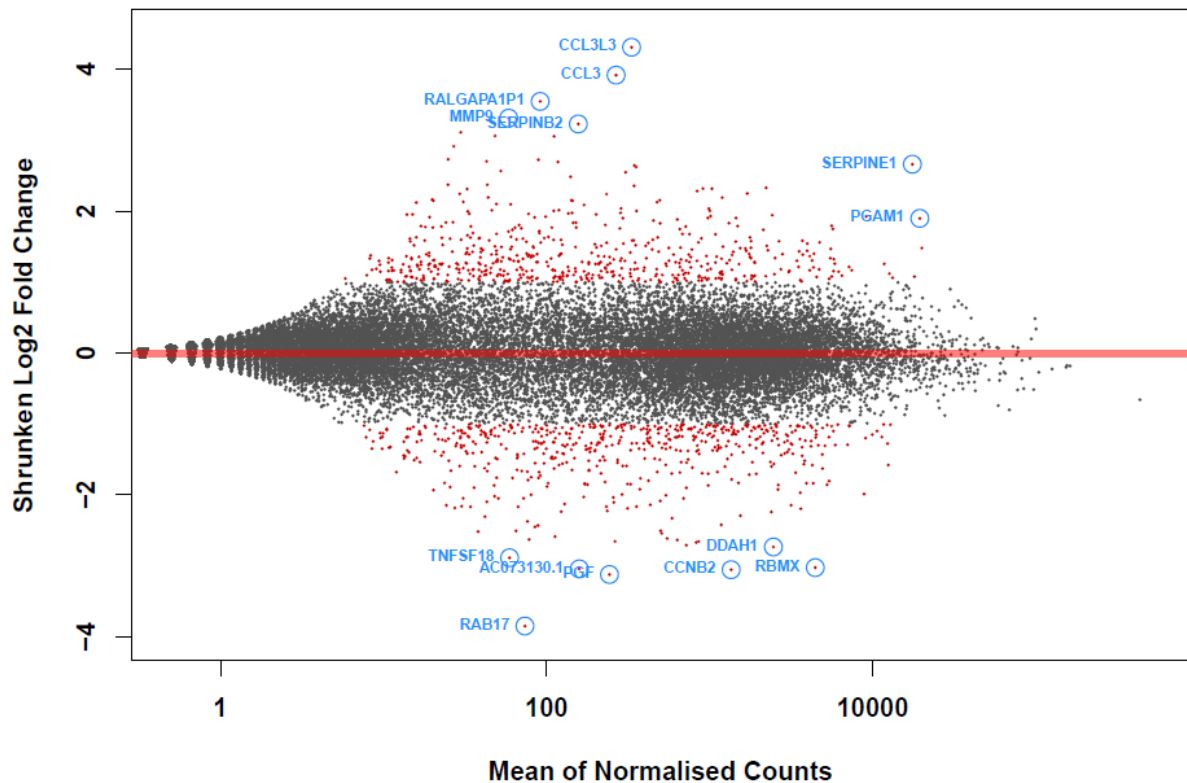
Figure 4-1 MA plot showing results of DESeq (Love et al., 2014) ***analysis after RBMX depletion.*** *The significantly differentially expressed genes with a false discovery rate less ≤ 0.05 are shown in red.*

4.3.2 DNA replication stress and cell cycle terms are enriched by RBMX depletion

In addition to looking at the differentially expressed genes in the MA plot, Gene set enrichment analysis (GSEA) was performed by Dr Kat Cheung. GSEA is a computational method that searches for common biological pathways regulated by a pre-defined set of genes, developed by the Broad Institute (Subramanian et al., 2005). RNA-seq data analysis using GSEA demonstrated that a high number of genes involved in the cell cycle and DNA replication pathways were significantly enriched when RBMX was depleted compared to control cells (Figure 4-2). This helps to identify biologically important genes and pathways controlled by RBMX.
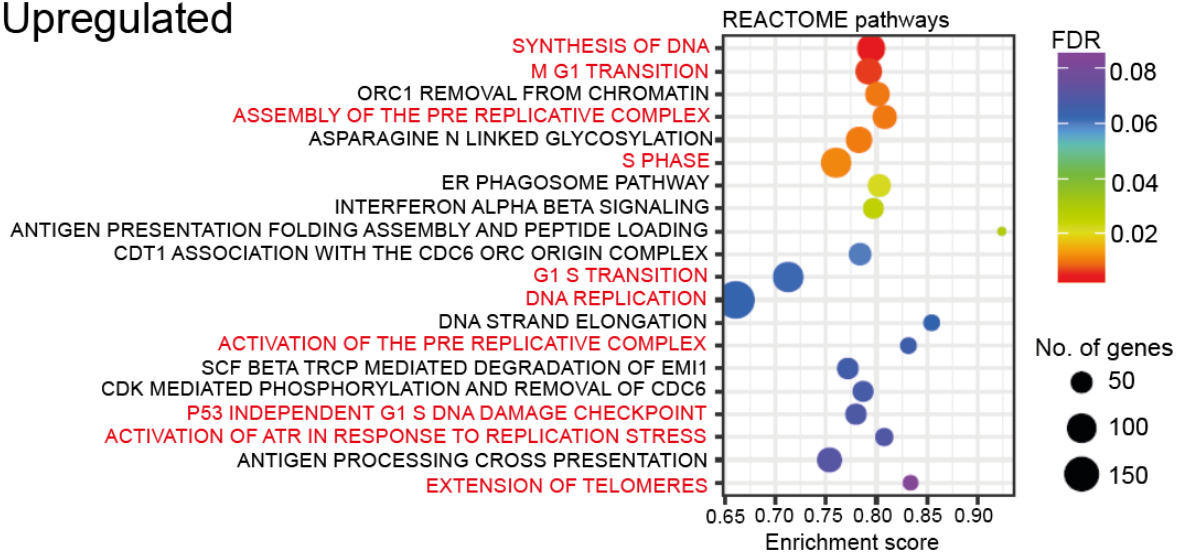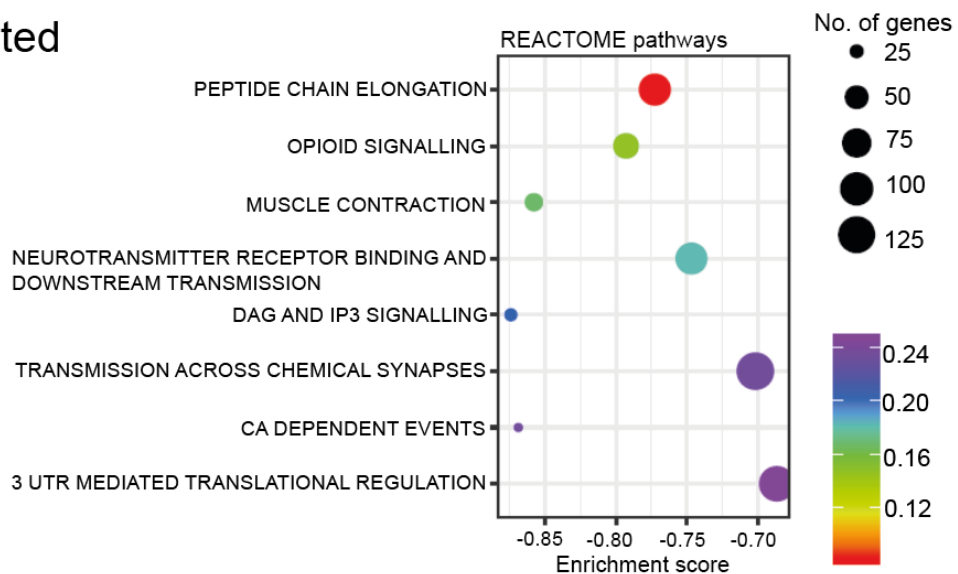
Figure 4-2 Gene set enrichment analysis (GSEA, Broad Institute, MIT) of RNAseq data after RBMX depletion. *This shows a bubble plot of upregulated and down regulated pathways in the absence of RBMX. The upregulated pathways involved in DNA replication and the cell cycle are highlighted in red.*

GSEA was used to identify any upregulated pathways obtained for RBMX knockdown versus the control. The enrichment score (ES) reflects how much a set of genes are overrepresented at the top or bottom of the pre-ranked list of genes. The graphs from the GSEA shown in Figure 4-3 demonstrate gene sets which are highly enriched in the RBMX knockdown dataset compared to the control.
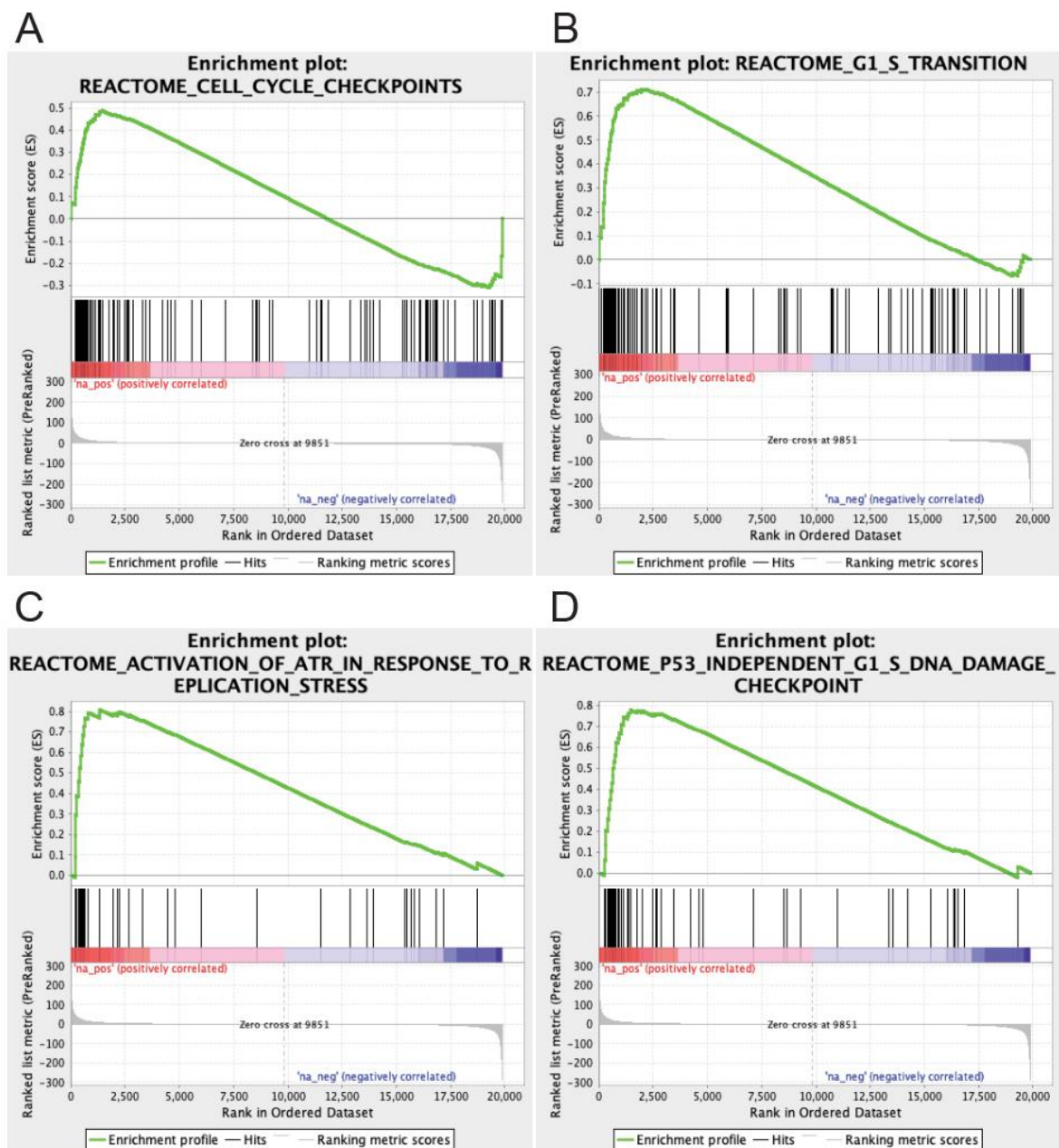
Figure 4-3 Gene set enrichment analysis (GSEA, Broad Institute, MIT) curves of pathways involved in cell cycle, DNA replication and DNA damage. *The genes that contribute the greatest to the enrichment score are represented in the leading edge genes by the black vertical lines below the graphs. The correlation with RBMX depletion, either positive or negative, is shown below the Leading Edge Gene Subset.*

4.3.3 Depletion of RBMX leads to increased use of cryptic RNA-processing sites in genes important for genome stability

The above GSEA indicated that gene expression pathways connected with genome stability changed after RBMX depletion. I next investigated the impact of RBMX knockdown in MDA-MB-231 cells on RNA processing and particularly splicing, since

RBMX has been previously identified as a splicing factor and also interacts with other splicing factor proteins. RNA-seq of three biological replicates for both RBMX-depleted and control cells were used. Preliminary differential splicing analysis using MAJIQ was done on single samples by Dr Yoseph Barash (University of Pennsylvania, Department of Computer and Information Science). Analyses using these single RNAseq datasets revealed some differential splicing targets (Hysenaj, 2020). However, splicing analyses of RNAseq data are much more powerful when performed on multiple samples. Thus triplicate RNA-seq samples were further analysed for RNA processing changes that depend on *RBMX* protein expression using SUPPA2 computational analysis by Dr Kat Cheung (Newcastle University). SUPPA2 calculates a percentage splicing inclusion (PSI) level between knockdown and control transcripts (Trincado et al., 2018). SUPPA2 predicted 6708 differentially processed RNA isoforms between the siRNA RBMX knocked down cells and siRNA control cells.

From this list of predicted splicing targets, we visually inspected the predicted events on IGV to confirm the presence of differential RNA processing and identified 611 events including alternative splice and polyadenylation (polyA) sites (these analyses were carried out by myself in collaboration with Dr. Sara Luzzi, Figure 4-4). The RNAseq replicates from control and RBMX-depleted cells were merged to enable more easy visual inspection of RNA processing events on IGV. Out of the 611 differential RNA processing events,22.75% were not previously annotated on Ensembl (https://www.ensembl.org/biomart/) (Figure 4-4, A). We also found that 51.55% events were alternatively spliced and 48.45% events were alternative 3′ end termination events. We identified several types of splicing patterns regulated by RBMX, shown in Figure 4-4, C.

A Annotated vs Non annotated RNA processing events in MDA-MB-231

Total = 611

- ■ 77.25% Annotated
- ■ 22.75% Non annotated

B Alt Splicing vs Alt Poly A RNA processing events in MDA-MB-231

Total = 611

- ■ 51.55% Splicing
- ■ 48.45% PolyA

C Types RNA processing events in MDA-MB-231

Total = 611

- ■ 22.75% Included exon
- ■ 18.82% Upstream termination site
- ■ 18.17% Upstream termination exon
- ■ 9.33% Skipped exon
- ■ 9.17% Crypic terminal exon
- ■ 6.06% Alternative first exon
- ■ 3.93% Intron retention
- ■ 3.44% Alternative 3' splice site
- ■ 2.62% Alternative 5' splice site
- ■ 1.64% Cryptic exon
- ■ 1.47% Alternative terminal exon
- ■ 1.31% Exitron
- ■ 0.82% Alternative termination site
- ■ 0.49% Mutually exclusive exons

Figure 4-4 Pie Charts representing the types of mRNA processing defects after RBMX knockdown in MDA-MB-231 cells by SUPPA2 analysis (Trincado et al., 2018)*. (A) RBMX RNA processing events that are annotated vs non-annotated events. (B) mRNA splicing events vs the polyA events. (C) Types of RNA processing events regulated by RBMX.*

The majority 82.49% of the RNA processing events we identified are repressed by RBMX (Figure 4-5). In Figure 4-5, B and C we show the detection frequency of different types of RNA processing events repressed and activated by RBMX.

63

A Activated vs repressed by RBMX RNA processing events in MDA-MB-231

Total = 611

■ 17.51% Activated
■ 82.49% Repressed

B RNA processing events Repressed by RBMX in MDA-MB-231

Total = 504

C RNA processing events Activated by RBMX in MDA-MB-231

Total = 107

■ Exon inclusion
■ Alt. last exon
■ Alt. polyA site
■ Alt. first exon
■ Intron retention
■ Alt. 5' splice site
■ Alt. 3' splice site
■ Exitron
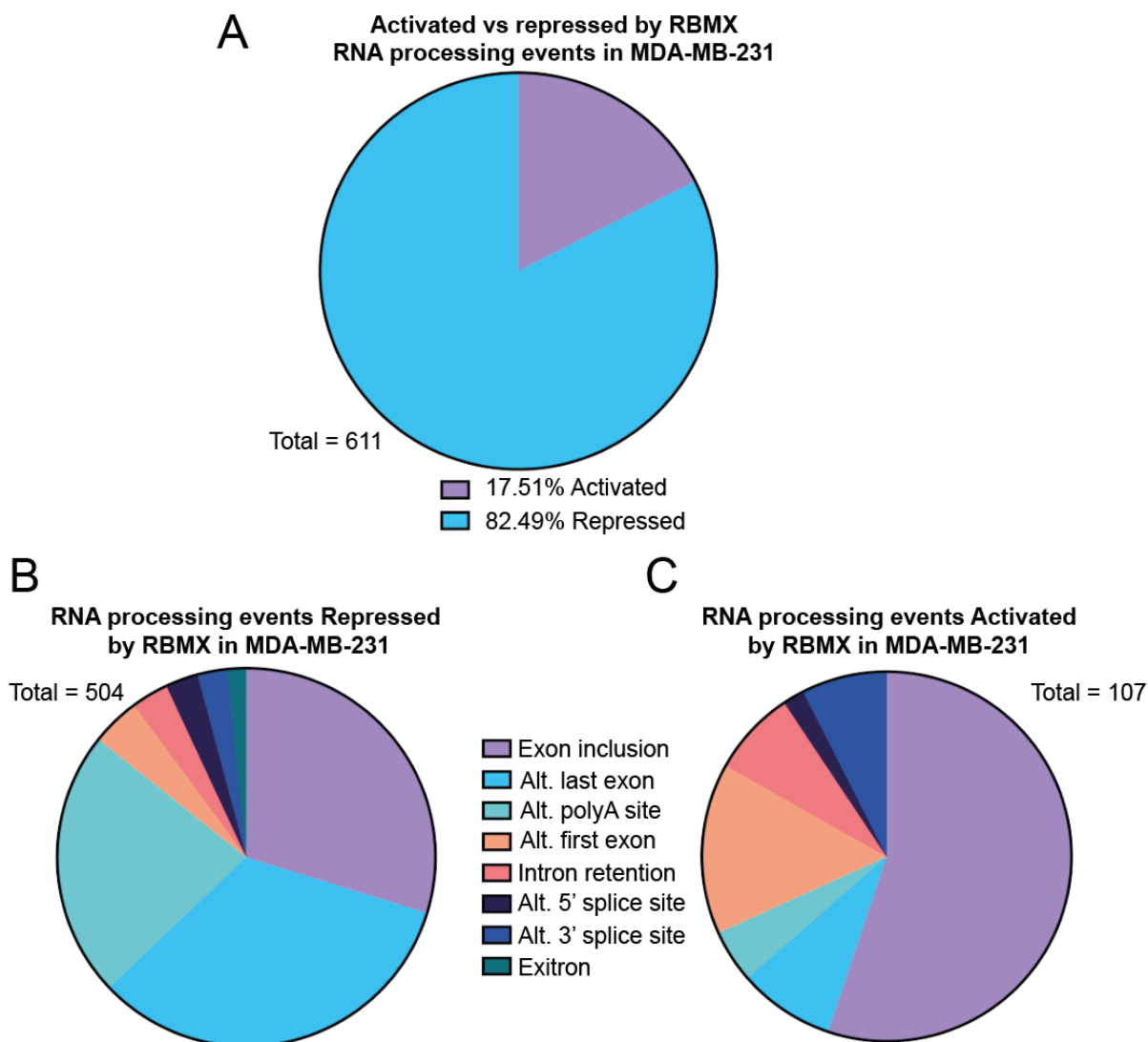
Figure 4-5 Pie charts representing the proportion of events either activated or repressed by RBMX after RBMX knockdown in MDA-MB-231 cells. (A) Total RNA processing events (611 events). (B) Repressed events. (C) Activated events.

4.3.3.1 Correlation of RNA processing events with exons with RBMX binding sites identified by iCLIP

I next wanted to find out if RBMX was directly binding near or within regulated exons/introns using the iCLIP data from chapter 3. I thus compared the iCLIP and RNA-seq datasets to identify RBMX binding sites in RBMX regulated exons.

The RBMX iCLIP data produced in chapter 3 was done in HEK293 cells but the RNA-seq data produced in this chapter was done in MDA-MB-231 breast cancer cells. In case of cell type specific differences in RNA processing I next searched for RNA

processing events regulated by RBMX that were present in HEK293 and MDA-MB-231 cells. This was possible using an online dataset containing RNAseq data after RBMX knockdown in HEK293 cells (Liu et al., 2017). Using this dataset I (with Dr Sara Luzzi) identified 63% (385 events) of RNA processing events in common between MDA-MB-231 and HEK 293 cells (Figure 4-6). Using the iCLIP data RBMX binding sites were identified by visual inspection on IGV in 26% (159 events) of RBMX regulated exons (Figure 4-6).
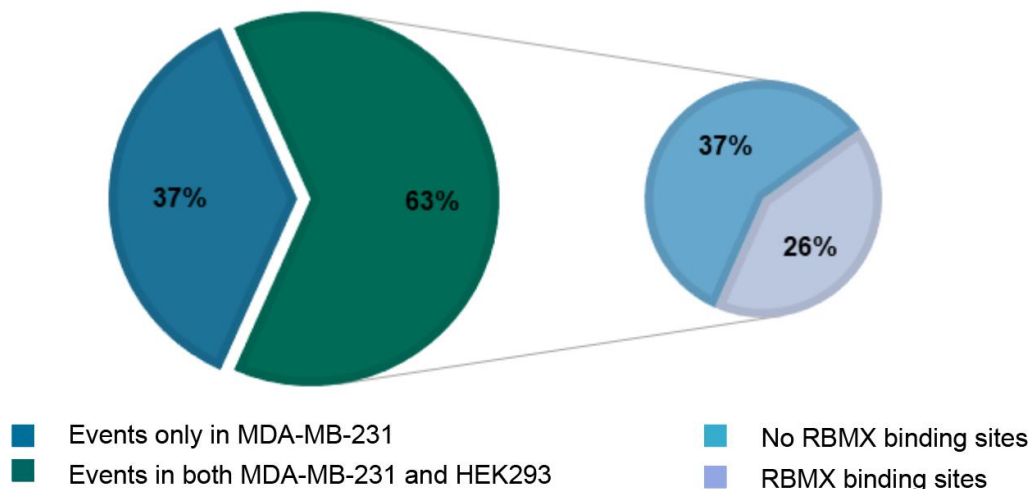
Figure 4-6 Pie charts representing the RBMX regulated RNA processing events in common between HEK293 and MDA-MB-231 with RBMX binding sites. *Left panel = Pie chart representing RBMX RNA processing events only in MDA-MB-231 (37%) and those in common between HEK293 and MDA-MB-231 (63%). Right panel = Pie chart representing events in common between HEK293 and MDA-MB-231 with RBMX binding sites (26%) and without RBMX binding sites (37%).*

During the identification of RNA processing events we noticed that RBMX was regulating a set of splicing events within exons over 1000kb which we have termed ultra long exons. The median exon size of all human exons is 120bp. The ultra-long RBMX regulated exons are much higher than this median size. I compared all the exon sizes that had RBMX RNA processing events and all iCLIP bound exons with the size of all the human exons. With the help of Dr Sara Luzzi, I showed that the median size of exons regulated by RBMX and exons bound by RBMX was significantly higher than the median sixe of all human exons. This analysis is shown in Figure 4-7.

RNAseq: events common in MDA-MB-231 and HEK293
iCLIP: exons with multiple tags were counted once
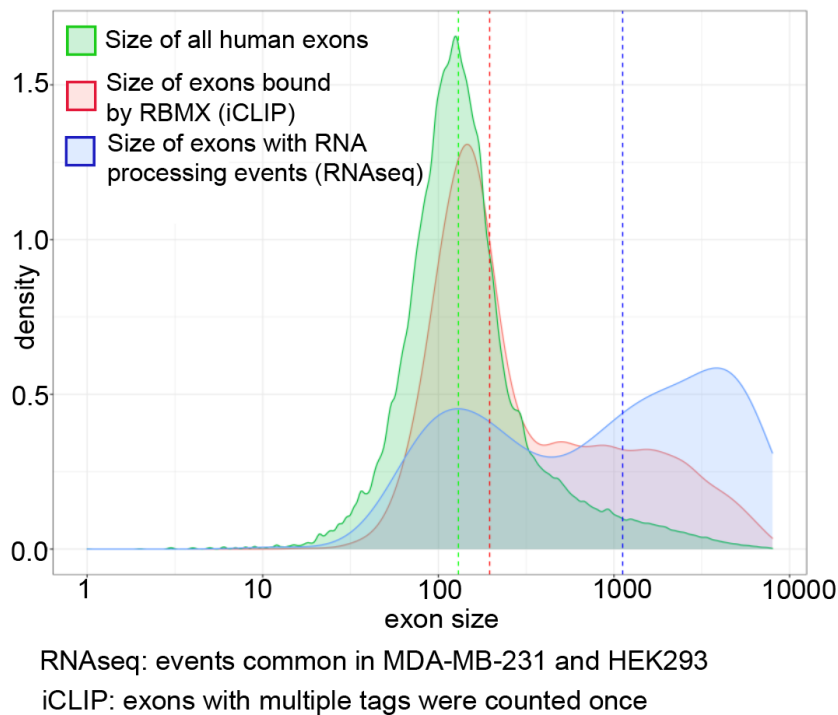
Figure 4-7 RBMX binds and regulates RNA processing events in ultralong exons.
Distribution plots representing the sizes of all human exons (shown in green), and exons bound (shown in red) and regulated by RBMX (shown in blue). Green = all human exons. The dotted green line denotes the median exon size (120 nt). Red = all RBMX bound genes. The dotted red line denotes the median exon size. Blue = RBMX RNA processing events in exons that are regulated by RBMX and that were identified by RNAseq in MDA-MB-231 and HEK293 cells. The dotted blue line denotes the median exon size.

4.3.3.2 RBMX protein represses cryptic splice sites in genes important for genome maintenance

To identify the pathways affected by the depletion of RBMX we carried out Gene Ontology analysis of the genes with differential RNA processing in RBMX depleted and control cells. The Gene Ontology analysis was performed by Dr Katherine James (Northumbria University) and identified enrichment of genes involved in meiotic cell cycle, replication fork processing, sister chromatic separation and meiotic replication Figure 4-8. Some of the target exons that are functionally important with regard to RBMX binding are shown in **Table 4-1** with the exon size and RNA processing event. These exons include within the *ETAA1* (*Ewings Tumour Associated 1*) gene, where RBMX directly binds to exon 5 to enable full-length exon inclusion by blocking selection of a cryptic 3′ splice site.

66

GO Terms

- replication fork processing
- meiotic cell cycle process
- double−strand break repair via homologous recombination
- DNA−dependent DNA replication maintenance of fidelity
- meiotic cell cycle
- recombinational repair
- meiotic nuclear division
- sister chromatid segregation
- carbohydrate derivative catabolic process
- DNA−dependent DNA replication
- nuclear chromosome segregation
- double−strand break repair

Figure 4-8 Chord plot showing Gene Ontology of enriched terms involving cell cycle and DNA repair in genes with RBMX regulated exons.

Table 4-1 Functionally important RBMX RNA processing targets

Gene	Function	RNA processing event	Exon length (nt)
NFX1	NFX1 (Nuclear Transcription Factor, X-Box Binding 1) is a DNA-binding transcription factor involved in the apoptosis and survival caspase cascade pathway (Xu et al., 2010)	Alternative terminal exon repressed by RBMX	1068
LRIF1	LRIF1 (Ligand Dependent Nuclear Receptor Interacting Factor 1) is a retinoic acid receptor binding. LRIF1-HP11α interaction is important for accurate cell division and mitotic cell division (Akram et al., 2018).	Skipped alternative exon activated by RBMX	1527
ATRX	ATRX (alpha thalassemia/mental retardation syndrome X-linked) protein is part of the chromatic remodelling family of proteins. It facilitates DNA replication and is involved in gene regulation at interphase (Nandakumar et al., 2017)	Exitron repressed by RBMX	3097
ABLIM3	ABLIM3 (actin-binding LIM) protein interacts with actin filaments and is involved in cell signalling pathways (Krupp et al., 2006).	Cryptic exon repressed by RBMX	2336
ETAA1	ETAA1 (Ewing'S Tumor-Associated Antigen 1) protein accumulates at sites of DNA damage and promotes progression at replication forks (Feng et al., 2016).	3'ss repressed by RBMX	2111
BRCA2	BRCA2 is an essential for DNA repair mediated by homologous recombination, it protects stalled replication forks from degradation (Andreassen et al., 2021).	Alternative last exon repressed by RBMX	4924
KNL1	KNL1 (kinetochore scaffold 1) is an important cell cycle protein, it is a large protein required for the assembly of the outer kinetochore proteins (Kern et al., 2015).	Alternative last exon repressed by RBMX	4987

4.3.4 Splicing analysis of endogenous target exons following depletion of RBMX breast cancer and lung cancer cell lines

The above bioinformatics analysis using GSEA, SUPPA2 and Gene Ontology indicated that RBMX is involved in the mRNA processing of genes important for genome maintenance. Strong splicing defects after RBMX depletion in the *ETAA1* and *REV3L* mRNAs were identified and verified by RT PCR to distinguish the mis-regulated RNA processing events. The RT-primers were designed to be complementary to flanking constitutive exons and an internal primer within the alternative exons of interest when required (if the amplified product for the alternative event would be otherwise too long to detect).

Splicing inclusion of cryptic mRNA processing events normally repressed by RBMX could severely impact the expression of genes involved in DNA repair and cell cycle control. Some genes such as *CD44*, *SGO2* and *TNC* had more than one mis-regulated RNA processing event. Analysis of exon junction reads on IGV are shown as indicated for some of the RBMX regulated exons in *ETAA1* (**Figure 4-9**, A), *REV3L* (Figure 4-13, A), *SGO2* (Figure 4-14, A), *ATRX* (Figure 4-15), *BRCA2* (Figure 4-16, A) *and KNL1* (Figure 4-16, B). I have used the iCLIP data to identify RBMX binding sites within or near the regulated exons. RBMX binding sites identified by iCLIP were seen in close proximity to the RNA processing events regulated by RBMX in these genes.

4.3.4.1 Repression of cryptic splice site selection in ETAA1 exon 5 is associated with a high density of RBMX binding

Sashimi plots of *ETAA1* transcripts showed that RBMX depletion resulted in increased selection of a weak 3′ splice site in the 2,111bp long exon 5. *ETAA1* encodes a protein involved in replication fork processing. ETAA1 binds to RPA and activates the ATR DNA repair pathway. In (Figure 4-9, A), RNAseq data from RBMX knockdown cells are coloured blue compared to RNAseq from wild type cells that are coloured red. RBMX depleted cells had increased use of the weak splice site compared to control cells. The use of the weak splice site in ETAA1 exon 5 after RBMX depletion was verified by RT-PCR in 3 different cancer cells lines: MDA-MB-231 (a triple negative breast cancer cell line), MCF7 (an estrogen responsive breast cancer cell line) and NCI-H520 (a squamous cell carcinoma lung cancer cell line) (Figure 4-9, C). iCLIP identified many

RBMX binding sites within *ETAA1* were concentrated in exon 5 compared to the rest of the gene.
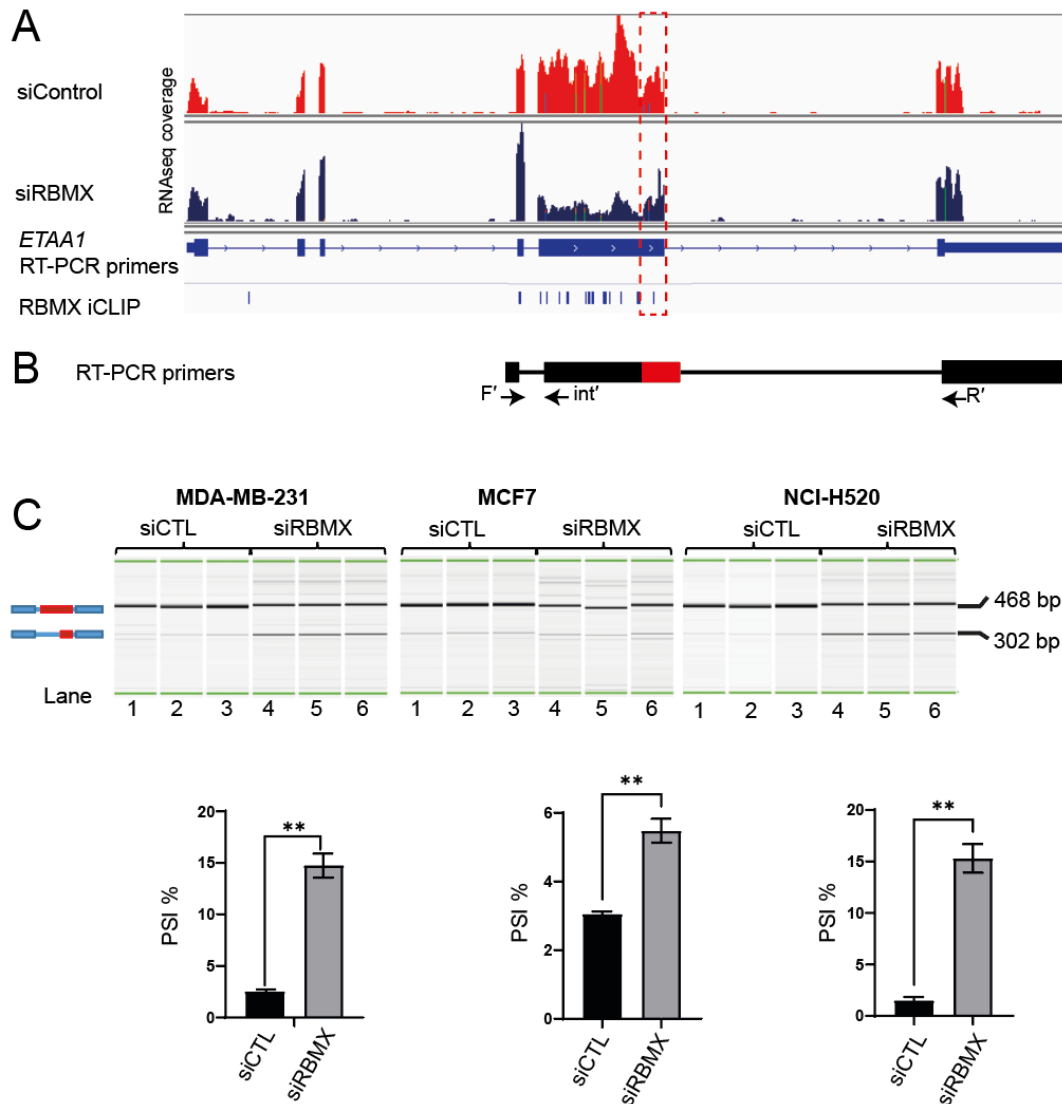
*Figure 4-9 RBMX depletion resulted in increased use of the weak splice site of ETAA1 exon 5 in MDA-MB-231, MCF7 and NCI-H520 cancer cell lines (A) IGV snapshot of the RNAseq reads from MDA-MB-231 cells. RNASeq coverage represents merged tracks from three (MDA-MB-231) biological replicates. Reads from the siRNA control cells are shown in red and the RBMX siRNA knockdown RNAseq reads are shown in blue. The ETAA1 Ensembl gene transcript isoforms indicate known isoforms of ETAA1. RBMX iCLIP tracks are shown below the gene. (B) Schematic showing the position of the RT-PCR primers. (C) Representative QIAxcel capillary gel electrophoretograms showing RT-PCR validation of the changing ETAA1 splicing pattern in 3 different cancer cell lines. The PSI was calculated from three biological replicates from negative control siRNA (black) and RBMX siRNA (grey). Lines in the bar plots represent mean ± SEM. P-values were produced by unpaired t-test, on GraphPad Prism v 9.31 where ** pvalue <0.05.*

This switch to a cryptic splicing pattern compromised ETAA1 protein expression. The use of the weak splice site reduces the size of *ETAA1* exon 5 included in the transcript. The amount of full-length ETAA1 protein was greatly reduced in the RBMX depleted cells compared to controls as shown in Figure 4-10. A western blot was carried out to detect the effect of RBMX depletion in MDA-MB-231 cells. No full-length ETAA1 protein was detected after RBMX depletion. The antibody used detects an epitope within the region *ETAA1* exon 5 spliced out in RBMX depleted cells.
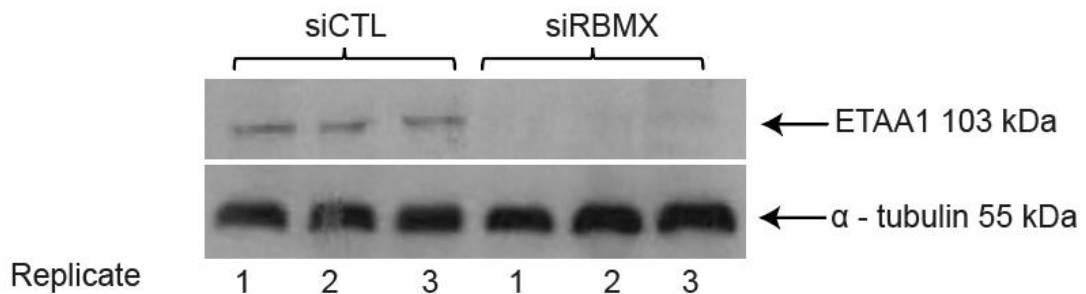
Figure 4-10 Western blot showing depletion of full-length ETAA1 protein after RBMX knockdown in MDA-MB-231 cells with an α-tubulin loading control.

ETAA1 exon 5 was identified as an RBMX target using siRNA depletion experiments. I also investigated the reciprocal conditions using the cell lines that enable RBMX over-expression, and that I had used for the iCLIP analysis. RBMX was over expressed in RBMX-FLAG Flp-In HEK293 cells and Flp-In HEK293 cells were used as a negative control. When RBMX was over expressed, I observed a splice site switch from the cryptic 3′ to the normal splice site (Figure 4-11). This resulted in increased inclusion of the full exon and reduced use of the cryptic 3′ splice site in the RBMX over expressed cells compared to controls. These experiments show that the cryptic splice site in *ETAA1* is dynamically responsive to RBMX protein levels.
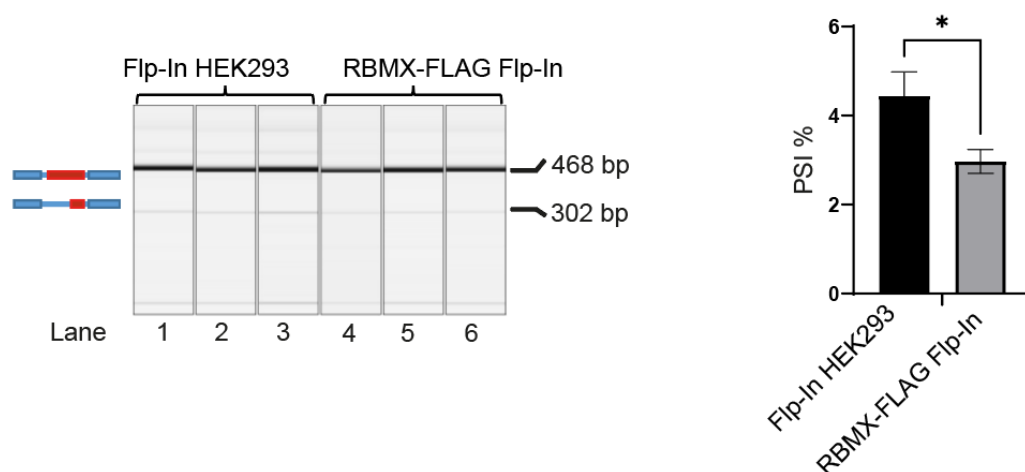
***Figure 4-11 The cryptic splice site in ETAA1 is repressed by increased
expression of RBMX. The cryptic splice site in ETAA1 is repressed by increased
expression of RBMX.*** *Representative QIAxcel capillary gel electrophoretograms
showing RT-PCR validation of the changing ETAA1 splicing pattern in HEK293 cells
overexpressing FLAG-RBMX. The PSI was calculated from three biological replicates
from negative control siRNA (black) and RBMX siRNA (grey). Lines in the bar plots
represent mean ± SEM. P-values were produced by unpaired t-test, on GraphPad
Prism v 9.31 where * pvalue <0.05.*

4.3.4.2 Experiments to analyse if ETAA1 exon 5 is also controlled by Tra2β

Tra2β/α are SR proteins that have been previously connected with RBMX in RNA
splicing control. RBMX has previously been previously shown to have an antagonistic
effect on splicing of some Tra2β-target exons expressed from minigenes (Nasim et
al., 2003). Alignment of iCLIP tags from Best et al., 2014 showed that ETAA1 exon 5
also contained iCLIP tags for Tra2β, indicating this exon may also be controlled by
Tra2β. To test this, I used siRNAs to knock down bothTra2β and Tra2α
(becauseTra2α can compensate for the loss of Tra2β). RT-PCR using RNA samples
purified from double knockdown cells and control cells indicated that Tra2β/α
depletion resulted in an increased the use of the 3′ cryptic splice sites in MDA-MB-
231, MCF7 and NCI-H520 cell lines (Figure 4-12). The use of the cryptic splice site is
higher in MDA-MB-231 cells than in MCF7 and NCI-H520. These data confirm that
splicing of ETAA1 exon 5 is also controlled by Transformer2 proteins.
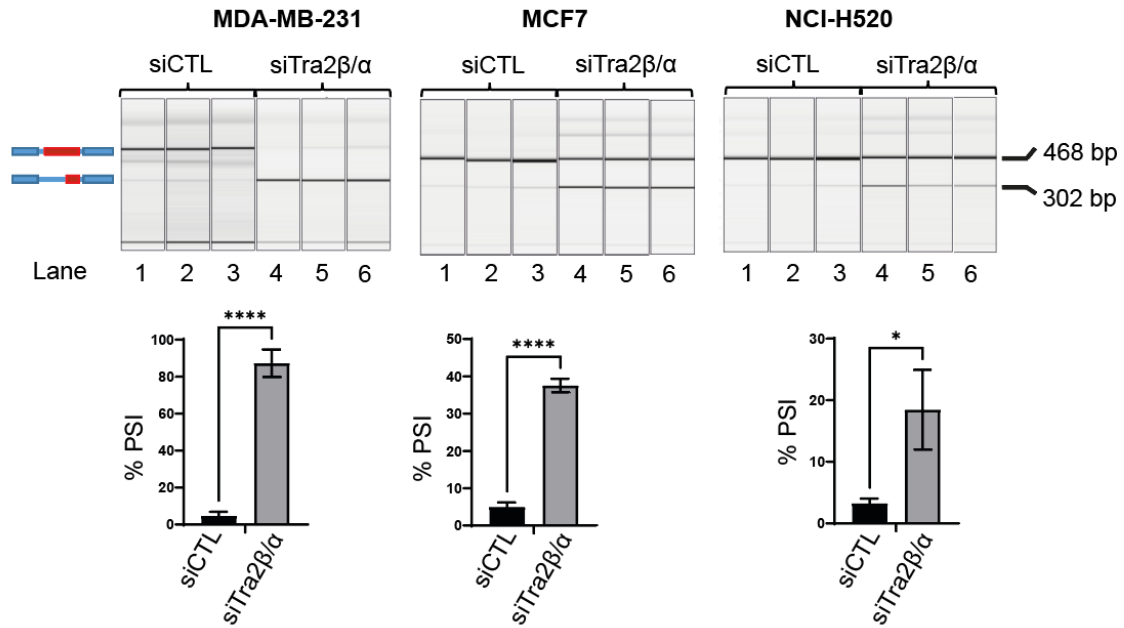
Figure 4-12 Tra2β/α joint depletion resulted in increased use of the weak splice site of ETAA1 exon 5 in MDA-MB-231, MCF7 and NCI-H520 cancer cell lines. *Representative QIAxcel capillary gel electrophoretograms showing RT-PCR validation of the changing ETAA1 splicing pattern in 3 different cancer cell lines. The PSI was calculated from three biological replicates from negative control siRNA (black) and RBMX siRNA (grey). Lines in the bar plots represent mean ± SEM. P-values were produced by unpaired t-test, on GraphPad Prism v 9.31 where ****pvalue < 0.0001, *pvalue = 0.0156.*

4.3.4.3 Repression of a cryptic splice site in REV3L exon 14 is associated with high levels of RBMX binding to RNA

In addition to *ETAA1*, I identified a more extensive list of genes involved in cell cycle, DNA damage and replication pathways which were differentially processed in the absence of RBMX. One of these, the *REV3L gene*, encodes for the catalytic subunit of DNA polymerase zeta complex involved in protection against DNA damage. In the *REV3L* gene, RBMX knockdown increased use of a 3′ cryptic splice site within the 4,162bp long exon 14 (Figure 4-13, A). Using the iCLIP data, multiple RBMX binding sites mapped within this RBMX regulated exon 14 of *REV3L*. I was able to validate increased use of the 3′ cryptic splice site within exon 14 of *REV3L* by RT-PCR in 3 different cell lines, MDA-MB-231, MCF7 and NCI-H520 (Figure 4-13, C).
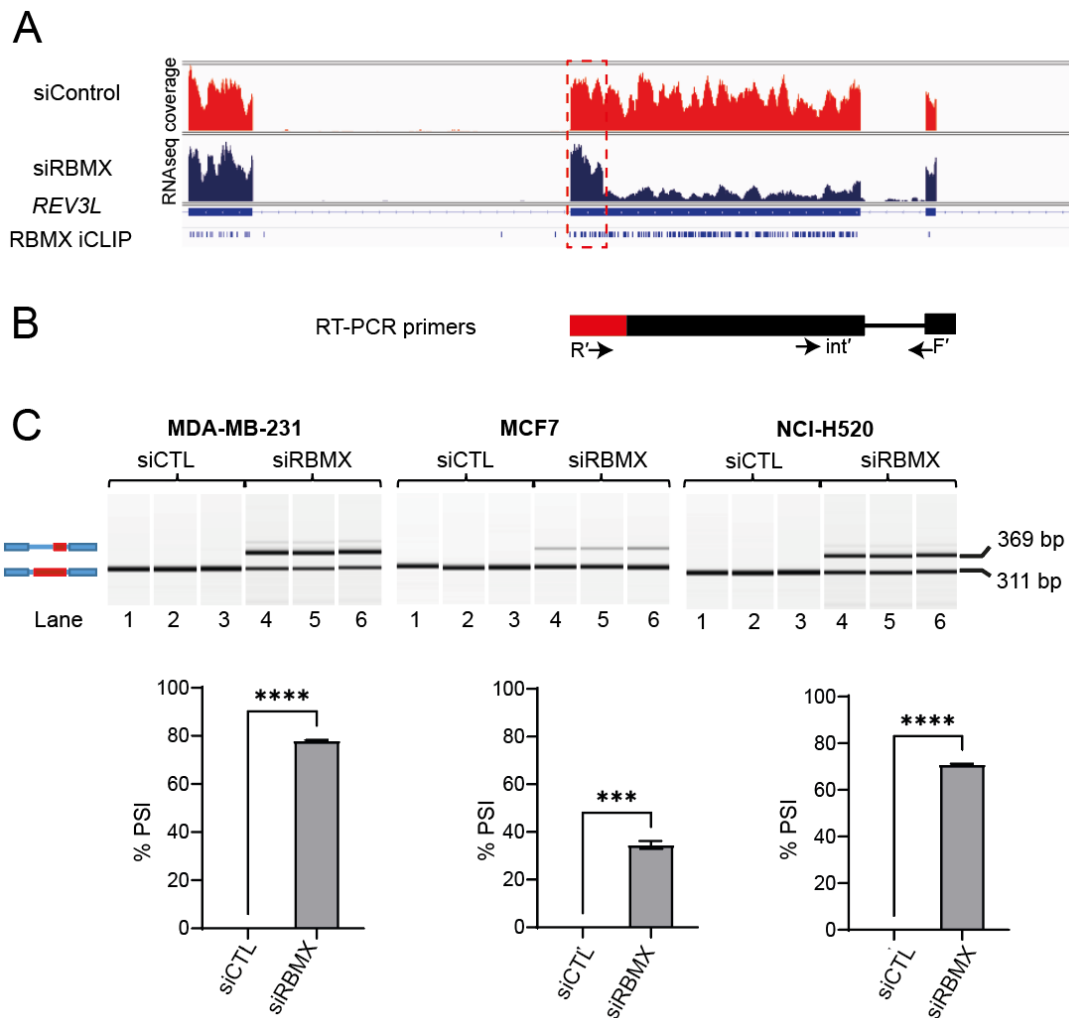
Figure 4-13 RBMX depletion resulted in increased use of the weak splice site of REV3L exon 14 in MDA-MB-231, MCF7 and NCI-H520 cancer cell lines *(A) IGV snapshot of the RNAseq reads from MDA-MB-231 cells. RNASeq coverage represents merged tracks from three (MDA-MB-231) biological replicates. Reads from the siRNA control cells are shown in red and the RBMX siRNA knockdown RNAseq reads are shown in blue. The REV3L Ensembl gene transcript isoforms indicate known isoforms of REV3L. RBMX iCLIP tracks are shown below the gene. (B) Schematic showing the position of the RT-PCR primers. (C) Representative QIAxcel capillary gel electrophoretograms showing RT-PCR validation of the changing REV3L splicing pattern in 3 different cancer cell lines. The PSI was calculated from three biological replicates from negative control siRNA (black) and RBMX siRNA (grey). Lines in the bar plots represent mean ± SEM. P-values were produced by unpaired t-test, on GraphPad Prism v 9.31 where **** pvalue < 0.0001, *** pvalue= 0.0007.*

4.3.4.4 Reduced shugoshin2 protein detected in RBMX depleted cells

RBMX depletion caused increased selection of 2 normally weakly used RNA processing sites in the *SGO2* gene (Figure 4-14, A), which encodes a protein important for meiotic sister chromatid cohesion. RBMX depletion caused use of an alternative termination site in the 232bp long exon 6 and an exitron in the 2,927bp long exon 7 in

SGO2 (Figure 4-14, A). There was high RBMX binding within *SGO2* exon 7 and 1 binding site in exon 6. Western analysis showed that RBMX depletion caused changes to SGO2 protein expression: When RBMX was depleted reduced levels of full length SGO2 protein were seen compared to controls (Figure 4-14, B). This could also be related to decreased levels of SGO2 mRNA expression after RBMX depletion (notice junction reads are reduced also after RBMX depletion, indicating a general down regulation of *SGO2* gene expression). Two different antibodies recognising two different epitopes were used for western blot protein detection.
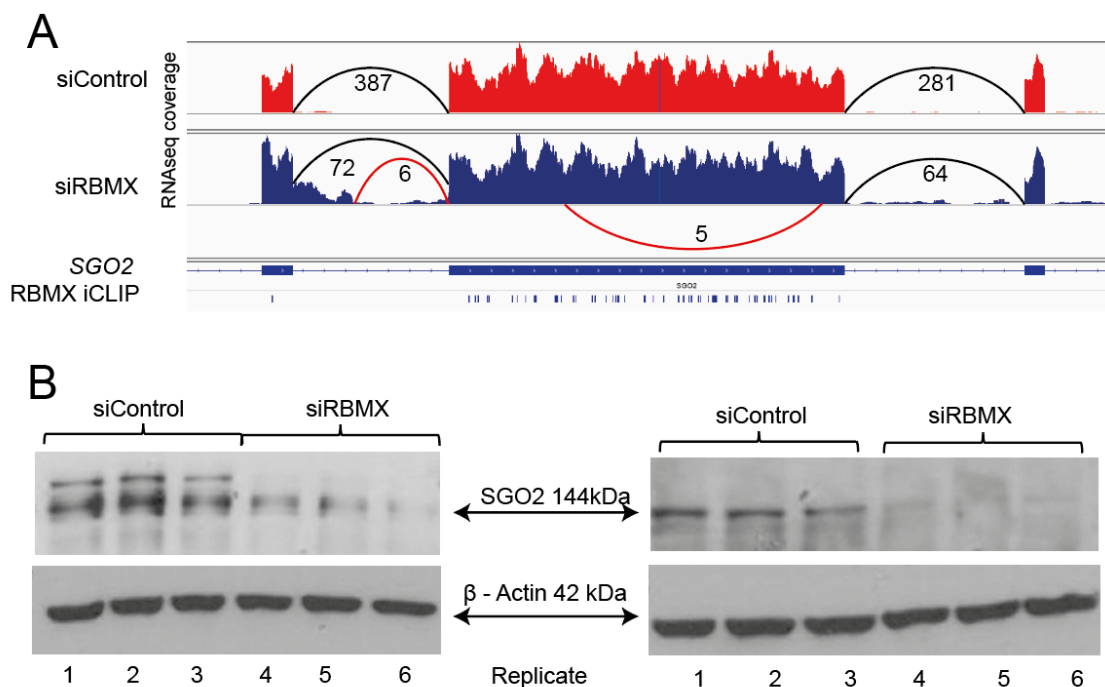
Figure 4-14 Validation of SGO2 exon 7 as an RBMX target exon in HEK293 and cancer cell lines *(A) IGV snapshot of part of the SGO2 gene showing RNAseq data obtained from MDA-MB-231 cells. RNASeq coverage represents merged tracks from three (MDA-MB-231) biological replicates. The RNAseq from the siRNA control samples are shown in red and RNAseq data from the RBMX knockdown cells shown in blue. Splice junctions are shown as curved lines with associated numbers of splice junction reads, and the alternative RNA processing splice junctions used in the presence absence of RBMX are shown in red. The SGO2 Ensembl gene transcript isoforms indicate known isoforms of SGO2. RBMX iCLIP tracks are shown below the gene. (B) Western blot showing depletion of full length SGO2 protein compared to levels of the β-actin loading control after RBMX knockdown in MDA-MB-231 cells. This experiment was performed using two different SGO2 antibodies that recognise 2 different epitopes.*

4.3.4.5 High levels of RBMX binding block exitron selection in the ATRX gene

Other genes involved in cell cycle, DNA damage and replication pathways were identified as having defective RNA processing patterns in RBMX depleted cells. RBMX depletion caused increased use of an exitron in the 3074bp long exon 9 of *ATRX*, a gene that encodes a protein that facilities DNA replication (Figure 4-15, A). Exitrons are regions of exons that switch to being introns. *ATRX* exon 9 has a high density of RBMX binding detected by iCLIP.
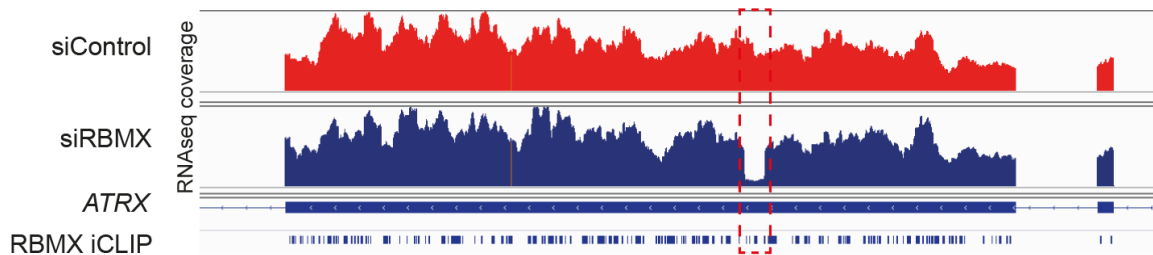
Figure 4-15 Snapshot from IGV browser of the ATRX gene. *The snapshot shows merged RNA-seq tracks from triplicate MDA-MB-231 cells after RBMX knockdown in blue and control cells in red and RBMX iCLIP binding sites.*

4.3.4.6 RBMX blocks the selection of cryptic polyA sites

The formation of 3′ ends of mRNAs (via polyadenylation) occurs during RNA processing (see Section 1.1). RBMX depletion caused increased use of alternative polyadenylation sites in the 4932bp long exon 11 of *BRCA2* (Figure 4-16 A) and 5001bp long exon 10 of *KNL1* (Figure 4-16, B). Both *BRCA2* exon 11 and *KNL1* exon 10 are ultra-long exons that had high concentrations of RBMX binding sites, indicating that they are bound by RBMX within cells. *BRCA2* encodes a protein that is involved in the repair of double stranded DNA breaks, *KNL1* encodes for a protein involved in spindle-assembly checkpoint signalling and correct chromosome segregation.
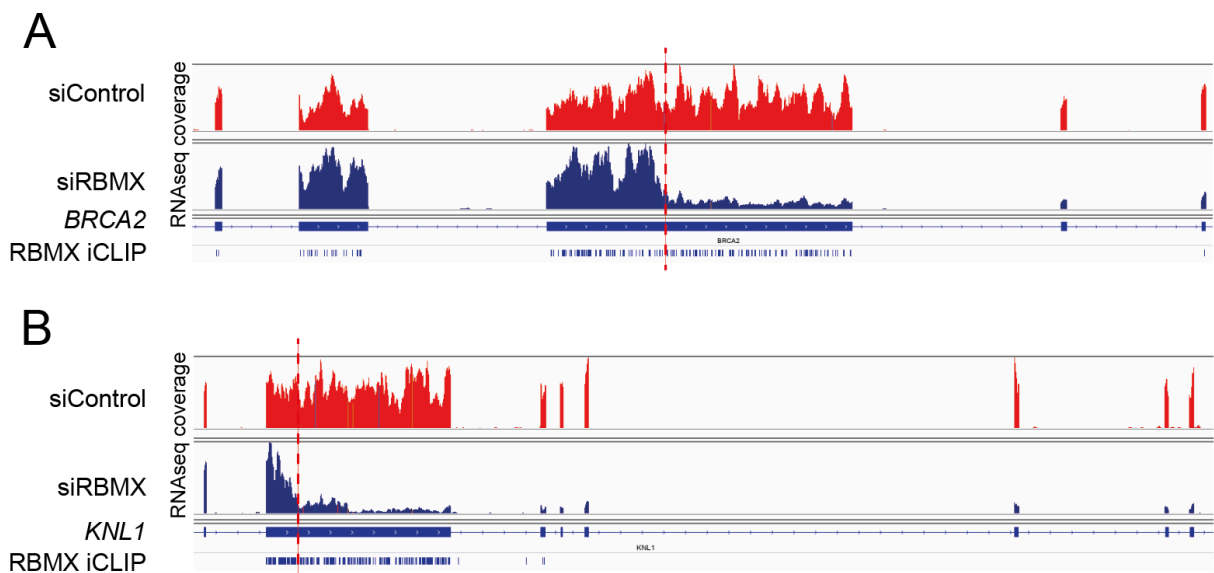
Figure 4-16 Snapshots from IGV browser of the mRNA transcripts taken from MDA-MB-231 cells. *RNASeq coverage represents merged tracks from three (MDA-MB-231) biological replicates siRNA control is shown in red and RBMX knockdown shown in blue. Ensembl gene transcript isoforms indicate known isoforms of (A) BRCA2 (B) KNL1 are represented. RBMX iCLIP tracks are shown below the genes.*

4.3.5 Minigene experiments suggest that RBMX and Tra2beta use their RRM domains to control the use of the weak splice site in ETAA1 exon 5

The above data showed that RBMX enables full-length inclusion of ultra-long exons by blocking selection of cryptic splice sites, and sometimes cryptic polyA sites. Many ultra-long exons are within genes involved in aspects of genome stability. For example RBMX depletion causes increased selection of a cryptic 3′ splice site in exon 5 of *ETAA1*. This is a large exon comprising 2111 nucleotides. *ETAA1* is also of particular interest since it encodes a key protein involved in DNA damage control and the replication fork. Loss of full length ETAA1 protein (demonstrated in Figure 4-10) after RBMX depletion could therefore contribute to the gene expression changes that I described at the start of this Chapter that were detected after RBMX knockdown. Using the RBMX iCLIP data I was able to identify RBMX crosslink sites close to the normal and weak splice sites. This may suggest that the regulation of ETAA1 exon 5 splicing is by direct RBMX binding. I therefore set out to test the mechanism of *ETAA1* splicing control using a minigene approach.

To investigate splicing mechanisms we used three separate minigenes with different sequence contents. In either case, exon 5 of the *ETAA1* gene and flanking intron sequences containing splicing signals were amplified by genomic PCR, and were cloned into the PXJ41 expression plasmid (see Figure 4-17 for the portion of *ETAA1* inserted into the plasmid). Successful minigene cloning was confirmed by Sanger sequencing. The minigenes were then transfected into HEK-293 cells with and without separate expression constructs.

Minigenes of 3 different sizes were designed. This first was a short minigene (295bp) containing only the weak splice site within ETAA1 exon 5 (Figure 4-17, top). The second minigene (521bp) contained more genomic DNA containing both the weak splice site and more of the exon 5 than the first minigene (Figure 4-17, middle). The third minigene (2847bp) was longer and contained both weak cryptic splice site as well as the far upstream 3′ splice site that is used when RBMX is present at normal concentrations in the cell (Figure 4-17, bottom).
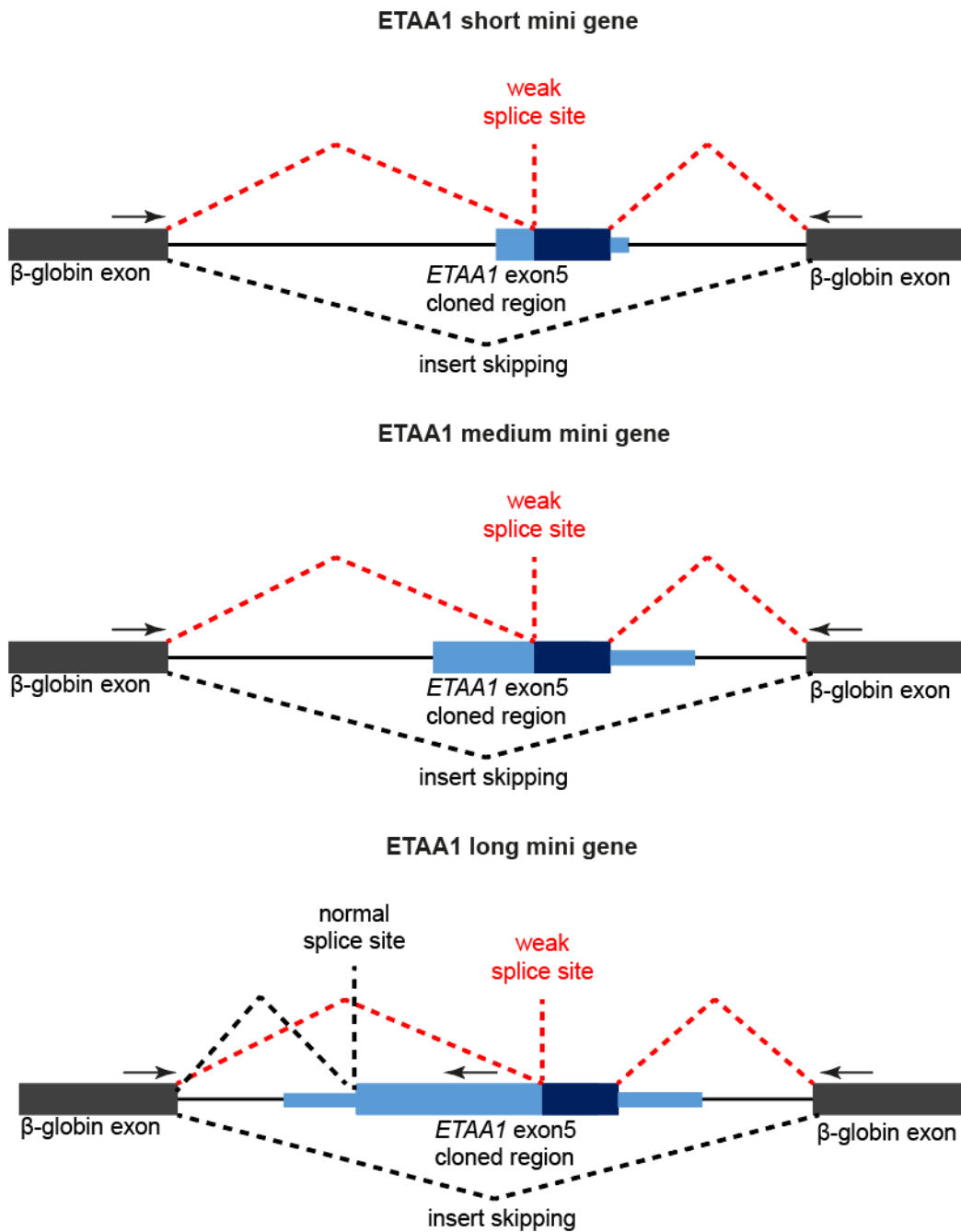
Figure 4-17 ETAA1 exon 5 mini gene design. The minigenes were cloned into the pXJ41 plasmid between 2 β-globin exons. *The top schematics are for the short and medium minigenes that only contain the weak splice site. The bottom schematic is the long minigene which contains both the normal (used for full length exon inclusion) and weak 3′ splice sites. The boxes represent exons and the line represents introns. The dotted lines represent different splicing patterns. The red dotted lines show the splicing pattern used when the weak splice site is used. The primers used for each mini gene are shown as arrows above the exons.*

The three minigenes gave different splicing patterns after transfection into HEK293 cells shown in Figure 4-18. Co-transfection of RBMX with the long minigene somewhat

increased selection of the short splice isoform that results from selection of the cryptic 3 splice site. Co-transfection of RBMX even more efficiently activated the medium version of the *ETAA1* exon 5, showing that the sequences involved in this repression must be included within this short minigene (Figure 4-18). The short mini gene was not used for further transfections, as the use of the cryptic exon was not seen. It appears the short mini gene blocks use of the cryptic splice site. This is perhaps because it was missing important sequences for cryptic splice site selection. In addition a higher band not seen in the other constructs is present in the small mini gene samples, this could be because of intron retention in the absence of splicing signals. There is also some non-specific amplification seen in all the GFP controls. This band was not seen in further transfections (Figure 4-21, Figure 4-22). RBMX is not seen to cause repression of the *ETAA1* medium and large minigenes.
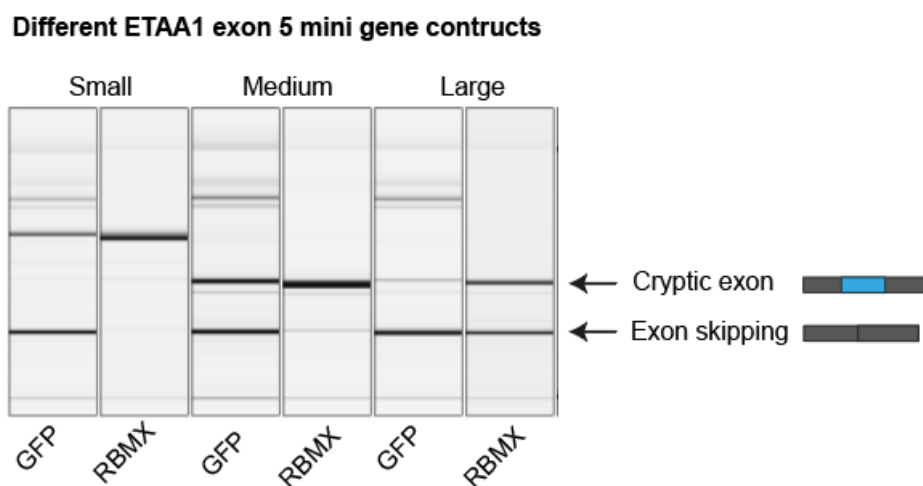
Figure 4-18 QIAxcel electrophoretogram showing splicing patterns of the ETAA1 minigenes. *The three minigenes (small, medium and large constructs) containing the cryptic exon and flanking intron sequences in HEK293 cells after transfection of ETAA1 mini genes with either GFP control or RBMX expression vectors.*

I next wanted to use minigenes to analyse the mechanism of *ETAA1* exon 5 splicing in more detail. Some evidence suggests that RBMX and Tra2β interact with each other to regulate the splicing of exons (Venables et al., 2000). A previously published analysis from our group (Best et al., 2014) provided Tra2-β iCLIP data that identified Tra2β binding sites within the *ETAA1* gene. The binding sites in exon 5 were upstream in the gene and downstream near the 3′ splice site that is repressed. Furthermore,

RNAseq data from this same study (Figure 4-12) showed that depletion of Tra2-β also affects splicing of ETAA1 exon 5.

To test the mechanism of *ETAA1* exon 5 splicing control, the minigenes were co-transfected with expression vectors for encoding GFP-only as a control. This showed the "normal" splicing pattern from the co-transfected minigenes. To dissect the mechanism of RBMX regulation, I also co-transfected *RBMX*-GFP, *RBMX* ΔRRM-GFP, *RBMX* ΔCterm-GFP (see **Figure 4-19**). These three constructs would show the effect of RBMX protein on the splicing pattern, and to what extent this was due to the RRM or the C-terminal regions of the RBMX proteins. I also co-transfected expression constructs encoding *RBMXL2*-GFP and *RBMXL2*-ΔRRM-GFP (see **Figure 4-19**) expression vectors to determine whether splicing regulation of *ETAA1* exon 5 could also be controlled by RBMXL2.
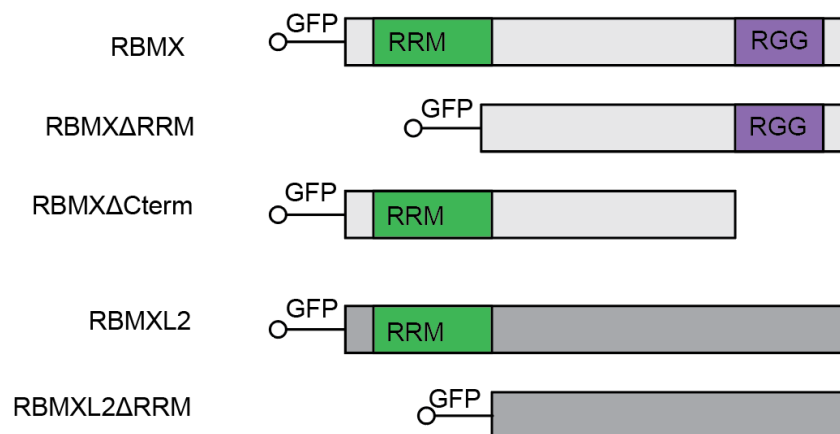
Figure 4-19 Domain structures of wild-type proteins, and proteins missing domains of RBMX and RBMXL2 proteins tagged with GFP. The structures of RBMX, RBMXΔRRM, RBMXΔCterm, RBMXL2 and RBMXL2ΔRRM are shown.

I also investigated the mechanism of Tra2β splicing control by co-transfecting *Tra2β*-GFP, *Tra2β* ΔRNP1-GFP, *Tra2β* ΔRNP1/2-GFP, *Tra2β* ΔRS1-GFP and *Tra2β* ΔRS2-GFP (see **Figure 4-20**). These experiments were aimed at establishing the effects of full length Tra2β protein on splicing patterns from the minigene constructs and to what extent these depended on the RS domains and the RRM of Tra2β. RS1 and RS2

domains are important for interaction with other proteins and splicing (David J. Elliott et al., 2000).
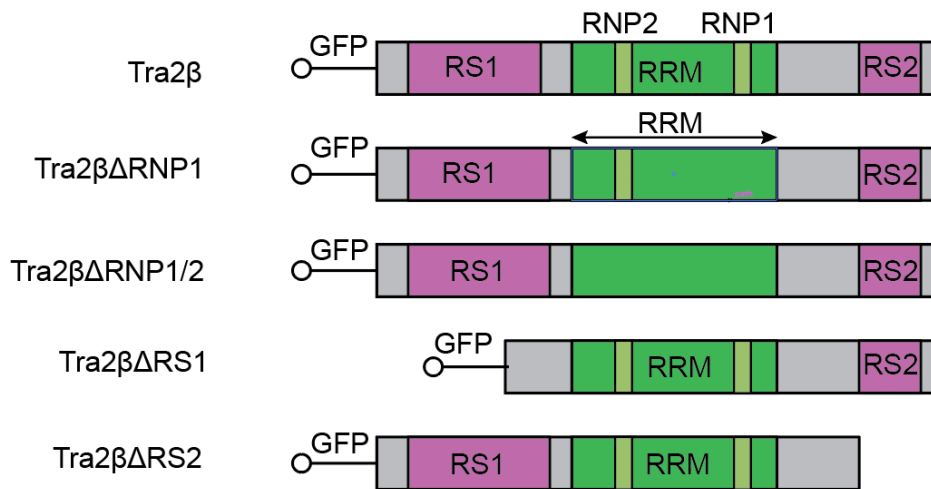
Figure 4-20 Domain structure of wild-type and missing domains of Tra2β proteins tagged with GFP expressed after transfection. *The structures of Tra2β, Tra2βΔRNP1, Tra2βΔRNP1/2, Tra2βΔRS1 and Tra2βΔRS2 are shown.*

Successful transfection of the vectors was confirmed by microscopy and western blot for the GFP tagged proteins. The vectors were transfected in different combinations (see Figure 4-21 and Figure 4-22). Co transfection with some of the expression vectors induced visible changes in mRNA splicing patterns from the minigene constructs. All minigene experiments were repeated a minimum of 3 times or more.

4.3.5.1 Splicing analysis of the ETAA1 exon 5 "medium size insert" mini gene

I first carried out splicing analysis of the *ETAA1* medium-sized insert mini gene. The medium sized minigene results shown in lanes 3 and 4 Figure 4-18. Co-transfection of *RBMX* caused slightly reduced inclusion of the cryptic exon (Figure 4-21, lane 2). A bigger decrease in cryptic exon splicing inclusion was seen when *RBMX* and *Tra2β*, and *RBMXL2* and *Tra2β* expression vectors, were co-transfected together than on their own (in Figure 4-21 compare lanes 2 and 4, lanes 5 and 7). This suggests that Tra2β cooperates with RBMX and RBMXL2 to have an inhibitory effect on the use of the weak splice site. Co-transfection of the *RBMXL2*ΔRMM construct caused more splicing inclusion of the *ETAA1* cryptic exon compared to transfection of the full length *RBMXL2*-GFP expression construct (in Figure 4-21 compare lanes 5 and 5). This

suggests that RBMXL2-RNA interactions are important for splicing regulation of the *ETAA1* cryptic exon.

I also dissected the mechanism of Tra2β splicing control of this minigene. In the absence of the RS domains co-transfected Tra2β was still able to carry out splicing control by encouraging exon skipping either when transfected alone or when transfected together with RBMX (in Figure 4-21 compare lanes 11 and 15, lanes 12 and 16). In the absence of the RNP protein binding domains, deleted versions of Tra2β co transfected with RBMX caused increased inclusion of the cryptic exon compared to when full length Tra2β was con transfected with RBMX (in Figure 4-21 compare lanes 9 and 10 with lane 4).
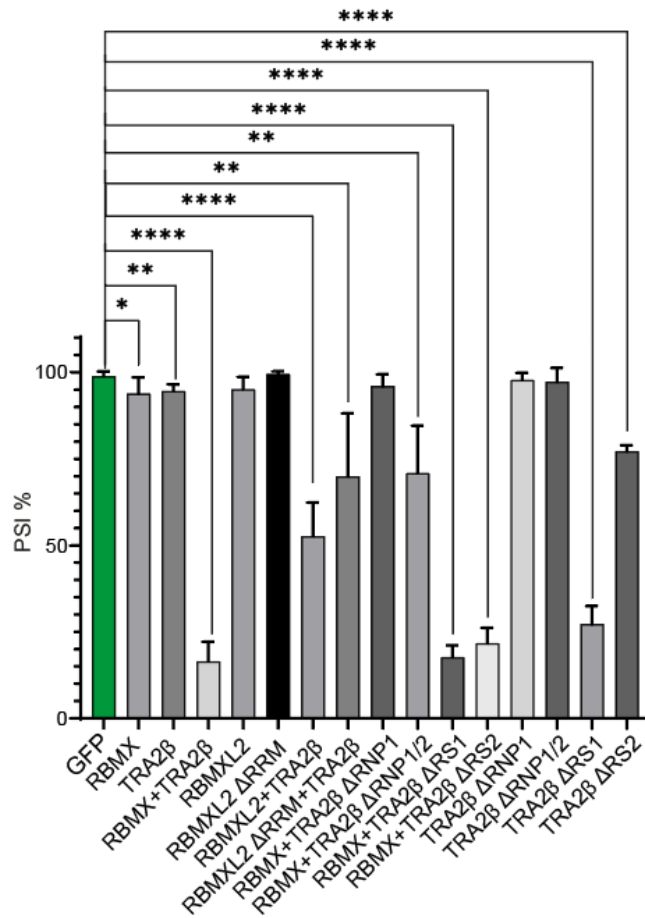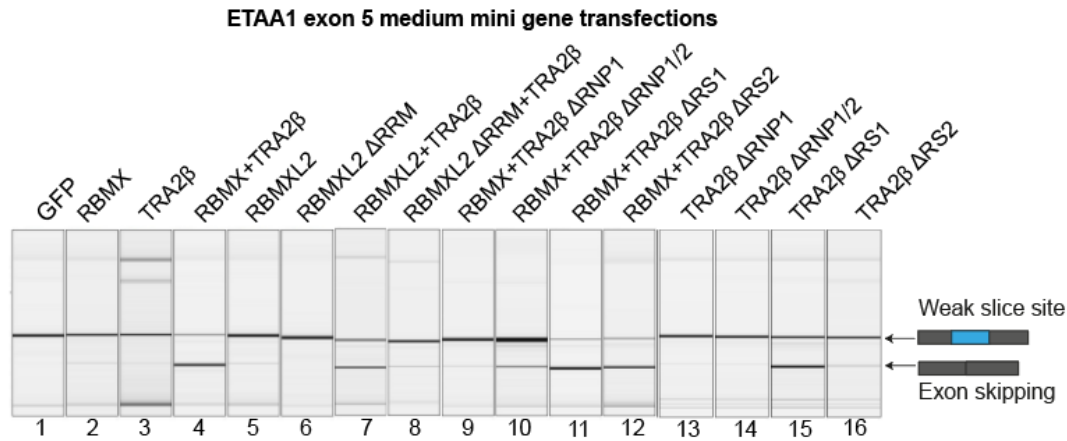
Figure 4-21 Tra2β has an antagonistic effect to RBMX on ETAA1 exon 5. These experiments used the ETAA1 "medium-sized" mini gene. *Representative QIAxcel capillary gel electrophoretograms showing RT-PCR validation of the "medium sized" minigene in the conditions indicated using PXJF and PXJB primers. The PSI was calculated from three or more biological replicates. Lines in the bar plots represent mean ± SEM. P-values were produced by unpaired t-test, on GraphPad Prism v 9.31 where*, **, ****, p value <0.05).*

4.3.5.2 Splicing analysis of the ETAA1 exon 5 mini gene containing the long insert

A disadvantage of the *ETAA1* minigene containing the medium size genomic insert in lanes 3 and 4 of Figure 4-18, was it only contained the cryptic splice site, and not the 3′ splice site that is normally used when RBMX is present in cells. This means that the medium size minigene can't be used to examine competition between these two different 3′ splice sites. However, by carrying out similar experiments using the "long" minigene I was able to test choices between the normal splice site and the weak cryptic splice sites.

Co-transfection of *RBMXL2* only with the long minigene repressed the use of the weak splice site (Figure 4-22, lane 7). This is mediated by its RRM because co-transfection of the RBMXL2ΔRRM expression vector caused increased splicing inclusion of the cryptic exon compared to co-transfection of RBMX-GFP (in Figure 4-22 compare lanes 8 and 2).

A major complication of these minigene experiments was that co-transfection of either *RBMX* or *RBMXL2* caused increased skipping of the full exon compared to GFP control (in Figure 4-22 compare lanes 2, 7 and 1). This exon skipping effect was even stronger than the repression of the shorter cryptic exon. However, when RBMX and RBMXL2 were missing the RRM the inclusion of the full exon increased compared to full length RBMX and RBMXL2 (in Figure 4-22 compare lanes 2 and 3, lanes 7 and 8). This means the inclusion of the full exon relies on the RRM, the inclusion of the cryptic splice site does not relay on the RRM possibly on its binding with other RBPs.

Inclusion of the full exon (i.e. use of the normal splice site) was highest when Tra2β was co-transfected (Figure 4-22, lane 5). The inclusion of the full-length exon was reduced when RBMX was missing the C-terminus and when Tra2β was missing its RNP RNA binding domain and protein binding RS domain (Figure 4-22 lanes 4, 11 and 12). This means that the RRM of Tra2β is required for splicing inclusion of the full-length ETAA1 exon 5, and splicing repression of the full-length *ETAA1* exon 5 is dependent on the C-terminus of RBMX.
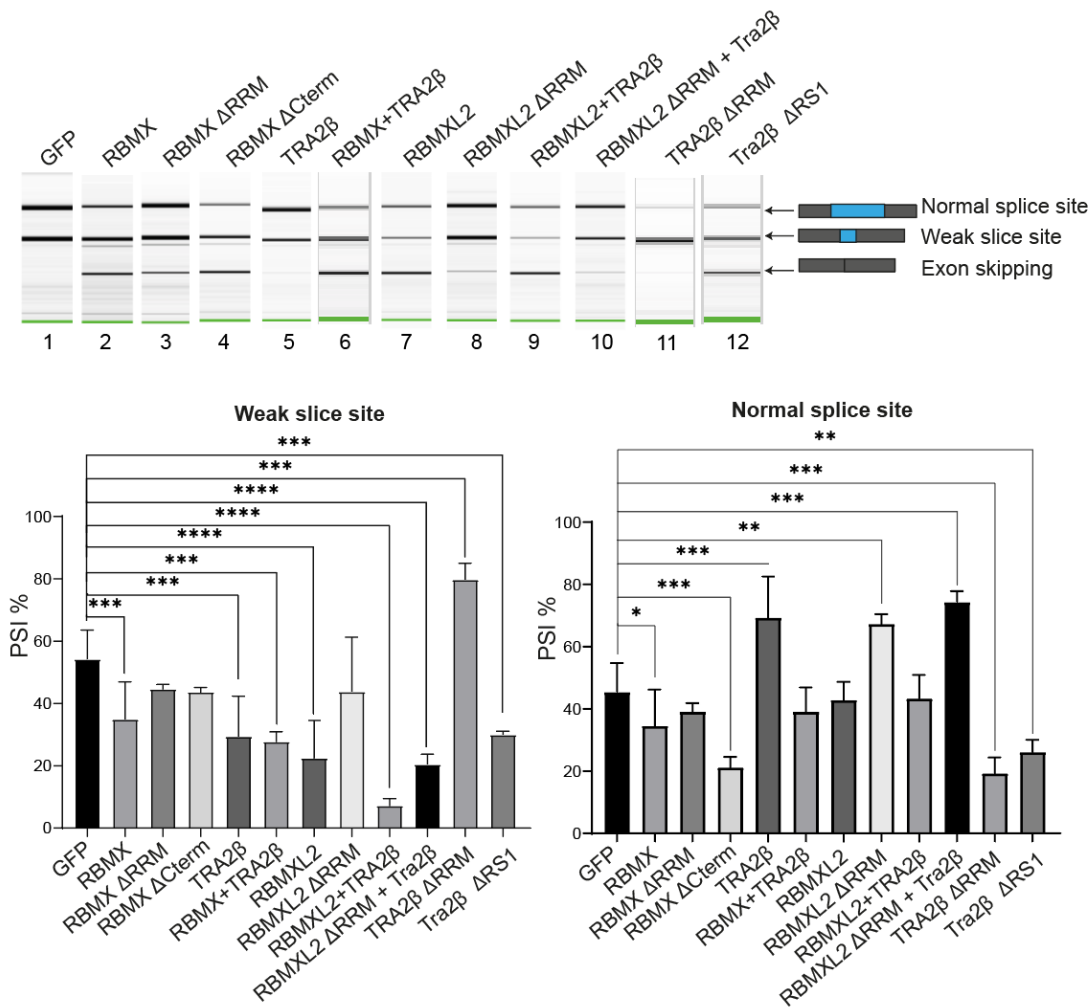
***Figure 4-22 RBMX and RBMXL2 encourage whole exon skipping of ETAA1 Exon
5.*** *Representative QIAxcel capillary gel electrophoretograms showing RT-PCR
validation of the "long-sized" mini gene in the conditions indicated using PXJF and
PXJB primers. The PSI was calculated from three or more biological replicates. Lines
in the bar plots represent mean ± SEM. P-values were produced by unpaired t-test, on
GraphPad Prism v 9.31 where*, **, ****, p value <0.05). (*, **, ***, ****, p value <0.05).*

4.3.5.3 Branch point mapping for the cryptic ETAA1 splice site

Overall these data are consistent with the experiments of Nasim et al that show that
RBMX and Tra2beta often have antagonistic effects on minigenes (Nasim et al., 2003).
One possibility how RBMX represses cryptic splice site selection could be via stearic
occlusion of key splice site sequences in the pre-mRNA, such as the branchpoint. Next,
I used the *ETAA1* long insert minigene to map the branch point that was used by the
cryptic splice event. The RNAseq showed that *ETAA1* exon 5 has two cryptic 3′ splice
sites one predominately used in MDA-MB-231 and two used in HEK293 cells.

Using the RNA from HEK293 cells transfected with the long *ETAA1* mini gene and either RBMX or RBMX ΔRRM, I was able to map the branch point sites used for the splicing of the cryptic exons identified in HEK293 and MD-MB-231 cells (Figure 4-23). I used the minigene transfected cells, as these would be expressing pre-mRNA transcribed from many plasmids per cell to help ensure that there were many lariat intermediates to be detected.

Experimental mapping of branch points takes advantage of the fact that RT polymerase can amplify across the 2′-5′ linkage present in lariat intermediates (Královičová et al., 2021). The primer design used to amplify across the lariat is shown in Figure 4-23 A. Two different sets of primers were used to ensure that the region containing the branch point was amplified. Figure 4-23 panel B shows the PCR amplifications using the two different primer sets. The PCR amplicons were cloned into pGEM-T-Easy vector (Promega) and sent for Sanger sequencing. I was able to map two different branch points within exon 5 357bp and 112bp from the splice sites used in HEK293 and MDA-MB-231 cells. Both branch point sites had RBMX iCLIP tags in close proximity (Figure 4-23, C). The presence of these branch points close to RBMX binding sites could mean that RBMX binding is involved in the repression of the splice sites in vivo.
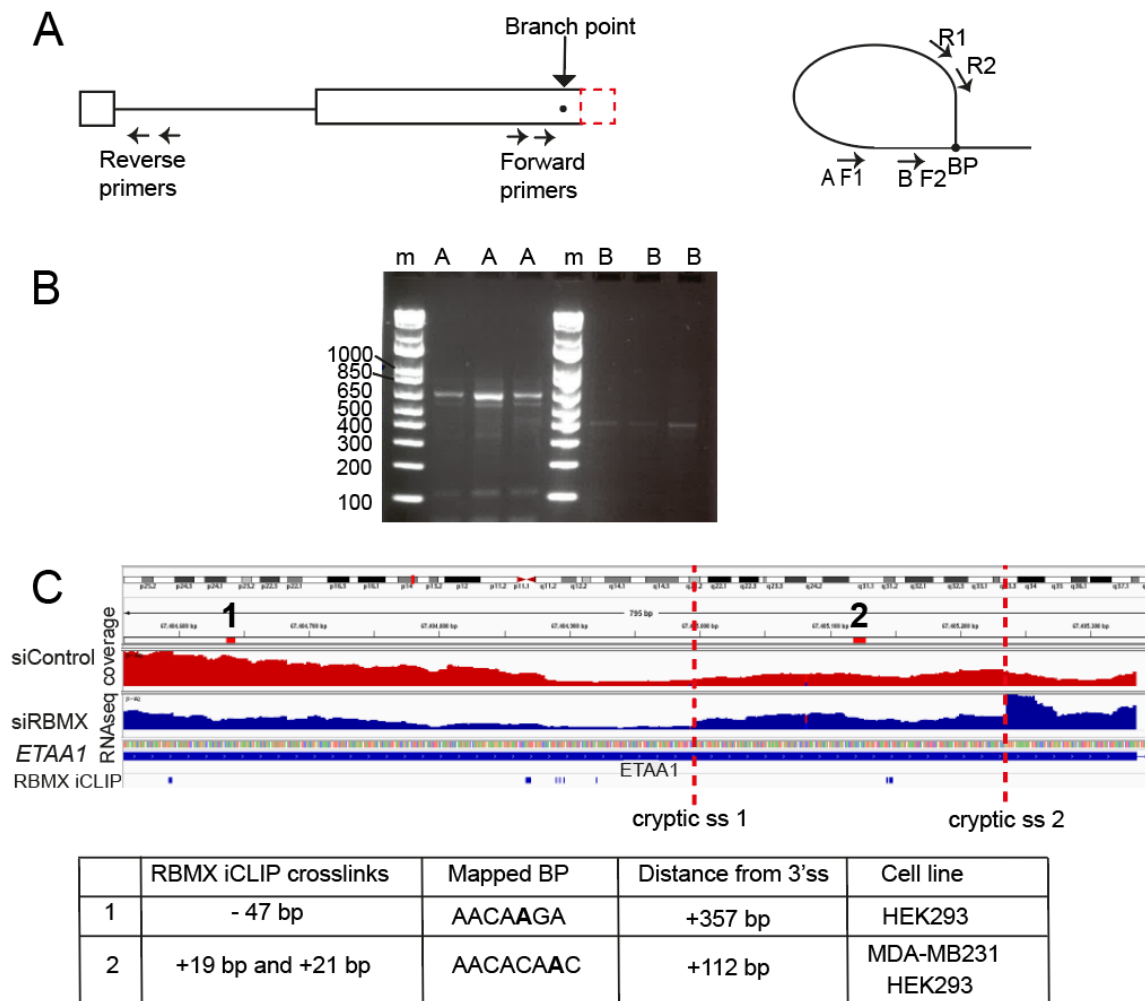
Figure 4-23 RBMX binds in close proximity to experimentally mapped branchpoints of ETAA1 exon 5 cryptic splice sites. *(A) Schematic of the RT-PCR approach to map the branch points across the lariat in ETAA1 exon 5 for the cryptic exon. (B) Agarose gel of PCR products amplified from the RNA lariat of ETAA1 exon 5. M, Invitrogen plus DNA marker; A, RNA from minigene transfected with full length RBMX lariat; B, RNA from minigene transfected with RBMX ΔRRM lariat. DNA marker sizes in base pairs are on the left. (C) IGV snapshot with mapped branch points and position of the splice sites in red dotted lines. Two different branch points were used by the upstream and downstream cryptic splice sites. The mapped branch point adenosines are indicated by bold letters. The distance of each branch point from the respective 3' splice sites and RBMX binding sites is indicated.*

If RBMX is repressing selection of the cryptic splice sites via binding to RNA, and blocking access to the spliceosome, I should be able to achieve a similar effect by blocking access using antisense oligonucleotides. To test this, I carried out some antisense oligonucleotide experiments. I designed an antisense oligo (ASO) to block the RBMX binding sites close to the 3′ splice site (relative position of ASO to RBMX

binding site and cryptic 3′ splice site is shown in Figure 4-24, A). I used an ASO in RBMX depleted cells to see if the ASO binding to the ETAA1 pre-mRNA could block access to the spliceosome in a similar way to RBMX. The ASO was transfected in MDA-MB-231 cells at the same time as RBMX siRNA for 72hours. RT-PCR showed decreased use of the cryptic splice site when RBMX was depleted in the presence of the ASO blocking the RBMX binding sites (Figure 4-24, B) although this was not significant. There was also an increase in the inclusion of the whole exon when the ASO was transfected compared to the control as shown on the QIAxcel gel (Figure 4-24, B).
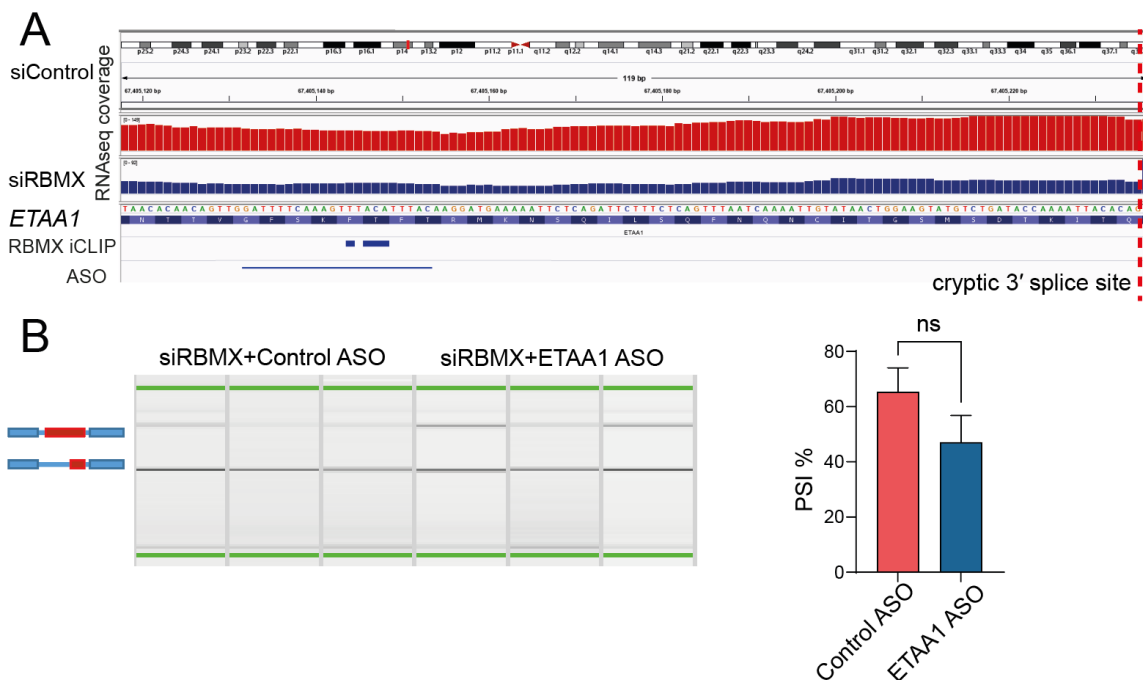
Figure 4-24 Use of an antisense oligo to test role of RBMX binding in RBMX depleted cells. *(A) IGV snapshot of the RNAseq reads from MDA-MB-231 cells. RNASeq coverage represents merged tracks from three (MDA-MB-231) biological replicates. Reads from the siRNA control cells are shown in red and the RBMX siRNA knockdown RNAseq reads are shown in blue. The ETAA1 Ensembl gene transcript isoforms indicate known isoforms of ETAA1. RBMX iCLIP tracks are shown below the gene. The position of the ASO in relation to the RBMX binding sites mapped by iCLIP are shown. (B) Representative QIAxcel capillary gel electrophoretograms showing RT-PCR validation of the ASO in MDA-MB-231 cells. The PSI was calculated from three biological replicates from negative control ASO (pink) and ETAA1 ASO (blue). Lines in the bar plots represent mean ± SEM. P-values were produced by unpaired t-test, on GraphPad Prism v 9.31 where ns= not significant.*

4.3.6 RBMX autoregulation

I also noticed that when RBMX was depleted in MDA-MB-231 cells *RBMX* was alternatively spliced. There was decreased skipping of the 3′ UTR in exon 9 of *RBMX*. This 3′ UTR region was dense in RBMX binding sites. When RBMX is depleted we notice decreased skipping of the 3′ UTR. A different UTR is preferred. This is related to the autoregulation of *RBMX*, RBMX is binding to its own 3′ UTR to regulate mRNA expression.
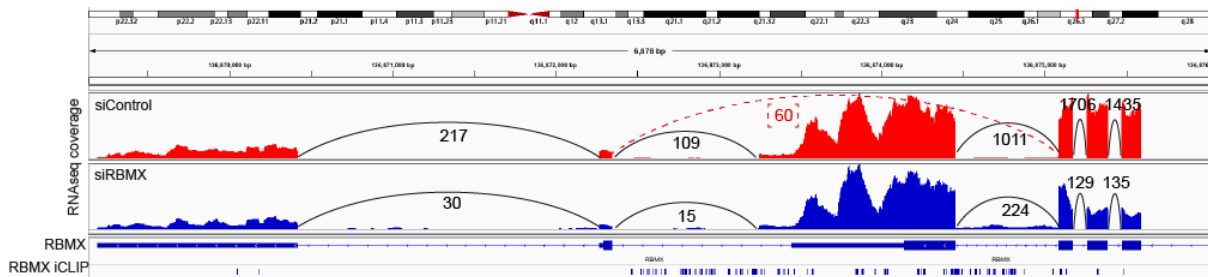
Figure 4-25 Sashimi plot of RBMX 3'UTR with RNAseq coverage of siRNA control in red and siRNA RBMX in blue. *A drastic drop in splice junction reads highlighted in a red dotted line can be seen when RBMX is depleted.*

4.4 Discussion

In this chapter, the iCLIP data was correlated with RNA-seq analysis after RBMX depletion to work out how it regulates RNA processing. RBMX has been shown to repress cryptic splice sites within genes important for DNA damage. Many of these genes have ultra-long exons (over 1000bp long). Longer exons are more likely to have cryptic splice sites and splicing repressors such as RBMX prevent the use of these cryptic splice sites. I was able to identify RBMX binding sites in RBMX regulated exons using iCLIP data from chapter 3 together with the RNA-seq analysis. I also carried out mini gene experiments to investigate the use the splicing of *ETAA1* in HEK293 cells.

4.4.1 RBMX represses the use of splice sites in exons of genes important for cell cycle, DNA replication and chromosome biology.

I detected transcript changes in 611 differentially expressed exons when RBMX was depleted in MDA-MB-231 cells. In chapter 3, I was able to show that RBMX binding sites were enriched in genes important for cell cycle, DNA replication and chromosome biology. The GSEA using the RNA-seq data described in this chapter further shows that when RBMX is absent there is an upregulation of cell cycle, DNA damage and replication associated gene expression pathways. In addition, Gene Ontology enrichment analysis has shown that RBMX regulates the RNA processing of differentially expressed exons within genes involved in chromosome biology and cell cycle processes.

My data presented in this Chapter is consistent with what is already known about RBMX biology. A previous study reported that RBMX has a role in the activation of ATR in concert with TopBP1 to repair damaged DNA (Zheng et al., 2020). In this chapter, we have shown that the *ETAA1* gene (which encodes a replication fork protein) has increased use of a cryptic 3′ splice site in 3 different cell lines. Experiments done in this chapter showed that alternative splicing of *ETAA1* was altered when RBMX and Tra2β/α were depleted from cells. In the absence of RBMX there is also reduced full length ETAA1 protein in MDA-MB-231 cells. This would therefore disrupt the ability of ETAA1 binding to RPA and its role in the DNA repair pathways, since ETAA1 is important for binding of RPA during DNA repair at stalled replication forks (Lee et al., 2016). Previous studies have shown that RBMXL2 (a paralog of RBMX) to be a

regulator of the splicing of *Brca2* in mouse testes. BRCA2 is involved in the homologous recombination pathway of DNA damage repair. A study done by Adamson et al. demonstrated that RBMX is a positive regulator of homologous recombination by promoting expression of BRCA2 (Adamson et al., 2012). The RNA-seq data in this chapter validates this finding. In this chapter, we have shown that RBMX is responsible for producing full length mRNA transcripts of *BRCA2*. The RNA-seq data demonstrates that when RBMX is absent there is increased use of an alternative Poly A site in exon 11, and this prematurely terminates the transcription of *BRCA2*. There is a high density of RBMX iCLIP tags at this alternative termination site, suggesting a model where RBMX is binding directly to prevent the use of this termination site. Other genes with RNA processing events regulated by RBMX also have direct RBMX binding sites. These include *REV3L*, *KNL1*, and *SGO2*, each of which are key DNA replication and chromosome biology genes.

My data indicates that RBMX represses 82% of RNA processing events that it regulates. Many of these repressed splice sites were cryptic RNA processing events. These data suggest that RBMX is involved in maintaining the fidelity of full mRNA transcripts in key DNA replication and chromosome biology genes. The RNA processing events regulated by RBMX include alternative splice sites and poly A sites. This is consistent with the current literature where RBMX has been identified as a splicing regulator in large scale proteomic analysis of interchromatin granule clusters (IGCs) and spliceosome associated factors (Mintz et al., 1999; Rappsilber et al., 2002). Co immunoprecipitation experiments have shown that RBMX is associated with splicing activator TRA2β in HeLa cells (Venables et al., 2000). RBMX and Tra2β protein interactions have been previously shown to promote the inclusion of SMN2 exon 7 (Hofmann & Wirth, 2002). For these reasons I over expressed Tra2β, RBMX and RBMXL2 on HEK293 cells transfected with ETAA1 exon 5 mini gene to observe their effect on alternative splicing.

4.4.2 Analysis of RBMX function using minigenes

Since RBMX binds to other proteins an important question is if RBMX binds directly to RNA to facilitate alternative splicing or does it rely on protein interactions. *ETAA1* exon 5 has a high density of RBMX crosslinks as demonstrated by the iCLIP experiments.

Mini gene experiments carried out using the minigenes with long and medium inserts showed that alternative splicing of *ETAA1* was altered by RBMX, RBMXL2 and Tra2β over expression in HEK293s. The overexpression of Tra2β reduced exon skipping, and this effect was dramatically reduced when Tra2β was missing its RS1 and not its RNP domains, therefore Tra2β may need to bind other proteins to encourage exon inclusion using both the weak and normal splice sites. RBMX decreased the use of the weak splice site as expected from the RNAseq data, yet it also increased exon skipping of the whole exon. The effect of RBMX on the *ETAA1* minigenes was not as strong as the RNAseq data after RBMX depletion. Still deletion of the C-terminus of RBMX reduced inclusion of the full exon. The C-terminus of RBMX had a RBD which is different from the N-terminus RRM (Kanhoush et al., 2010a). Deletion of the RBMX RRM did not have a big difference. Deletion of the RBMXL2 RRM increased use of the weak splice site and reduced exon skipping compared to full length RBMXL2.

Since we did not see a big change between RBMX transfection and RBMX missing the RRM, we speculate that RBMX control of splicing may be affected by chromatin state. The binding of other proteins on chromatin may work together with RBMX to regulate the splicing of ETAA1 exon 5. In the lab Saad Aldalaqan has, overexpressed RBMXL2 and RBMY in RBMX depleted HEK293 cells. Interestingly both RBMXL2 and RBMY were seen to compensate for the loss of RBMX in these cells (Aldalaqan and Siachisumo, data not shown).

4.4.3 Identification of Branch points for cryptic splice sites in ETAA1 exon 5

Since we were able to identify iCLIP binding sites close to the 3′ splice site of ETAA1 we wanted to identify the branch site and also see if RBMX direct binding to the endogenous ETAA1 affected splicing. Computational and experimental methods have been used to predict branch point locations within introns with great accuracy (Taggart et al., 2017; Leman et al., 2020)(Briese et al., 2019). Using RT-PCR branch point mapping experiments we were able to identify a branch point 112bp from the splice site of the cryptic exon highly used in MDA-MB-231 cells compared to HEK293 cells. There are RBMX iCLIP crosslink sites +19bp and +21bp upstream of the branch point, and therefore RBMX direct binding to ETAA1 mRNA could affect the use of the cryptic splice site. We used an ASO to block the RBMX binding sites close to that branch

point. Blocking the RBMX binding sites led to increased inclusion of the full-length ETAA1 exon 5 and reduced inclusion of the cryptic exon in RBMX depleted cells, although this latter experiment was not statistically significant. Moving forward, mutating the branch point sequence in RBMX depleted cells could be used to test the importance of the mapped branch points in cryptic exon selection.

4.4.4   RBMX autoregulation

When RBMX was depleted there was reduced exon skipping of the 3′UTR in exon 9 favouring expression of a shorter more stable isoform of *RBMX* mRNA. This could be a regulatory mechanism to compensate for the loss of RBMX protein. Since multiple RBMX binding sites were present in the alternatively splice 3′ UTR, RBMX could be involved in directly regulating itself through an auto regulatory pathway. When there is a high amount of RBMX protein in cells an alternative 3′ UTR is skipped expressing a less stable version of *RBMX*.

4.4.5   Model of RBMX function in repressing cryptic RNA processing events

In our RNA-seq data, we shown that RBMX is involved in suppressing the use of cryptic splice sites within exons and in some cases introns. The iCLIP data identified a high number of RBMX crosslinks on mRNA suggesting that RBMX binding might be essential for exon regulation. An important question to address is how RBMX stops cryptic splice sites from being included in the full-length transcript. RBMX could be doing this is one of two ways shown in Figure 4-26. (1) RBMX may sterically block splice sites and as a consequence binding of another RBP (such as Tra2β) that may be acting as activators. In addition RBMX binding may interact with other RNA binding proteins involved in pre-mRNA processing. RBMX has been shown to interact with HNRNPK, A1, H1, L, M, PA2B1, R, C, PTBP1 and TRA2B RNA binding proteins. Known interacting proteins can be identified by co-immunoprecipitation assays. Previously unknown interactions can be discovered using pulldown assays together with mass spectrometry analysis. (2) RBMX may directly binds to splice site RNA and acts as a repressor inhibiting the use of RNA processing sites.

A

RBMX acts similarly to RBMXL2 and represses the use of cryptic splice sites
1. RBMX inhibits splicing activator Tra2β
2. RBMX sterically blocks access to cryptic RNA processing sites

B

Absence of RBMX leads to the use of cryptic RNA processing sites
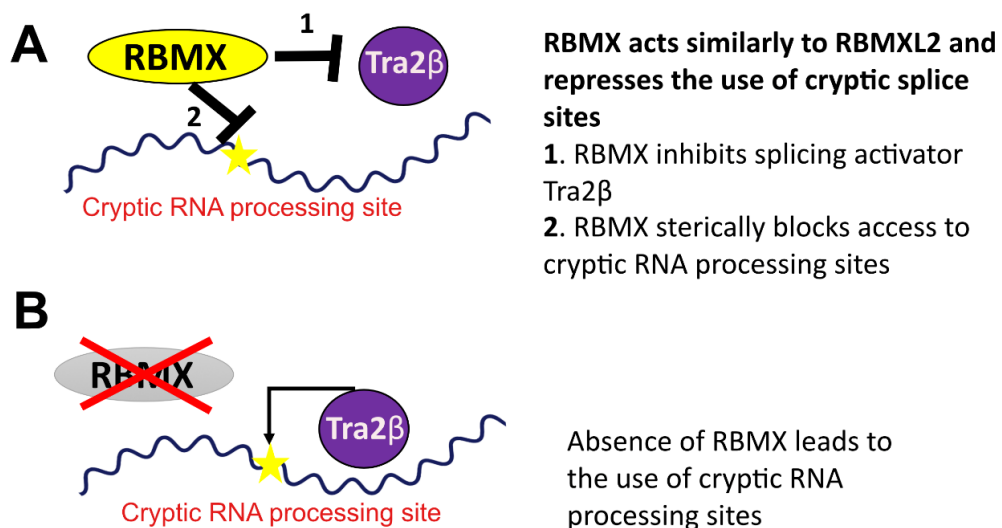
Figure 4-26 *Model for RBMX function RBMX is important for the use of proper splice isoform. A) Inhibition of RNA processing events by RBMX B) The absence of RBMX will result in an increase of use of cryptic splice sites, similar to RBMX retrogene RBMXL2* (Ehrmann et al., 2019)

4.5   Chapter summary

In this chapter I have tested the hypothesis that RBMX is responsible for the RNA processing of exons in genes important for genome stability. RBMX mainly has a repressive function in preventing use of cryptic mRNA processing sites, thus promoting correct DNA damage response and genome maintenance. RBMX regulated exons were frequently identified in genes important for genome stability. GSEA and Gene Ontology analysis also supported by the Gene Ontology done in chapter 3 on RBMX bound exons.

Analysis was carried out to investigate the effect of RBMX over expression, RBMX knockdown and Tra2-α/β knockdown on the splicing of endogenous *ETAA1*. There was a slight but significant decrease in the inclusion of the small exon when RBMX was overexpressed. There was a large significant increase in the inclusion of the cryptic exon of ETAA1 when RBMX and Tra2-α/β were depleted compared to the negative controls. This together with the minigene data supports ETAA1 exon 5 being a jointly regulated by RBMX and Tra2β.

The data described in this Chapter suggested that depletion of RBMX would interfere with genome stability. I describe experiments to test this in the next Chapter.

Chapter 5. Testing biological effect of RBMX knockdown on genome integrity

5.1 Introduction

In chapters 3 and 4 I have shown that RBMX binds and regulates exons within genes important for DNA damage repair and chromosome regulation. This is supported by current literature that has shown that loss of RBMX is important for maintaining sister chromatid cohesion. It has also been shown that RBMX is localised to DNA breaks. These functions do not relay on the ability of RBMX to bind RNA. Eukaryotes have evolved many tightly regulated processes in response to DNA damage. During every cell cycle, accumulation of DNA breaks halts replication forks. ETAA1 (a target of RBMX identified by RNAseq in chapter 4) is functionally important for the recruitment of RPA and binds to ssDNA at stalled replication forks.

DNA damage repair mechanisms sense and signal genotoxic events to either repair DNA or cause cell death if the DNA cannot be repaired. The DNA damage response (DDR) is mediated by protein kinases. These include ATM (ataxia telangiectasia-mutated) and ATR (ATM and Rad3-related) (Cimprich & Cortez, 2008). ATR and ATM have overlapping targets critical for DSB repair. These include kinases such as Chk1 and Chk2. The ATR pathway is activated once RPA is recruited to ssDNA at stalled replication forks or intermediates at sites of double strand breaks (Ball et al., 2005). RPA is made up of 3 protein subunits RPA14, RPA32 and RPA70. RPA is essential for binding nearly all ssDNA at DNA processing events. RPA is exchanged for Rad51 to sites of ssDNA. Rad51 recruitment is a highly regulated process predominantly mediated by BRCA2 (Liu et al., 2010). The replication fork protein ETAA1 is recruited to sites of DNA damage via its RPA binding domain (Figure 5-1) (Lee et al., 2016).

Activation of ATM leads to the phosphorylation of a number of crucial proteins at sites of DSB this includes histone H2AX at serine 139 (Burma et al., 2001) in its phosphorylated state H2AX (phosphorylated H2AX is referred to as γH2AX) recruits additional DDR proteins to the site of DSB (Figure 5-1).
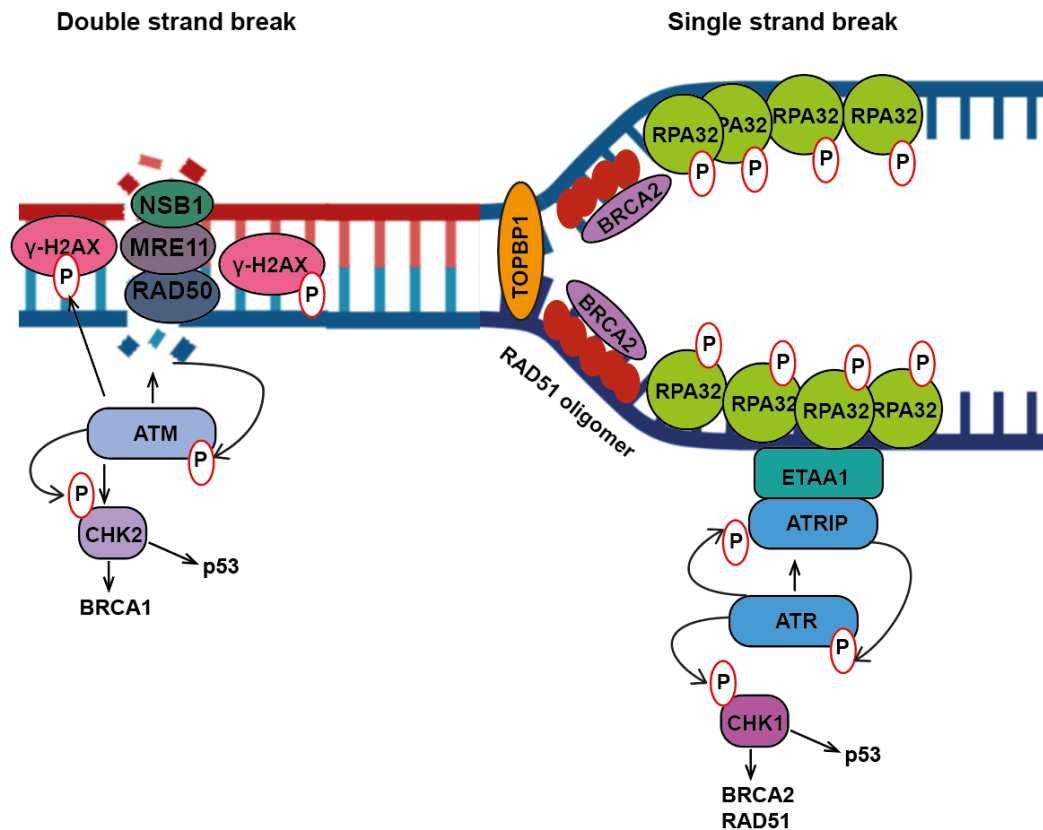
Figure 5-1 DNA damage repair signalling pathways. *The ATM DNA damage repair pathway is activated in response to dsDNA breaks and ATR DNA damage repair pathway is activated in response to ssDNA at stalled replication forks. γH2AX accumulates at sites of dsDNA breaks and RPA32 accumulates at sites of ssDNA breaks, BRCA2 mediates RAD51 recruitment to ssDNA breaks replacing RPA32. Activated ATM and ATR phosphorylate CHK2 and CHK1 effectors respectively, both lead to activation of additional downstream factors such as p53 to mediate DNA repair or eliminate the damaged cell by apoptosis or senescence.*

Studies have shown that inactivating checkpoints that prevent cell cycle progression in the presence of unrepaired DNA damage can increase p53-independent death. In one instance, p53-deficient cancer cells needed ATM, ATR, Chk1, and p38MAPK/MK2 signalling for cell cycle arrest. When this response was lost, DNA damage resulted in caspase-3 activation and mitotic catastrophe (Reinhardt et al., 2007). The p53 protein is mutated in several cancers and functions as a transcription factor. When cells are stressed (DNA damage) p53 is stabilised and facilitates the transcription of multiple genes involved in growth arrest, DNA repair and apoptosis if the DNA damage cannot be repaired in the cell depending on the cellular context (Pucci & Giordano, 1999), (Ozaki & Nakagawara, 2011).

DNA damage effects upon RBMX depletion have been shown in the literature in HeLa and U2OS cells. When RBMX was depleted high levels of phosphorylated CHK1 (CHK1 pS345) and phosphorylated RPA2 were seen (Zheng et al., 2020). An increase of γH2AX and an increase in chromosomal defects were also seen when RBMX was depleted (Adamson et al., 2012), (Matsunaga et al., 2012).

It has been suggested that RBMX may play a role as a tumour suppressor (Shin et al., 2008). However, its mechanism of action is still unknown. RBMX plays a role in apoptosis of breast cancer cells where its expression has a positive correlation with pro-apoptotic Bax gene (Martínez-Arribas et al., 2006). RBMX has also been shown to reduce tumorigenicity of human oral small cell carcinomas in vivo. An increase in RBMX expression led to an upregulation of Txnip, which is a tumour suppressor gene involved in maintaining genome stability (Shin et al., 2006). RBMX localises to sites of DNA damage and is necessary for resistance to DNA damage. In Chapters 3 and 4 of this thesis I identified that RBMX bound to and regulated genes important for DNA damage and repair. RBMX may be regulating DNA damage by enabling RNA processing of full-length protein isoforms that maintain genome stability at replication fork Figure 5-2.
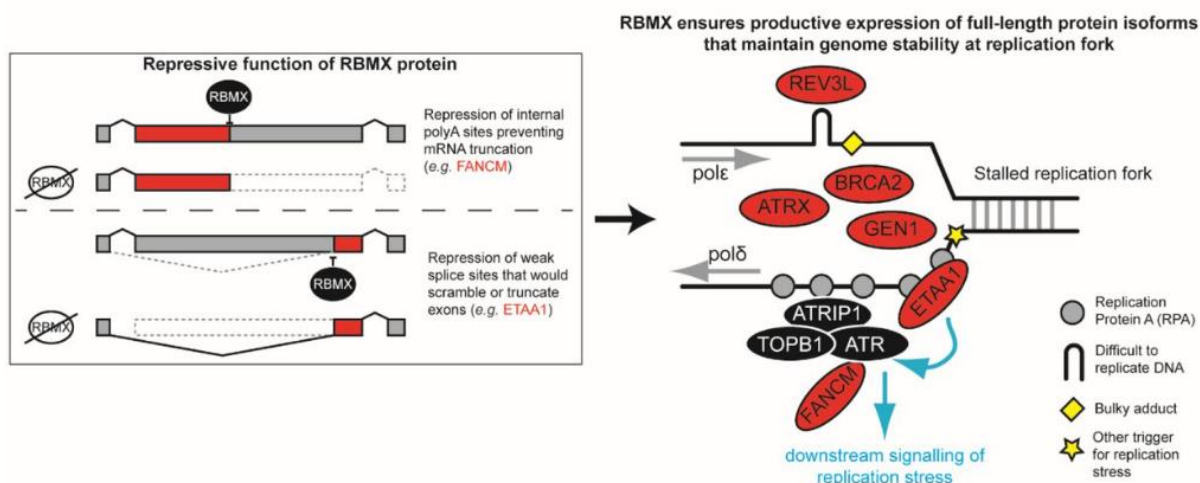
***Figure 5-2 RNA processing by RBMX controls key genes for replication fork stabili**ty. Schematic representation of the repressive function of RBMX in preventing the use of cryptic mRNA processing sites therefore promoting correct replication stress response and genome maintenance.*

In this chapter we examine the effects of RBMX depletion on the cell cycle and DNA damage response in MDA-MB-231 cells. We hypothesised that RBMX depletion

sensitises cells to DNA damage. To investigate this, we tested p53, phospho RPA32 (S4/S8), γH2AX (S139), phospho Chk1 (S345) and RAD51 protein expression in RBMX knockdown cells compared to controls. Genotoxic drugs exacerbate replication stress. We wanted to find out if this replication stress could be further increased by RBMX depletion. To investigate the genotoxic effect on RBMX depleted cells we tested protein expression of DNA damage markers after treatment with Camptothecin (CPT) and hydroxyurea (HU).

5.2 Aims

The aims of this chapter were to:

1. Monitor the cell cycle effects after RBMX depletion in MDA-MB-231 cells
2. Investigate the DNA damage and replication defects after RBMX depletion.
3. Investigate if RBMX depletion sensitises cells to treatment of genotoxic drugs.

5.3 Results

5.3.1 *Depletion of RBMX is associated with an increase in numbers of cells in mitosis*

In the literature RBMX has been shown to be involved in DNA replication and cell cycle. The results in our own DESeq analysis show that when RBMX is depleted in breast cancer MDA-MB-231 cells there is low expression of Cyclin B2 and an enrichment of cell cycle and DNA replication pathways. Taken together this tells us that there may be a defect in cell cycle regulation in MDA-MB-231 cells when RBMX is depleted. In our GO and GSEA analysis of RBMX bound targets and RBMX regulated exons, I have showed that terms relating to DNA damage, replication and cell cycle were enriched upon RBMX depletion. In our GSEA data, P53 independent G1 S DNA Damage checkpoint was significantly upregulated. This pathway is ascribed to cells arresting in G1 phase of the cell cycle. So we carried out flow cytometry analysis of cell cycle proportions in RBMX depleted cells versus the control cells. I found that there was no significant difference in the number of cells in S, G1, G2 and G0 phases of the cell cycle between RBMX siRNA depleted and siRNA control cells (Figure 5-3). A significant increase was seen in the number of cells in mitosis in the RBMX siRNA depleted cells compared to siRNA control cells (Figure 5-3).
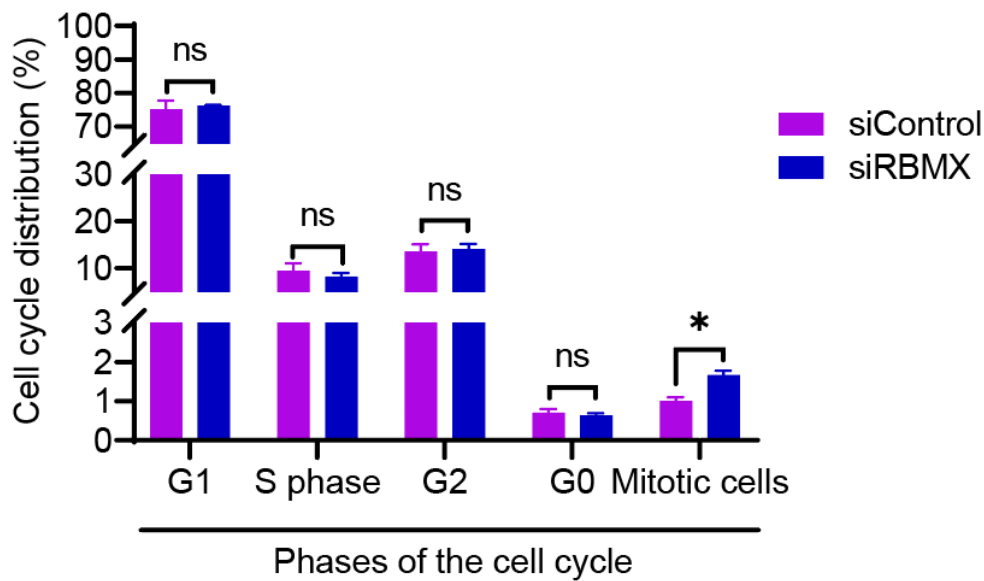
Figure 5-3 Bar chart showing the percentage of cells in the various stages of the cell cycle. *There is a significant increase in the number of mitotic cells after RBMX depletion in MDA-MB-231 cells. Mean values of the replicates were plotted with error bars indicating standard deviation. Statistical significance was calculated using GraphPad Prism v 9.31 with unpaired t test (* pvalue< 0.05, ns=not significant).*

5.3.2 Depletion of RBMX caused disruption of the p53 pathway and accumulation of γH2AX DNA damage marker

RBMX has been found to be mutated in Shashi syndrome (Shashi et al., 2015). It has been found that a deletion mutation in RBMX low complexity RRG/RG region is the cause of Shashi syndrome (Cai et al., 2021). This deletion caused aberrant activation of the p53 pathway and defects in alternative splicing, an increase in p53 protein was observed but there was a decrease in *TP53* mRNA transcripts. P53 is a transcription factor and tumour suppressor protein. When RBMX was depleted in MDA-MB-231 cells I observed a decrease of *TP53* at the mRNA by 2.2 fold this is the same observation in the Cai., et al study. However at the protein level I surprisingly observed a decrease in p53 protein (Figure 5-4, A). Interestingly in MDA-MB-231 cells I found that loss of RBMX lead to an increase in p53 target *CDKN1A* at the mRNA level. I observed a 2.8 fold increase in *CDKN1A* in RBMX depleted MDA-MB-231. This was also observed in RBMX depleted U2OS cells (Cai et al., 2021).

It has been shown that loss of RBMX leads to genome instability and persistent activation of the DNA damage response. Increased levels of PCNA (a marker of DNA

replication stress) was seen to be increased in U2OS and HeLa cells (Zheng et al., 2020). In our study loss of RBMX did not lead to an increase of PCNA compared to control cells by western blot in MDA-MB-231 cells (Figure 5-4, A).

To detect DNA damage after RBMX depletion in MDA-MB-231 cells I also performed a γH2AX (S139) western to detect DNA damage, and a phospho RPA32 (S4/8) western to detect replication stress. γH2AX plays an important role in activating the DNA damage response (Podhorecka et al., 2010). When RBMX was depleted an increase in DNA damage was detected by γH2AX using western blot (Figure 5-4, B) this result was qualitative as it was not quantified with a total H2AX detection. An increase in γH2AX foci was also observed by fluorescent staining in nuclei of siRNA RBMX depleted cells compared to siRNA controls (Figure 5-5) this result was qualitative as the number of foci were not quantified. RBMX depleted cells had a higher signal of γH2AX compared to a weak signal in control cells, suggesting that in there may increased DNA damage in the absence of RBMX. To verify this effect more quantifiable experiments would need to be carried out.

RPA32 plays an important role in cellular response to DNA damage. RPA32 forms a complex with other proteins involved in DNA metabolism. When RBMX was depleted I did not observe any difference in phospho RPA32 (Figure 5-4, B) suggesting that the absence of RBMX does not prevent DNA break repair from proceeding properly.
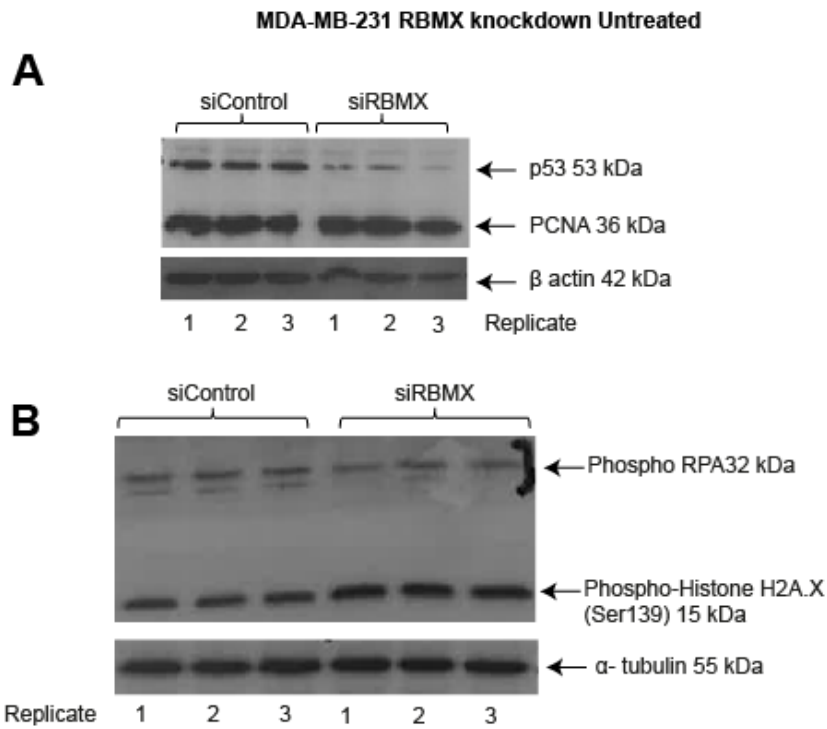
Figure 5-4 DNA damage and repair effects. *Western blot carried out in siRNA control and siRNA RBMX MDA-MB-231 showing (A) P53, PCNA and a β-actin loading control protein expression. (B) Phospho RPA32 and γH2AX and α-tubulin loading control protein expression.*
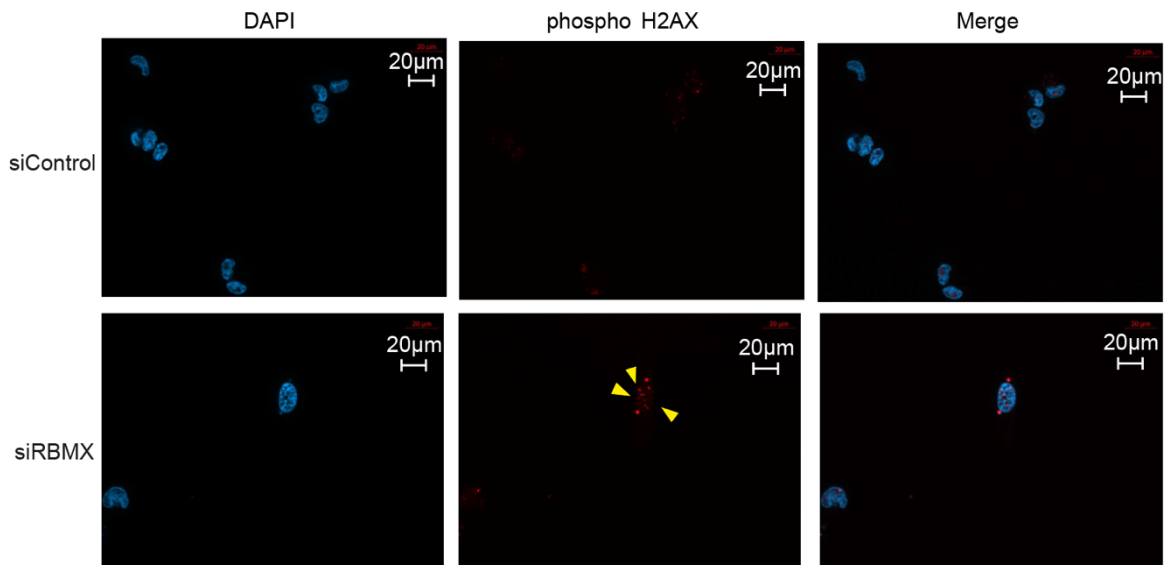
Figure 5-5 RBMX depletion is associated with accumulation of γH2AX foci within nuclei in MDA-MB-231 cells. *siRNA RBMX knockdown and siRNA control cells stained for γH2AX. An increase in γH2AX foci in red is seen in RBMX depleted cells. 40 x magnification, 20µm scale cars shown. Images analysed using ZEN 3.0 (blue edition).*

In work carried out by others (Zheng et al., 2020) it has been shown that when RBMX was depleted in U2OS cells more DNA fragments were detected by comet assay. To further investigate the effect of RBMX depletion in MDA-MB-231 cells I started to optimise a comet assay. Preliminary results from the comet assay are shown in figure Figure 5-6. I observed bigger comets in the RBMX depleted MDA-MB-231 cells. These results together with the γH2AX western and staining support the idea that RBMX may play an essential role in the cellular response to DNA damage in MDA-MB-231 cells. However to draw conclusions from the comet assay further optimisation of the lysis step needs to be carried out to obtain sufficient data that can be quantified.
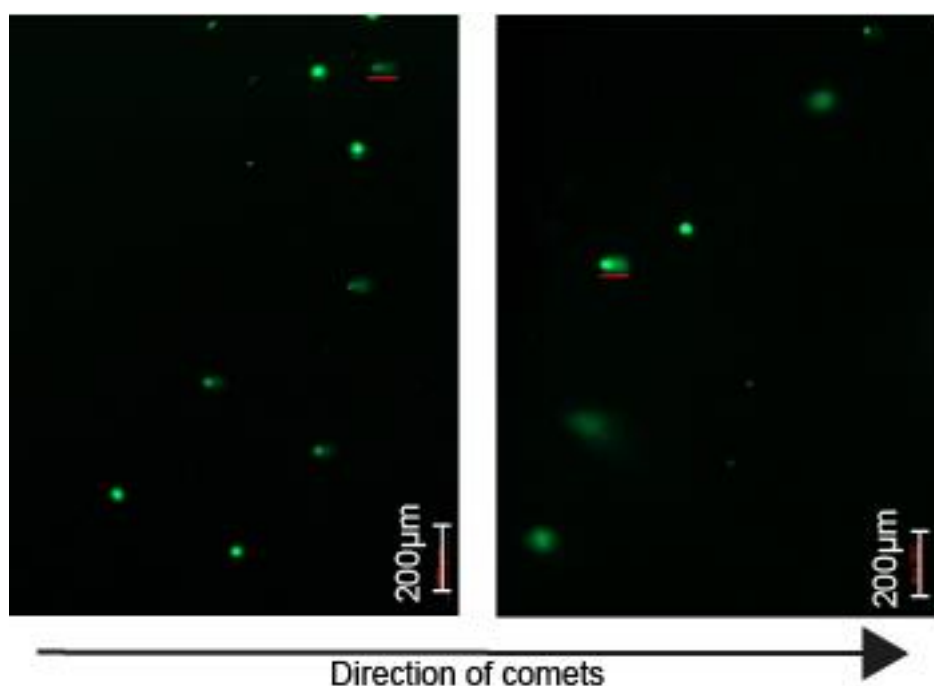
Figure 5-6 Comet assay detection of DNA fragments in control and RBMX depleted MDA-MB-231 cells. *Scale bars, 200μm.*

5.3.3 Investigating the effect of genotoxic drugs on RBMX depletion

In previous Chapters I have shown that RBMX is important for the productive splicing of genes important for DNA damage and replication. I hypothesised that RBMX depletion may sensitise cells to the effects of genotoxic drugs. Previous work has shown that depletion of RBMX in U2OS osteosarcoma cells sensitises them to DNA damage caused by ionising radiation and genotoxic drugs including Camptothecin (Adamson et al., 2012). To this end I performed preliminary experiments to treat RBMX depleted MDA-MB-231 cells with hydroxyurea (HU) and Camptothecin (CPT).

Treatment of cells with Camptothecin inhibits the activity of Topoisomerase I causing replication fork collapse (Berniak et al., 2013) and hydroxyurea treatment causes replication fork arrest (Singh & Xu, 2016).

5.3.3.1 *Investigating the effect of joint hydroxyurea and RBMX depletion on DNA damage*

The preliminary DNA damage experiments include western blots of γH2AX staining, phospho RPA32 and RAD51 on RBMX depleted cells treated with either 4mM hydroxyurea or 1 µM Camptothecin. We wanted to test if DNA damage and replication stress would be increased after treating RBMX depleted cells with genotoxic drugs. To test this I treated RBMX depleted and control cells with 4mM of hydroxyurea for 0, 5.5 and 10 hours.

Previously I have been able to show an increase in γH2AX in RBMX depleted cells by western blot (Figure 5-4) and immunofluorescence staining (Figure 5-5). Using this assay, hydroxyurea treatment did not exacerbate the effect of RBMX depletion. There was added no hydroxyurea dependent increase in DNA damage seen between 0 hours and 5.5 hours of hydroxyurea treatment by western blot (Figure 5-7). After 10 hours of hydroxyurea treatment both RBMX depleted and control cells had very high levels of DNA damage indicated by high levels of γH2AX (Figure 5-7).

It has been shown that RPA32 phosphorylation does stimulate the repair of chromosomal DNA damage(Anantha et al., 2007). There was no difference seen in the amount of phospho RPA32 between RBMX depleted and control cells by western blot (Figure 5-7). However, an increase in RPA32 was seen by western blot after treatment with hydroxyurea in both conditions (Figure 5-7). This substantiates the presence of replication stress in these hydroxyurea treated cells. RPA32 is required for the recruitment of RAD51 a double strand break repair factor (Shi et al., 2010). I also tested for RAD51 after RBMX depletion and treatment with hydroxyurea and saw no changes in total RAD51.

By immunofluorescence staining after 5.5 hours of treatment with hydroxyurea there is a slight increase in RPA32 staining in RBMX depleted cells compared to 0 hours and 10 hours of hydroxyurea treatment.
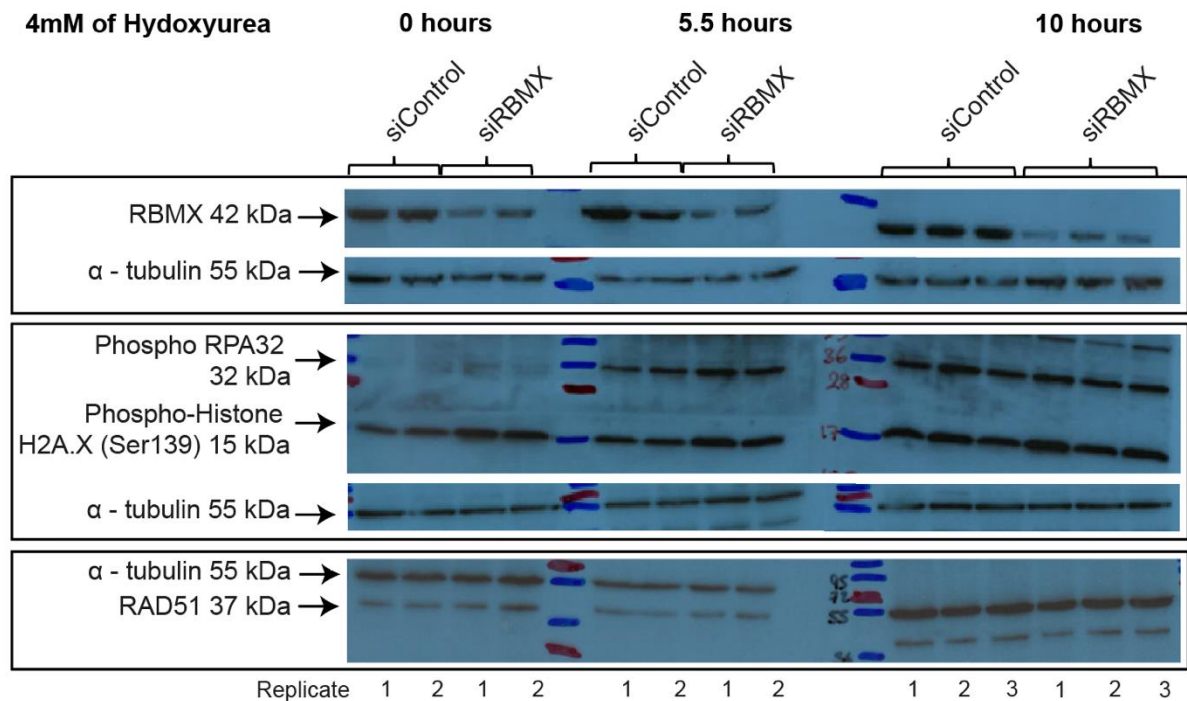
Figure 5-7 Hydroxyurea does not enhance the DNA damage effect of RBMX depleted cells. *MDA-MB—231 cells treated with 4mM of hydroxyurea after 72 hours of siRNA RBMX or siRNA control. Protein expression of RBMX, phospho RPA32, γH2AX, RAD51 and β-tubulin loading control. NB: RBMX in the top right panel appears at a different size due to the placement of the nitrocellulose membrane in the cassette.*
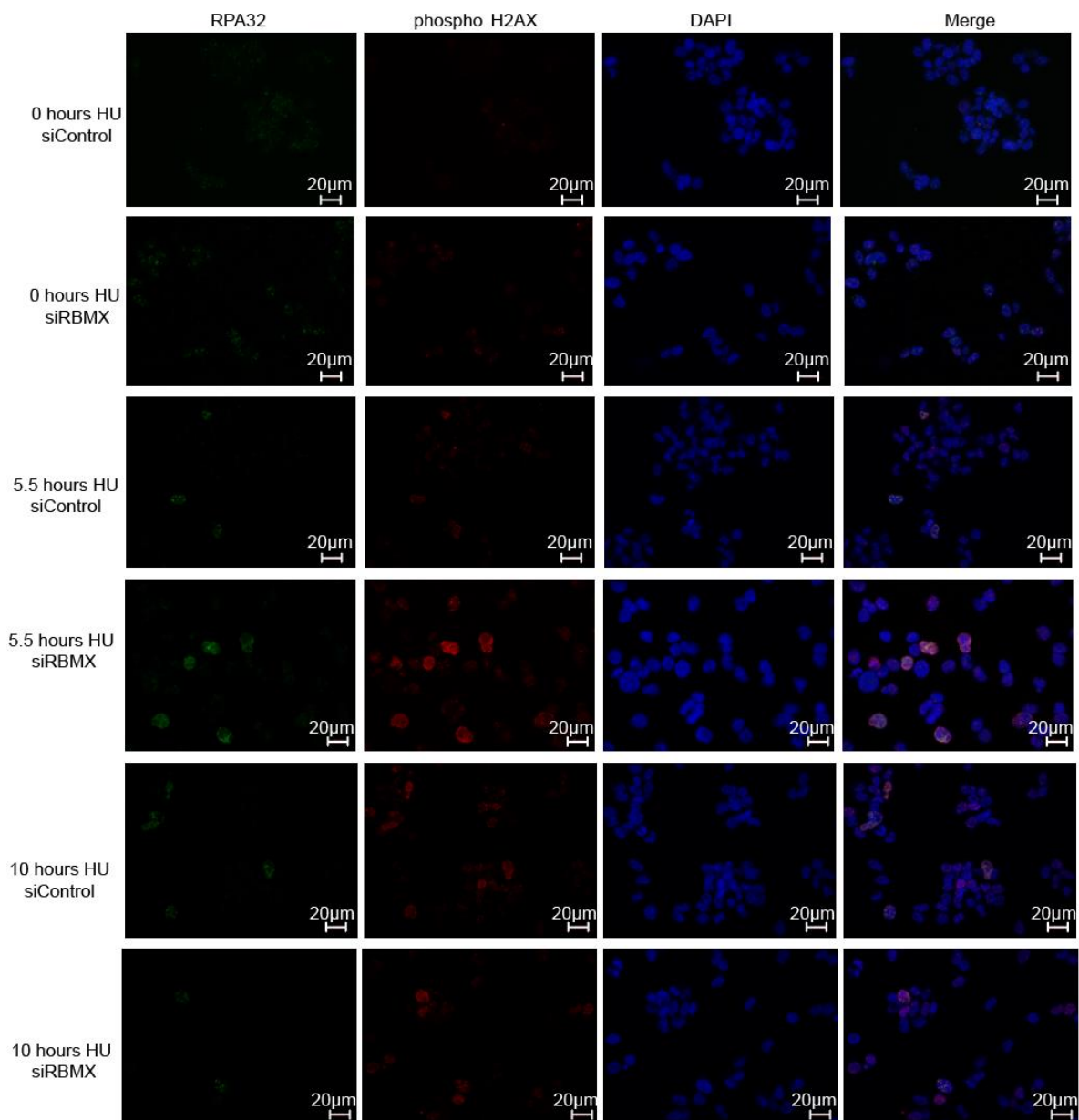
Figure 5-8 Immunofluorescence staining of DNA damage markers phospho RPA32, γH2AX and DAPI in MDA-MB231. MDA-MB-231 cells depleted of RBMX and treated with 4mM hydroxyurea for 0hrs, 5.5 hours and 10 hours. Increased phospho RPA32 and γH2AX staining is seen after 5.5 hours of treatment when RBMX was depleted. 40 x magnification, 20µm scale cars shown. Images analysed using ZEN 3.0 (blue edition).

Replication stress activates the ATR/ATM DNA repair pathways. Chk1 and RPA32 are downstream targets of the activated ATR pathway, and are phosphorylated at sites of replication stress (Shiotani et al., 2013). Western blots were carried out using Licor secondary antibodies to enable quantification of phospho and total proteins. To help ascertain that the increase in phosphorylated proteins was due to activation and not an increase in gene expression I used antibodies to detect total protein and the

phosphorylated protein. Depletion of RBMX resulted in no significant changes in phosphorylated Chk1 (Chk1 pS345) (Figure 5-9, A) and RPA32 (Figure 5-9, B) after hydroxyurea treatment. RPA32 western was repeated to normalise the phospho RPA32 over the total RPA32.
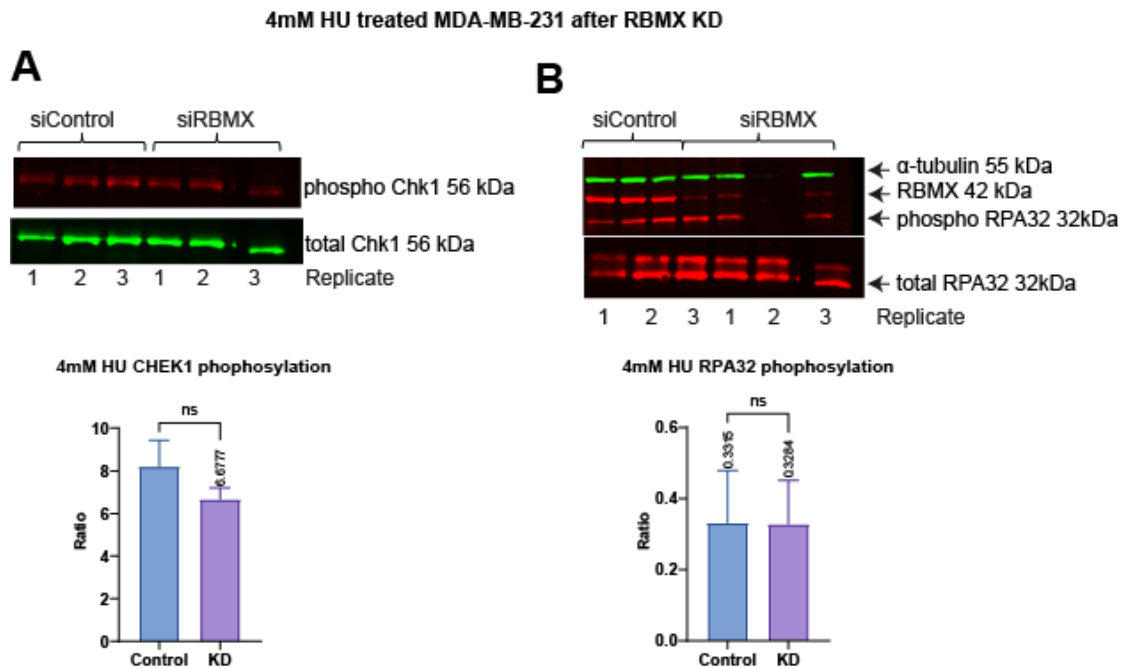
Figure 5-9 Treatment with hydroxyurea does not have an effect on Chk1 and RPA32 in RBMX depleted MDA-MB-231 cells. *Western blot and quantification of ATR activation in RBMX depleted MDA-MB-231 cells treated with 4mM hydroxyurea for 5 hours (A) Protein analysis of Chk1 phosphorylation and total Chk1. (B) Analysis of RPA32 phosphorylation, RPA32 total, RBMX and α-tubulin expression. Mean values of the replicates were plotted with error bars indicating standard deviation. Statistical significance was calculated using GraphPad Prism v 9.31 with unpaired t test (ns = not significant).*

5.3.3.2 Investigating the effect of joint Camptothecin and RBMX depletion on DNA damage

I carried out preliminary experiments to test if RBMX depletion with Camptothecin sensitises cells to DNA damage. This was done by treating cells after RBMX siRNA depletion with 1μM of Camptothecin for 7hours. An initial experiment was done to monitor the amount of phospho Chk1 in Camptothecin treated cells after RBMX depletion. This showed a significant increase of phospho Chk1 in RBMX depleted cells compared to controls (Figure 5-10, A). An increase in Chk1 phosphorylation is

associated with a reduced replication associated breaks. All samples were treated with Camptothecin. There were no untreated controls carried out at the time the thesis was put together.

There were no significant changes seen in the amount of γH2AX (Figure 5-10, B) and phospho RPA32 (Figure 5-10, C) in RBMX depleted cells compared to the control cells when treated with Camptothecin. The amount of γH2AX and phosphorylated RPA32 in RBMX depleted cells is lower than in control cells. If these change were significant compared to untreated controls, it would suggest that there is reduced DNA damage in RBMX depleted cells after treatment with Camptothecin.
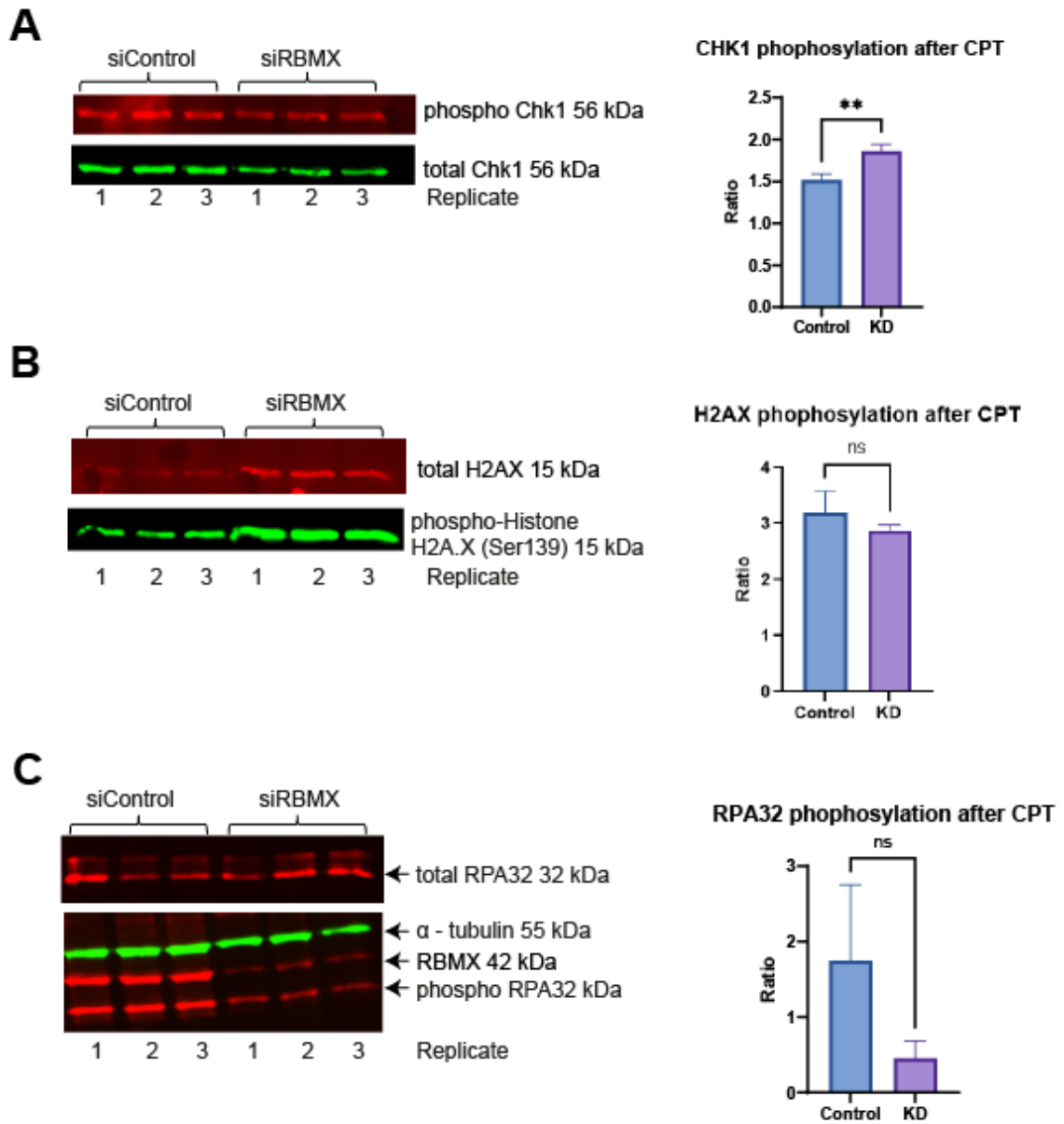
Figure 5-10 Treatment with Camptothecin in RBMX depleted MDA-MB231 causes an increase in Chk1 phosphorylation. *Western blot and quantification of ATR activation in RBMX depleted MDA-MB-231 cells treated with 1μM Camptothecin for 7 hours (A) Chk1 phosphorylation (B) γH2AX (C) RPA32 phosphorylation (ns=not significant)*

5.4   Discussion

The work in this chapter aimed to identify the genome instability effects after RBMX depletion. The experiments were preliminary. The work in this chapter shows that there is an increase in γH2AX indicating increased DNA damage when RBMX is depleted from MDA-MB-231 cells. However this increase is not intensified by the treatment of genotoxic drugs such as hydroxyurea and Camptothecin. The DNA damage assays were unable to all be finished within the constraints of this project due to the university closures during the Covid-19 pandemic. However, each of the preliminary experiments carried out and discussed in this chapter offer a potential future route to understanding the role of RBMX protein in the DNA damage response.

5.4.1   Depletion of RBMX leads to an increase in mitotic cells

Flow cytometry analysis of the proportion of cells in the different stages of the cell cycle showed an increase of cells in mitosis following RBMX depletion. Since RBMX is a known regulator of DNA damage, the cells were expected to arrest during the cell cycle. An arrest in mitosis is expected, due to faulty DNA repair caused by the absence of RBMX and replication fork proteins controlled by RBMX such as ETAA1. In chapter 4 I have shown that when RBMX is depleted ETAA1 is incorrectly spliced and does not give a full length protein. However there was no difference in the proportion of cells in the other stages of the cell cycle between RBMX depleted cells and control cells. It would be useful to repeat the cell cycle analysis in RBMX depleted cells treated with genotoxic drugs. This would investigate if RBMX depleted cells can effectively re-join the cell cycle after disruption by replication stress or if there is accumulation of cells in any of the phases of the cell cycle.

5.4.2   DNA damage effects in untreated RBMX depleted cells

Visual inspection of transcripts on IGV on the RNA-seq data showed that when RBMX was depleted there was a decrease in *TP53* mRNA and increase in *CDKN1A* which encodes for p21 protein a downstream target of p53. A reduction in p53 protein was seen in RBMX depleted cells **Figure 5-4**. This would suggest that the p53 pathway is disrupted after RBMX depletion, and so this could results in genomic instability. An increase of DNA damage marker γH2AX was seen upon RBMX depletion. No change was observed in the amount of PCNA protein when RBMX was depleted. PCNA is a

marker of proliferating cells and is a component of the DNA replication and repair pathway. Since no difference in PCNA was seen this suggests that the depletion of RBMX does not have a striking effect on the proliferation of MDA-MB-231 cells confirming the results shown by flow cytometry (Figure 5-3). I also showed that there was no change in RPA32 when RBMX was depleted from cells.

5.4.3 Joint RBMX depletion and treatment with genotoxic drugs did not increase DNA damage effects

I next wanted to test the effect of genotoxic drugs on cells that were depleted for RBMX using hydroxyurea and Camptothecin. In chapter 4 I showed that RBMX regulates exons in genes important for DNA repair. Thus we hypothesised that treatment with genotoxic drugs would increase DNA damaging effects after RBMX depletion and impair DNA repair. There was no evidence to support this hypothesis. I tested RBMX depleted cells before and after treatment with hydroxyurea and showed that there was no difference in the amount of phospho RPA32 by western blot or immunofluorescence. Interestingly there was an increase in RPA32 between untreated cells and hydroxyurea treated cells but this was independent of RBMX depletion. RPA32 is required for the recruitment of RAD51 a double strand break repair factor (Shi et al., 2010). There was no change in the amount of RAD51 when RBMX was depleted or when RBMX was depleted and cells treated with 4mM of hydroxyurea. An increase in γH2AX was seen after RBMX depletion but this was not further increased by HU treatment. However after 10 hours of hydroxyurea treatment there was inconsistent increase in γH2AX in the replicates of both RBMX depleted and control cells **Figure 5-7**. Future experiments carrying out repeats at different time points with smaller increments in time between collections may add more insight into the effects of hydroxyurea on RBMX depleted cells. Nevertheless after 5.5 hours of hydroxyurea treatment there was an increase in γH2AX in RBMX depleted cells compared to controls by immunofluorescence however, this was not quantified due to time restraints. Preliminary experiments showed a significant increase in Chk1 phosphorylation. A reduction on γH2AX and RPA32 was seen when RBMX depleted cells were treated with Camptothecin however, this was not significant. Chk1 phosphorylation is an indicator of activation in vivo in response to genotoxic stress. Form the literature it is unclear if the modest changes seen in Chk1 phosphorylation would cause phenotypic changes. However if the increase in Chk1 levels were to

cause a phenotypic change it would cause cells to accumulate prior to mitosis. It has been suggested that Chk1 plays an important role in DNA replication control required for cell proliferation (Wilsker et al., 2008).

Perhaps we have not seen a dramatic increase in DNA damage as expected because ETAA1 only regulates a subset of ATR targets and it's depletion has no discernible effect on CHK1 phosphorylation a well-recognised ATR target (Haahr et al., 2016).

5.4.4 **Future directions to investigate DNA damage effects cause after RBMX depletion**

It has been previously suggested by (Zheng et al., 2020) that depletion of RBMX causes an increase in defects of sister chromatid exchanges and an increase in DNA fragments indicating DNA damage seen by comet assay in HeLa cells (Zheng et al., 2020). Future experiments with an optimised comet assay, may prove valuable in determining the DNA damaging effects of RBMX in MDA-MB-231 cells. It has been shown that there is low ATR activation, low RPA32 and Chk1 phosphorylation in RBMX depleted U2OS cells (Zheng et al., 2020). This suggests that RBMX is unlikely to be involved in ATR activation. Conclusions by (Zheng et al., 2020) say that RBMX may be binding to ssDNA at sites of DNA Damage. RBMX localised in close proximity to RPA foci at sites of ssDNA damage in U2OS cells treated with Camptothecin (Zheng et al., 2020). RBMX used its RBM1CTR domain for binding to ssDNA. Future experiments carrying out immunofluorescence in MDA-MB-231 cells to monitor the localisation of RBMX to sites of DNA damage by staining cells with RBMX and γH2AX may give us an insight to the role of RBMX in these cells. In addition live immunofluorescence imaging of RBMX depleted cells to observe cell cycle progression may further elucidate the role of RBMX in the cell cycle.

A limitation of this study is the use of siRNA. The use of siRNA usually does not completely shut off the gene. Some functional RNA remains and is translated at lower levels so the protein is limited, but not eliminated. The DNA damage phenotypes would be more evident if the gene was completely knocked out. Using a gene knockout that involves the complete erasing of target genes such as CRISPR/Cas9 may display more drastic phenotypes.

5.5   Chapter summary

RBMX has been shown to have a role in chromosome biology and DNA damage response pathways. The role of RBMX in the DNA damage response pathway has been investigated in this chapter. Even though a slight increase in γH2AX was seen in RBMX depleted cells, it is unclear whether RBMX depletion has an impact on the DNA damage response pathway in MDA-MB-23 cells since due to time constraints the experiment's had to be cut short.

Chapter 6. Concluding remarks and future work

During gene expression RNAs undergo multiple steps of processing such capping, polyadenylation and mRNA splicing. Understanding protein-RNA interactions is crucial to understanding mechanisms of gene expression and protein synthesis in disease. Pre-mRNA splicing is an important step in the regulation of gene expression. During mRNA splicing protein-coding mRNA with multiple exons have the introns removed and exons joined together in varying patterns producing proteins with different abilities and functions this is called alternative splicing. Alternative splicing is coordinated by various RNA binding proteins. Identifying the RNA targets of these proteins helps put into focus their role in splicing. In my PhD I aimed to identify transcriptome wide targets of RNA binding protein RBMX in an attempt to decipher its role in mRNA splicing.

6.1 Identification of direct targets of RBMX using iCLIP

The work in this thesis describes work looking at transcriptome wide analysis of RBMX RNA binding targets using iCLIP in HEK293 cells. *RBMX* is expressed ubiquitously and is important in splicing control, transcription and genome integrity (Elliott et al., 2019). To investigate the direct binding targets of RBMX I carried out iCLIP. I used a modified version of iCLIP to identify the transcriptome wide binding sites of RBMX in RBMX-FLAG Flp-In HEK293 cells. The version of iCLIP I used incorporated methods from the original iCLIP (Huppertz et al., 2014), irCLIP (Brian J Zarnegar et al., 2016) and eCLIP (Van Nostrand et al., 2016) methods. This modified approach reduced the amount of RNAs lost during the purification processes. My results clearly indicate that RBMX binding was high in intronic regions of protein coding genes and intergenic regions. Interestingly gene ontology analysis shows that RBMX protein binding is enriched in transcripts from genes involved in DNA replication and cell cycle. The RBMX iCLIP also identified enrichment of RBMX binding sites in long exons. Importantly RBMX has been recognized as protecting cells from DNA damage in literature: RBMX is a positive regulator of homologous recombination and regulates chromosome segregation by its interaction with NORAD (Adamson et al., 2012), (Munschauer et al., 2018). My data suggests that some of these roles in DNA repair could also be indirect through regulation of genes that encode proteins involved in genome stability. Future studies could involve studying the binding sites of other RBMX

family of proteins and investigate if members of this family bind to the same transcript regions and if they have a similar role in mRNA splicing.

A possible limitation of my iCLIP analysis is that the number of CLIP-tags is influenced by gene expression and transcript levels within the cell. In the future normalisation of the iCLIP sequencing with an RNA-seq data set generated in the same cell line (RBMX-FLAG Flp-In HEK293 cells) will prove more powerful for studying the RBMX-RNA interactions. Both iCLIP-seq and RNA-seq methods have been used to understand TDP-43 binding and regulation (Hallegger et al., 2021).

6.2 Transcriptome analysis of mRNA seq data

A major aim of my work was to try and correlate patterns of RNA binding with the RNA processing patterns controlled by RBMX. I used RNA-seq performed prior to the start of this project to detect changes in the transcriptome of MDA-MB-231 cells after RBMX siRNA depletion. I was able to identify and characterise RBMX regulated exons. I also used the iCLIP data I generated from HEK293 cells to visually inspect the RNA processing targets of RBMX for RBMX binding sites close to the RNA processing event. Interestingly, RBMX was found to regulate exons within genes important for DNA replication and cell cycle where it is involved in protecting cells from DNA damage. Interestingly a paper from the Elliott lab showed that RBMXL2, a paralogue of RBMX, is important in preventing the use of cryptic splice sites in germ cells of male mice. Similarly in this study I have shown that RBMX is responsible for preventing the use of cryptic splice sites. RBMX and RBMXL2 proteins have 73% similarity RBMXL2 may be replacing RBMX during meiosis when RBMX is inactivated in the XY body. Thus it is likely that RBMX and RBMXL2 have similar functions in their respective cell types (somatic cells versus meiotic and post-meiotic germ cells).

One of the most interesting splicing and binding targets of RBMX in this study was *ETAA1* which encodes a key protein important for activation of ATR at stalled replication forks (Achuthankutty et al., 2019). Separate depletion of RBMX and Tra2β/α resulted in the use of a cryptic 3′ splice site in exon 5 of *ETAA1* in 3 different cancer cell lines. During my investigation into RBMX responsive minigenes I surprisingly found that RBMX overexpression increased exon skipping of the whole exon 5. The effect of RBMX on the ETAA1 mini genes was not as dramatic as the RNAseq data and PCR after RBMX depletion. However deletion of the RBMXL2 RRM increased use of the

weak splice site and reduced exon skipping compared to full length RBMXL2. These data suggest that minigenes do not adequately model ETAA1 cryptic splicing patterns. In particular, the minigene system does not take account chromatin state, and RBMX also interacts with RNA polymerase II suggesting it might function normally co-transcriptionally on endogenous genes. Further investigation is required to determine whether chromatin state affects the splicing of *ETAA1* exon 5.

6.3   Biological effect of RBMX knockdown on genome integrity

My global studies of RBMX function identified splicing changes in genes involved in genome stability. A number of lines of evidence suggest that low levels of RBMX expression are linked to poor survival in persons with cancer. Genome sequencing in lung cancer patients revealed truncation mutations in the *RBMX* gene indicating RBMX as a possible tumour suppressor (Renieri et al., 2014). Recently (Munschauer et al., 2018) have identified RBMX as being important for genome stability and prevention of cell cancer. RBMX has also been previously described as a tumour suppressor via its association with the hnRNP A1 protein together they modulate the alternative splicing of PKM inhibiting progression of bladder cancer (Yan et al., 2021). As a result we carried out some experiments to determine if RBMX depleted cells would have high DNA damage and cause increased replication defects. Depletion of RBMX was associated with accumulation of γH2AX foci in the nucleus suggesting that RBMX depletion increased DNA damage, however these results were not quantified. Surprisingly when RBMX depleted cells were treated with exogenous DNA damage using hydroxyurea or Camptothecin no additional increase in DNA damage was seen compared to cells that contain RBMX. These were only preliminary experiments. Future work involving treating RBMX depleted cells with the hydroxyurea or Camptothecin but for a shorter period of time such as 1 hour would be necessary to observe any transient increase in DNA damage. In addition the number of γH2AX foci in the RBMX depleted and control cells would need to be quantified before any conclusions can be made. The main challenges faced when carrying out the DNA damage analysis was that the phosphorylated proteins indicating defects in DNA damage and replication were difficult to detect, possibly due to the amount of protein used for the SDS-PAGE and western blotting, and the presence of phosphatases that destroyed the epitopes. Future experiments with more protein to detect defects in the ATM and ATR DNA repair pathways would be essential. In addition cell viability assays

to monitor the impact of RBMX depletion in cells would prove impactful to help understand the role of RBMX in cells.

6.4 Final conclusion

In this thesis I have identified RBMX as being involved in the mRNA processing of genes involved in DNA damage and replication. I have shown that RBMX is involved in directly binding within long exons of genes involved in DNA damage preventing use of cryptic mRNA processing sites thus promoting correct replication stress response and genome maintenance. Splicing changes have been identified as a hallmark of cancer. These cancer-specific splicing changes provide great avenues for new prognostic markers in cancer, as well as providing novel therapeutic targets (Gimeno-Valiente et al., 2022). RBMX could provide a viable cancer therapeutic target. RBMX depletion does lead a change of the transcriptome landscape and genes involved in DNA damage. Further experiments to investigate the important domains of RBMX such as the RRM and RGG domains would prove exciting.

Bibliography

Achuthankutty, D., Thakur, R.S., Haahr, P., Hoffmann, S., Drainas, A.P., Bizard, A.H., Weischenfeldt, J., Hickson, I.D. & Mailand, N. (2019) 'Regulation of ETAA1-mediated ATR activation couples DNA replication fidelity and genome stability', *The Journal of Cell Biology*, 218(12), p. 3943.

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', *Nature Cell Biology*, 14(3), pp. 318–328.

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Ech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., et al. (2018) 'The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update', *Nucleic Acids Research*, 46(W1), pp. W537–W544.

Akram, S., Yang, F., Li, J., Adams, G., Liu, Y., Zhuang, X., Chu, L., Liu, Xu, Emmett, N., Thompson, W., Mullen, M., Muthusamy, S., Wang, W., Mo, F. & Liu, Xing (2018) 'LRIF1 interacts with HP1α to coordinate accurate chromosome segregation during mitosis.', *Journal of molecular cell biology*, 10(6), pp. 527–538.

Anantha, R.W., Vassin, V.M. & Borowiec, J.A. (2007) 'Sequential and synergistic modification of human RPA stimulates chromosomal DNA repair', *Journal of Biological Chemistry*, 282(49), pp. 35910–35923.

Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. & Komorowski, J. (2009) 'Nucleosomes are well positioned in exons and carry characteristic histone modifications', *Genome Research*, 19(10), p. 1732.

Andreassen, P.R., Seo, J., Wiek, C. & Hanenberg, H. (2021) 'Understanding BRCA2 Function as a Tumor Suppressor Based on Domain-Specific Activities in DNA Damage Responses', *Genes*, 12(7), .

Anna, A. & Monika, G. (2018) 'Splicing mutations in human genetic disorders: examples, detection, and confirmation.', *Journal of applied genetics*, 59(3), pp.

253–268.

Ball, H.L., Myers, J.S. & Cortez, D. (2005) 'ATRIP binding to replication protein A-single-stranded DNA promotes ATR-ATRIP localization but is dispensable for Chk1 phosphorylation', *Molecular biology of the cell*, 16(5), pp. 2372–2381.

Bathke, J., Gauernack, A.S., Rupp, O., Weber, L., Preusser, C., Lechner, M., Rossbach, O., Goesmann, A., Evguenieva-Hackenberg, E. & Klug, G. (2020) 'iCLIP analysis of RNA substrates of the archaeal exosome', *BMC Genomics*, 21(1), .

Becker, J.S., McCarthy, R.L., Sidoli, S., Donahue, G., Kaeding, K.E., He, Z., Lin, S., Garcia, B.A. & Zaret, K.S. (2017) 'Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes', *Molecular Cell*, 68(6), pp. 1023-1037.e15.

Berniak, K., Rybak, P., Bernas, T., Zarebski, M., Biela, E., Zhao, H., Darzynkiewicz, Z. & Dobrucki, J.W. (2013) 'Relationship between DNA damage response, initiated by camptothecin or oxidative stress, and DNA replication, analyzed by quantitative 3D image analysis', *Cytometry Part A*, 83(10), pp. 913–924.

Best, A., James, K., Dalgliesh, C., Hong, E., Kheirolahi-Kouhestani, M., Curk, T., Xu, Y., Danilenko, M., Hussain, R., Keavney, B., Wipat, A., Klinck, R., Cowell, I.G., Cheong Lee, K., Austin, C.A., Venables, J.P., Chabot, B., Santibanez Koref, M., Tyson-Capper, A., et al. (2014) 'Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons', *Nature Communications 2014 5:1*, 5(1), pp. 1–15.

Black, D.L. (1995) 'Finding splice sites within a wilderness of RNA.' RNA (New York, N.Y.) 1 (8) p.pp. 763–771.

Black, D.L. (2003) 'Mechanisms of Alternative Pre-Messenger RNA Splicing', *Annual Review of Biochemistry*, 72(1), pp. 291–336.

Briese, M., Haberman, N., Sibley, C.R., Faraway, R., Elser, A.S., Chakrabarti, A.M., Wang, Z., König, J., Perera, D., Wickramasinghe, V.O., Venkitaraman, A.R., Luscombe, N.M., Saieva, L., Pellizzoni, L., Smith, C.W.J., Curk, T. & Ule, J. (2019) 'A systems view of spliceosomal assembly and branchpoints with iCLIP', *Nature Structural and Molecular Biology*, 26(10), pp. 930–940.

Burma, S., Chen, B.P., Murphy, M., Kurimasa, A. & Chen, D.J. (2001) 'ATM

phosphorylates histone H2AX in response to DNA double-strand breaks', *The Journal of biological chemistry*, 276(45), pp. 42462–42467.

Busch, A., Brüggemann, M., Ebersberger, S. & Zarnack, K. (2020) 'iCLIP data analysis: A complete pipeline from sequencing reads to RBP binding sites', *Methods*, 178pp. 49–62.

Cai, T., Cinkornpumin, J.K., Yu, Z., Villarreal, O.D., Pastor, W.A. & Richard, S. (2021) 'Deletion of RBMX RGG/RG motif in Shashi-XLID syndrome leads to aberrant p53 activation and neuronal differentiation defects', *Cell Reports*, 36(2), p. 109337.

Chow, L.T., Gelinas, R.E., Broker, T.R. & Roberts, R.J. (1977) 'An Amazing Sequence Arrangement at the 5' Ends of Adenovirus 2 Messenger RNA'. Cell 12.

Cieply, B. & Carstens, R.P. (2015) 'Functional roles of alternative splicing factors in human disease.', *Wiley Interdisciplinary reviews. RNA*, 6(3), pp. 311–326.

Cimprich, K.A. & Cortez, D. (2008) 'ATR: An Essential Regulator of Genome Integrity', *Nature reviews. Molecular cell biology*, 9(8), p. 616.

Cooper, T.A. (2005) 'Use of minigene systems to dissect alternative splicing elements', *Methods*, 37(4), pp. 331–340.

Dreumont, N., Bourgeois, C.F., Lejeune, F., Liu, Y., Ehrmann, I.E., Elliott, D.J. & Stévenin, J. (2010) 'Human RBMY regulates germline-specific splicing events by modulating the function of the serine/arginine-rich proteins 9G8 and Tra2-β', *Journal of Cell Science*, 123(1), pp. 40–50.

Ehrmann, I., Crichton, J.H., Gazzara, M.R., James, K., Liu, Y., Grellscheid, S.N., Curk, T., de Rooij, D., Steyn, J.S., Cockell, S., Adams, I.R., Barash, Y. & Elliott, D.J. (2019) 'An ancient germ cell-specific RNA-binding protein protects the germline from cryptic splice site poisoning', *eLife*, 8.

Ehrmann, I., Dalgliesh, C., Tsaousi, A., Paronetto, M.P., Heinrich, B., Kist, R., Cairns, P., Li, W., Mueller, C., Jackson, M., Peters, H., Nayernia, K., Saunders, P., Mitchell, M., Stamm, S., Sette, C. & 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), pp. 2803–2818.

Elguindy, M.M., Kopp, F., Goodarzi, M., Rehfeld, F., Thomas, A., Chang, T.C. & Mendell, J.T. (2019) 'PUMILIO, but not RBMX, binding is required for regulation of genomic stability by noncoding RNA NORAD', *eLife*, 8.

Elliott, D. & Ladomery, M. (2016) *Molecular biology of RNA*.

Elliott, D.J. (2004) 'The role of potential splicing factors including RBMY, RBMX, hnRNPG-T and STAR proteins in spermatogenesis*', *International Journal of Andrology*, 27(6), pp. 328–334.

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*, 108pp. 1–6.

Elliott, D.J., Venables, J.P., Newton, C.S., Lawson, D., Boyle, S., Eperon, I.C. & Cooke, H.J. (2000) 'An evolutionarily conserved germ cell-specific hnRNP is encoded by a retrotransposed gene'. Human Molecular Genetics 9 (14).

Feng, S., Zhao, Y., Xu, Y., Ning, S., Huo, W., Hou, M., Gao, G., Ji, J., Guo, R. & Xu, D. (2016) 'Ewing Tumor-associated Antigen 1 Interacts with Replication Protein A to Promote Restart of Stalled Replication Forks', *Journal of Biological Chemistry*, 291(42), pp. 21956–21962.

Gimeno-Valiente, F., López-Rodas, G., Castillo, J. & Franco, L. (2022) 'Alternative Splicing, Epigenetic Modifications and Cancer: A Dangerous Triangle, or a Hopeful One?', *Cancers*, 14(3), .

Gonatopoulos-Pournatzis, T. & Cowling, V.H. (2014) 'Cap-binding complex (CBC)', *Biochem. J*, 457pp. 231–242.

Granneman, S., Kudla, G., Petfalski, E. & Tollervey, D. (2009) 'Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs', *Proceedings of the National Academy of Sciences of the United States of America*, 106(24), pp. 9613–9618.

Guo, J., Wang, X., Jia, J. & Jia, R. (2020) 'Underexpression of SRSF3 and its target gene RBMX predicts good prognosis in patients with head and neck cancer', *Journal of Oral Science*, 62(2), pp. 175–179.

Haahr, P., Hoffmann, S., Tollenaere, M.A.X., Ho, T., Toledo, L.I., Mann, M., Bekker-Jensen, S., Räschle, M. & Mailand, N. (2016) 'Activation of the ATR kinase by

the RPA-binding protein ETAA1', *Nature Cell Biology*, 18(11), pp. 1196–1207.

Hallegger, M., Chakrabarti, A.M., Lee, F.C.Y., Lee, B.L., Amalietti, A.G., Odeh, H.M., Copley, K.E., Rubien, J.D., Portz, B., Kuret, K., Huppertz, I., Rau, F., Patani, R., Fawzi, N.L., Shorter, J., Luscombe, N.M. & Ule, J. (2021) 'TDP-43 condensation properties specify its RNA-binding and regulatory repertoire', *Cell*, 184(18), pp. 4680-4696.e22.

Hartman, M.L., Sztiller-Sikorska, M., Gajos-Michniewicz, A. & Czyz, M. (2020) 'Dissecting Mechanisms of Melanoma Resistance to BRAF and MEK Inhibitors Revealed Genetic and Non-Genetic Patient- and Drug-Specific Alterations and Remarkable Phenotypic Plasticity', *Cells*, 9(1), p. 142.

Heinrich, B., Zhang, Z., Raitskin, O., Hiller, M., Benderska, N., Hartmann, A.M., Bracco, L., Elliott, D., Ben-Ari, S., Soreq, H., Sperling, J., Sperling, R. & Stamm, S. (2009) 'Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regions in pre-mRNA', *The Journal of biological chemistry*, 284(21), .

Hemani, Y. & Soller, M. (2012) 'Mechanisms of Drosophila Dscam mutually exclusive splicing regulation', *Biochemical Society Transactions*, 40(4), pp. 804–809.

Hofmann, Y. & Wirth, B. (2002) 'hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1', *Human Molecular Genetics*, 11(17), pp. 2037–2049.

https://usegalaxy.org/ (n.d.) *https://usegalaxy.org/*. [Online] [online]. Available from: https://usegalaxy.org/ (Accessed 18 October 2022).

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J. & Ule, J. (2014) 'iCLIP: Protein–RNA interactions at nucleotide resolution', *Methods (San Diego, Calif.)*, 65(3), .

Hysenaj, G. (2020) *Investigating the regulation of alternative splicing by Tra2 proteins and RBMX in triple negative breast cancer cells*,

imaps.genialis.com (n.d.) *https://imaps.genialis.com/ - Genialis Workspace*. [Online] [online]. Available from: https://imaps.genialis.com/iclip (Accessed 18 October 2022).

Jeong, S. (2017) *SR Proteins: Binders, Regulators, and Connectors of RNA*, 40(1), .

JMardon, H., Sebastio, G. & EBaralle Sir William, F. (1987) *Nucleic Acids Research A rote for exon sequences in alternative splicing of the human flbronectin gene*.

Kalsotra, A. & Cooper, T.A. (2011) 'Functional consequences of developmentally regulated alternative splicing', *Nature Reviews Genetics*, 12(10), pp. 715–729.

Kanhoush, R., Beenders, B., Perrin, C., Moreau, J., Bellini, M. & Penrad-Mobayed, M. (2010a) 'Novel domains in the hnRNP G/RBMX protein with distinct roles in RNA binding and targeting nascent transcripts.', *Nucleus (Austin, Tex.)*, 1(1), pp. 109–22.

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

Kapustin, Y., Chan, E., Sarkar, R., Wong, F., Vorechovsky, I., Winston, R.M., Tatusova, T. & Dibb, N.J. (2011) 'Cryptic splice sites and split genes', *Nucleic Acids Research*, 39(14), pp. 5837–5844.

Katz, Y., Wang, E.T., Silterra, J., Schwartz, S., Wong, B., Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., Airoldi, E.M. & Burge, C.B. (2014) 'Sashimi plots: Quantitative visualization of alternative isoform expression from RNA-seq data', *bioRxiv*, p. 002576.

Kern, D.M., Kim, T., Rigney, M., Hattersley, N., Desai, A. & Cheeseman, I.M. (2015) 'The outer kinetochore protein KNL-1 contains a defined oligomerization domain in nematodes', *Molecular Biology of the Cell*, 26(2), pp. 229–237.

Kheirollahi Kouhestani, M. & Kouhestani, M.K. (2013) *Analysis of RNA binding proteins using stable cell lines*. [Online]. University of Newcastle Upon Tyne.

Kim, Y.-J. & Kim, H.-S. (2012) 'Alternative Splicing and Its Impact as a Cancer Diagnostic Marker', *Genomics & Informatics*, 10(2), p. 74.

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S. & Ahringer, J. (2009) 'Differential chromatin marking of introns and expressed exons by H3K36me3', *Nature genetics*, 41(3), p. 376.

König, J., Zarnack, K., Luscombe, N.M. & Ule, J. (2012) 'Protein–RNA interactions:

new genomic technologies and perspectives', *Nature Reviews Genetics*, 13(2), pp. 77–83.

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

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M. & Ule, J. (2010) 'iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution', *Nature Structural & Molecular Biology*, 17(7), pp. 909–915.

Královičová, J., Borovská, I., Pengelly, R., Lee, E., Abaffy, P., Šindelka, R., Grutzner, F. & Vořechovský, I. (2021) 'Restriction of an intron size en route to endothermy', *Nucleic Acids Research*, 49(5), pp. 2460–2487.

Krupp, M., Weinmann, A., Galle, P.R. & Teufel, A. (2006) 'Actin binding LIM protein 3 (abLIM3)', *International Journal of Molecular Medicine*, 17(1), pp. 129–133.

Lee, F.C.Y., Chakrabarti, A.M., Hänel, H., Monzón-Casanova, E., Hallegger, M., Militti, C., Capraro, F., Sadée, C., Toolan-Kerr, P., Wilkins, O., Turner, M., König, J., Sibley, C.R. & Ule, J. (2021) 'An improved iCLIP protocol', *bioRxiv*, p. 2021.08.27.457890.

Lee, Y.C., Zhou, Q., Chen, J. & Yuan, J. (2016) 'RPA-Binding Protein ETAA1 Is an ATR Activator Involved in DNA Replication Stress Response', *Current Biology*, 26(24), pp. 3257–3268.

Leman, R., Tubeuf, H., Raad, S., Tournier, I., Derambure, C., Lanos, R., Gaildrat, P., Castelain, G., Hauchard, J., Killian, A., Baert-Desurmont, S., Legros, A., Goardon, N., Quesnelle, C., Ricou, A., Castera, L., Vaur, D., Le Gac, G., Ka, C., et al. (2020) 'Assessment of branch point prediction tools to predict physiological branch points and their alteration by variants', *BMC Genomics*, 21(1), pp. 1–12.

Li, S., Zhang, J., Huang, S. & He, X. (2018) 'Genome-wide analysis reveals that exon methylation facilitates its selective usage in the human transcriptome', *Briefings in bioinformatics*, 19(5), pp. 754–764.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P. & Mesirov, J.P. (2011) 'Molecular signatures database (MSigDB) 3.0',

Bioinformatics, 27(12), pp. 1739–1740.

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

Lingenfelter, P.A., Delbridge, M.L., Thomas, S., Hoekstra, H.E., Mitchell, M.J., Marshall Graves, J.A. & Disteche, C.M. (2001) 'Expression and conservation of processed copies of the RBMX gene', *Mammalian Genome*, 12(7), pp. 538–545.

Liu, J., Doty, T., Gibson, B. & Heyer, W.D. (2010) 'Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA', *Nature structural & molecular biology*, 17(10), pp. 1260–1262.

Liu, N., Zhou, K.I., Parisien, M., Dai, Q., Diatchenko, L. & Pan, T. (2017) 'N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein', *Nucleic Acids Research*, 45(10), pp. 6051–6063.

Liu, Y., Bourgeois, C.F., Pang, S., Kudla, M., Dreumont, N., Kister, L., Sun, Y.-H., Stevenin, J. & Elliott, D.J. (2009) 'The germ cell nuclear proteins hnRNP G-T and RBMY activate a testis-specific exon.' Gregory P. Copenhaver (ed.), *PLoS genetics*, 5(11), p. e1000707.

Love, M.I., Huber, W. & Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), pp. 1–21.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M. & Misteli, T. (2010) 'Regulation of Alternative Splicing by HistoneModifications', *Science (New York, N.Y.)*, 327(5968), p. 996.

Lutz, C.S. (2008) 'Alternative Polyadenylation: A Twist on mRNA 3′ End Formation', *ACS Chemical Biology*, 3(10), pp. 609–617.

Ma, K., Inglis, J.D., Sharkey, A., Bickmore, W.A., Hill, R.E., Prosser, E.J.J., Speed, R.M., Thomson, E.J., Jobling, M., Taylor, K., Wolfe, J., Cooke, H.J., Hargreave, T.B. & Chandley, A.C. (1993) 'A Y chromosome gene family with RNA-binding protein homology: Candidates for the azoospermia factor AZF controlling human spermatogenesis', *Cell*, 75(7), pp. 1287–1295.

Ma, L., Jiang, Q.A., Sun, L., Yang, X., Huang, H., Jin, X., Zhang, C. & Wang, J.H.

(2020) 'X-linked rna-binding motif protein modulates hiv-1 infection of cd4+ T cells by maintaining the trimethylation of histone h3 lysine 9 at the downstream region of the 5= long terminal repeat of hiv proviral dna', *mBio*, 11(2), .

Matlin, A.J., Clark, F. & Smith, C.W.J. (2005) 'Understanding alternative splicing: towards a cellular code', *Nature Reviews Molecular Cell Biology*, 6(5), pp. 386–398.

Matsunaga, S., Takata, H., Morimoto, A., Hayashihara, K., Higashi, T., Akatsuchi, K., Mizusawa, E., Yamakawa, M., Ashida, M., Matsunaga, T.M.M., Azuma, T., Uchiyama, S. & Fukui, K. (2012) 'RBMX: A Regulator for Maintenance and Centromeric Protection of Sister Chromatid Cohesion', *Cell Reports*, 1(4), pp. 299–308.

McCarrey, J.R. & Thomas, K. (1987) 'Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene', *Nature*, 326(6112), pp. 501–505.

McClory, S.P., Lynch, K.W. & Ling, J.P. (2018) 'HnRNP L represses cryptic exons.', *RNA (New York, N.Y.)*, 24(6), pp. 761–768.

Meyer, K., Köster, T., Nolte, C., Weinholdt, C., Lewinski, M., Grosse, I. & Staiger, D. (2017) 'Adaptation of iCLIP to plants determines the binding landscape of the clock-regulated RNA-binding protein AtGRP7', *Genome Biology*, 18(1), .

Mintz, P.J., Patterson, S.D., Neuwald, A.F., Spahr, C.S. & Spector, D.L. (1999) 'Purification and biochemical characterization of interchromatin granule clusters', *The EMBO Journal*, 18(15), pp. 4308–4320.

Mokry, M., Feitsma, H., Nijman, I.J., de Bruijn, E., van der Zaag, P.J., Guryev, V. & Cuppen, E. (2010) 'Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries', *Nucleic Acids Research*, 38(10), p. e116.

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), p. 6659.

Munschauer, M., Nguyen, C.T., Sirokman, K., Hartigan, C.R., Hogstrom, L., Engreitz, J.M., Ulirsch, J.C., Fulco, C.P., Subramanian, V., Chen, J., Schenone, M., Guttman, M., Carr, S.A. & Lander, E.S. (2018) 'The NORAD lncRNA assembles

a topoisomerase complex critical for genome stability', *Nature*, 561(7721), pp. 132–136.

Nandakumar, P., Mansouri, A. & Das, S. (2017) 'The role of ATRX in glioma biology'. Frontiers in Oncology 7 (SEP) p.p. 236.

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), pp. 1337–1348.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M. & Yeo, G.W. (2016) 'Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)', *Nature Methods*, 13(6), pp. 508–514.

Ozaki, T. & Nakagawara, A. (2011) 'Role of p53 in cell death and human cancers'. Cancers 3 (1) p.pp. 994–1013.

Podhorecka, M., Skladanowski, A. & Bozko, P. (2010) 'H2AX phosphorylation: Its role in DNA damage response and cancer therapy'. Journal of Nucleic Acids 2010.

Prieto, C., Nguyen, D.T.T., Liu, Z., Wheat, J., Perez, A., Gourkanti, S., Chou, T., Barin, E., Velleca, A., Rohwetter, T., Chow, A., Taggart, J., Savino, A.M., Hoskova, K., Dhodapkar, M., Schurer, A., Barlowe, T.S., Vu, L.P., Leslie, C., et al. (2021) 'Transcriptional control of CBX5 by the RNA-binding proteins RBMX and RBMXL1 maintains chromatin state in myeloid leukemia', *Nature Cancer 2021 2:7*, 2(7), pp. 741–757.

Pucci, B. & Giordano, A. (1999) 'Cell cycle and cancer'. Clinica Terapeutica 150 (2) p.pp. 135–141.

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

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

Reinhardt, H.C., Aslanian, A.S., Lees, J.A. & Yaffe, M.B. (2007) 'p53-Deficient Cells Rely on ATM- and ATR-Mediated Checkpoint Signaling through the

p38MAPK/MK2 Pathway for Survival after DNA Damage', *Cancer Cell*, 11(2), pp. 175–189.

Renieri, A., Mencarelli, M.A., Cetta, F., Baldassarri, M., Mari, F., Furini, S., Piu, P., Ariani, F., Dragani, T.A. & Frullanti, E. (2014) 'Oligogenic germline mutations identified in early non-smokers lung adenocarcinoma patients', *Lung cancer (Amsterdam, Netherlands)*, 85(2), pp. 168–174.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. & Mesirov, J.P. (2011) 'Integrative genomics viewer', *Nature Biotechnology*, 29(1), pp. 24–26.

Shashi, V., Xie, P., Schoch, K., Goldstein, D.B., Howard, T.D., Berry, M.N., Schwartz, C.E., Cronin, K., Sliwa, S., Allen, A. & Need, A.C. (2015) 'The RBMX gene as a candidate for the Shashi X-linked intellectual disability syndrome', *Clinical Genetics*, 88(4), pp. 386–390.

Shi, W., Feng, Z., Zhang, Jiuqin, Gonzalez-Suarez, I., Vanderwaal, R.P., Wu, X., Powell, S.N., Roti Roti, J.L., Gonzalo, S. & Zhang, Junran (2010) 'The role of RPA2 phosphorylation in homologous recombination in response to replication arrest', *Carcinogenesis*, 31(6), pp. 994–1002.

Shiotani, B., Nguyen, H.D., Håkansson, P., Maréchal, A., Tse, A., Tahara, H. & Zou, L. (2013) 'Two Distinct Modes of ATR Activation Orchestrated by Rad17 and Nbs1', *Cell Reports*, 3(5), pp. 1651–1662.

Sibley, C.R., Blazquez, L. & Ule, J. (2016) 'Lessons from non-canonical splicing.', *Nature reviews. Genetics*, 17(7), pp. 407–421.

Simon, J.M., Hacker, K.E., Singh, D., Brannon, A.R., Parker, J.S., Weiser, M., Ho, T.H., Kuan, P.F., Jonasch, E., Furey, T.S., Prins, J.F., Lieb, J.D., Rathmell, W.K. & Davis, I.J. (2014) 'Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects', *Genome Research*, 24(2), p. 241.

Singh, A. & Xu, Y.J. (2016) 'The Cell Killing Mechanisms of Hydroxyurea', *Genes*, 7(11), .

Sivaramakrishnan, M., McCarthy, K.D., Campagne, S., Huber, S., Meier, S., Augustin, A., Heckel, T., Meistermann, H., Hug, M.N., Birrer, P., Moursy, A., Khawaja, S., Schmucki, R., Berntenis, N., Giroud, N., Golling, S., Tzouros, M.,

Banfai, B., Duran-Pacheco, G., et al. (2017) 'Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers', *Nature Communications*, 8(1), pp. 1–13.

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

Stark, R., Grzelak, M. & Hadfield, J. (2019) 'RNA sequencing: the teenage years', *Nature Reviews Genetics*,

Stoss, O., Stoilov, P., Hartmann, A.M., Nayler, O. & Stamm, S. (1999) 'The in vivo minigene approach to analyze tissue-specific splicing', *Brain research. Brain research protocols*, 4(3), pp. 383–394.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. & Mesirov, J.P. (2005) 'Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles', *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), pp. 15545–15550.

Suzanne, C. (2008) 'RNA Splicing: Introns, Exons and Spliceosome', *Nature Education*, 1(1), p. 31.

Taggart, A.J., Lin, C.L., Shrestha, B., Heintzelman, C., Kim, S. & Fairbrother, W.G. (2017) 'Large-scale analysis of branchpoint usage across species and cell lines', *Genome Research*, 27(4), pp. 639–649.

Takemoto, T., Nishio, Y., Sekine, O., Ikeuchi, C., Nagai, Y., Maeno, Y., Maegawa, H., Kimura, H. & Kashiwagi, A. (2007) 'RBMX is a novel hepatic transcriptional regulator of SREBP-1c gene response to high-fructose diet', *FEBS Letters*, 581(2), pp. 218–222.

Tian, B., Hu, J., Zhang, H. & Lutz, C.S. (2005) 'A large-scale analysis of mRNA polyadenylation of human and mouse genes', *Nucleic acids research*, 33(1), pp. 201–212.

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), pp. 1–11.

Trincado, J.L., Entizne, J.C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D.J. & Eyras, E. (2017) 'SUPPA2 provides fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions', *bioRxiv*, p. 086876.

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

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

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

Uyar, B., Yusuf, D., Wurmus, R., Rajewsky, N., Ohler, U. & Akalin, A. (2017) 'RCAS: An RNA centric annotation system for transcriptome-wide regions of interest', *Nucleic Acids Research*, 45(10), p. 91.

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', *Human Molecular Genetics*, 9(5), pp. 685–694.

Vogel, J., Hess, W.R. & Börner, T. (1997) 'Precise branch point mapping and quantification of splicing intermediates', *Nucleic acids research*, 25(10), pp. 2030–2031.

Wahl, M.C., Will, C.L. & Lührmann, R. (2009) 'The spliceosome: design principles of a dynamic RNP machine', *Cell*, 136(4), pp. 701–718.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. & Burge, C.B. (2008) 'Alternative isoform regulation in human tissue transcriptomes', *Nature*, 456(7221), pp. 470–476.

Wang, M., Wang, W., Ding, J., Wang, J. & Zhang, J. (2020) 'Downregulation of Rab17 promotes cell proliferation and invasion in non-small cell lung cancer through STAT3/HIF-1α/VEGF signaling', *Thoracic Cancer*, 11(2), p. 379.

Wang, P.J. (2004) 'X chromosomes, retrogenes and their role in male reproduction', *Trends in Endocrinology and Metabolism*, 15(2), pp. 79–83.

Wang, Y., Liu, J., Huang, B.O., Xu, Y.-M., Li, J., Huang, L.-F., Lin, J., Zhang, J., Min, Q.-H., Yang, W.-M. & Wang, X.-Z. (2015) 'Mechanism of alternative splicing and its regulation.', *Biomedical reports*, 3(2), pp. 152–158.

Wang, Z. & Burge, C.B. (2008) 'Splicing regulation: From a parts list of regulatory elements to an integrated splicing code', *RNA*, 14(5), p. 802.

Wilhelm, B.T., Marguerat, S., Aligianni, S., Codlin, S., Watt, S. & Bähler, J. (2011) 'Differential patterns of intronic and exonic DNA regions with respect to RNA polymerase II occupancy, nucleosome density and H3K36me3 marking in fission yeast', *Genome Biology*, 12(8), p. R82.

Xu, M., Katzenellenbogen, R.A., Grandori, C. & Galloway, D.A. (2010) 'NFX1 Plays a Role in Human Papillomavirus Type 16 E6 Activation of NFκB Activity', *Journal of Virology*, 84(21), pp. 11461–11469.

Yan, Q., Zeng, • Peng, Zhou, • Xiuqin, Zhao, X., Chen, • Runqiang, Qiao, J., Feng, L., Zhu, Z., Zhang, G. & Chen, • Cairong (2021) 'RBMX suppresses tumorigenicity and progression of bladder cancer by interacting with the hnRNP A1 protein to regulate PKM alternative splicing', *Oncogene*, 40pp. 2635–2650.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D. & Gage, F.H. (2009) 'An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells', *Nature Structural and Molecular Biology*, 16(2), pp. 130–137.

Zarnegar, Brian J., Flynn, R.A., Shen, Y., Do, B.T., Chang, H.Y. & Khavari, P.A. (2016) 'IrCLIP platform for efficient characterization of protein-RNA interactions', *Nature Methods*, 13(6), pp. 489–492.

Zarnegar, Brian J, Flynn, R.A., Shen, Y., Do, B.T., Chang, H.Y. & Khavari, P.A. (2016) 'irCLIP platform for efficient characterization of protein–RNA interactions', *Nature Methods*, 13(6), pp. 489–492.

Zheng, T., Zhou, H., Li, X., Peng, D., Yang, Y., Zeng, Y., Liu, H., Ren, J. & Zhao, Y. (2020) 'RBMX is required for activation of ATR on repetitive DNAs to maintain genome stability', *Cell Death & Differentiation*, 27(11), pp. 3162–3176.

Zhou, K.I., Shi, H., Lyu, R., Wylder, A.C., Matuszek, Ż., Pan, J.N., He, C., Parisien, M. & Pan, T. (2019) 'Regulation of Co-transcriptional Pre-mRNA Splicing by m6A through the Low-Complexity Protein hnRNPG', *Molecular Cell*, 76(1), pp. 70-

81.e9.

Appendix A: List of publications and conference attendances associated with this thesis

Saad Aldalaqan, Caroline Dalgliesh, Sara Luzzi, **Chileleko Siachisumo**, Louise N Reynard, Ingrid Ehrmann, and David J. Elliott. 'Cryptic splicing: common pathological mechanisms involved in male infertility and neuronal diseases', Cell Cycle. 2022 Feb;21(3):219-227. doi: 10.1080/15384101.2021.2015672. Epub 2021 Dec 20.

Sara Luzzi; Gerald Hysenaj; **Chileleko Siachisumo**; Ivaylo Yonchev; Kathleen Cheung; Matthew Gazzara; Katherine James; Caroline Dalgliesh; Mahsa Kheirollahi Chadegani; Ingrid Ehrmann; Graham R Smith; Simon J Cockell; Jennifer Munkley; Yoseph Barash; Stuart A Wilson; and David J Elliott. 'RBMX enables productive RNA processing of ultra-long exons important for genome stability', bioRxiv 2020.10.09.333039; doi: https://doi.org/10.1101/2020.10.09.333039.

North East Postgraduate Conference (NEPG) 2019 – Presentation talk

North East Postgraduate Conference (NEPG) 2020 – Presentation talk and chaired the EDI panel

3rd International Conference in Splicing, Caparica, Portugal, 2020 – Presentation talk – (winner of presentation talk prize)

4th International Conference in Splicing, Caparica, Portugal, 2021 – Presentation talk

RNA UK meeting 2022 – Presentation talk funded by the BSCB Honor Fell Travel Grant

Appendix B: List of all RBMX splicing and polyadenylation targets identified in MDA-MB-231 cells

Gene_name	event_location_hg38	Type_change	Annotated	RBMX_function	Canonical_PAS	Exon_size	CLIP_event_exon	CLIP_density
A4GALT	chr22:42,693,998-42,695,631	Alternative 5' splice site	Yes	Repressed by RBMX	-	141	-	-
ABCC5	chr3:183,981,875-183,982,773	Skipped exon	No	Activated by RBMX	-	174	-	-
ABHD6	chr3:58,237,917-58,256,561	Included exon	Yes	Repressed by RBMX	-	65	-	-
ABLIM3	chr5:149,239,889-149,244,880	Tandem included exon	Yes	Repressed by RBMX	-	99	-	-
ABLIM3	chr5:149,239,889-149,244,880	Tandem included exon	Yes	Repressed by RBMX	-	48	-	-
ABLIM3	chr5:149,200,429-149,205,108	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
ACIN1	chr14:23,081,837-23,090,521	Skipped exon	Yes	Activated by RBMX	-	120	-	-
ACOX1	chr17:75,957,567-75,978,533	Mutually exclusive exons	Yes	Repressed by RBMX	-	161	-	-
ACTR3B	chr7:152,759,767-152,801,620	Alternative first exon	Yes	Repressed by RBMX	-	160	-	-
ACYP1	chr14:75,053,660-75,063,469	Included exon	Yes	Repressed by RBMX	-	79	-	-
ADAM15	chr1:155,060,833-155,062,244	Skipped exon	Yes	Activated by RBMX	-	75	-	-
ADAM15	chr1:155,060,833-155,062,244	Skipped exon	Yes	Activated by RBMX	-	72	-	-
ADARB1	chr21:45,134,833-45,137,584	Upstream terminal exon	Yes	Repressed by RBMX	Yes	504	-	-
ADARB1	chr21:45,134,833-45,147,285	Upstream terminal exon	No	Repressed by RBMX	Yes	2655	Yes but near splice site	Low
AGAP3	chr7:151,120,146-151,123,638	Alternative terminal exon	Yes	Activated by RBMX	Yes	914	exon	Low
AGPAT4	chr6:161,129,979-161,136,634	Upstream termination site	No	Repressed by RBMX	No	6656	400nt downstream	Low
AHSA2P	chr2:61,181,406-61,184,711	Cryptic exon	Yes	Repressed by RBMX	-	-	-	-
AK4	chr1:65,226,063-65,232,145	Upstream termination site	No	Repressed by RBMX	Yes	6083	-	-
ALDH3A2	chr17:19,671,957-19,675,557	Included exon	Yes	Repressed by RBMX	-	125	-	-
ALG11	chr13:52,028,319-52,033,600	Upstream termination site	Yes	Repressed by RBMX	No	5282	In TW and downstream	Medium
ALG14	chr1:95,062,194-95,064,865	Cryptic upstream terminal exon	No	Repressed by RBMX	No	-	-	-
ALMS1	chr2:73,489,634-73,519,774	Alternative 5' splice site	Yes	Repressed by RBMX	-	1865	Yes	Medium
ALMS1	chr2:73,385,758-73,419,318	Alternative first exon	Yes	Activated by RBMX	-	435	-	-
ANKMY1	chr2:240,494,943-240,499,957	Upstream terminal exon	No	Repressed by RBMX	Yes	2495	-	-

ANKRD10	chr13:110,892,698-110,906,032	Upstream terminal exon	Yes	Repressed by RBMX	No	566	the intron upstream	Medium
ANKRD49	chr11:94,496,952-94,498,070	Alternative 5' splice site	Yes	Repressed by RBMX	-	515	-	-
ANTXR1	chr2:69,124,644-69,146,365	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1040	-	-
ANTXR1	chr2:69,124,644-69,135,140	Upstream terminal exon	Yes	Repressed by RBMX	Yes	4512	to splice site not TW	Low
ANTXR1	chr2:69,124,644-69,146,365	Cryptic exon	No	Repressed by RBMX	-	49	-	-
AP4E1	chr15:50,908,929-50,911,645	Cryptic exon	No	Repressed by RBMX	-	-	-	-
ARFIP1	chr4:152,863,715-152,872,451	Skipped exon	Yes	Activated by RBMX	-	96	-	-
ARHGAP26	chr5:143,222,358-143,229,007	Upstream termination site	Yes	Repressed by RBMX	No	6650	In TW and downstream	Low
ARHGAP26	chr5:143,057,748-143,069,369	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	-	-
ARHGAP26	chr5:142,770,377-142,873,399	Alternative first exon	No	Repressed by RBMX	-	177	-	-
ASPH	chr8:61,606,461-61,618,977	Upstream terminal exon	Yes	Repressed by RBMX	Yes	862	-	-
ASPH	chr8:61,624,605-61,633,682	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1676	-	-
ASPM	chr1:197,100,431-197,105,185	Upstream termination site	No	Repressed by RBMX	Yes	4755	Yes	High
ASPM	chr1:197,100,431-197,105,185	Upstream termination site	No	Repressed by RBMX	Yes	4755	Yes	High
ASPM	chr1:197,142,331-197,143,956	Alternative 3' splice site	No	Repressed by RBMX	-	1480	Yes	High
ASRGL1	chr11:62,392,079-62,393,415	Upstream termination site	Yes	Repressed by RBMX	No	1337	-	-
ASRGL1	chr11:62,337,457-62,356,986	Alternative first exon	No	Activated by RBMX	-	118	-	-
ASRGL1	chr11:62,357,145-62,383,247	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	-	-
ATF7IP	chr12:14,497,654-14,502,935	Upstream termination site	Yes	Repressed by RBMX	Yes	5282	-	-
ATMIN	chr16:81,043,161-81,047,350	Upstream termination site	Yes	Repressed by RBMX	No	4190	-	-
ATOH8	chr2:85,764,183-85,775,213	Cryptic upstream terminal exon	Yes	Repressed by RBMX	No	-	-	-
ATP11C	chrX:139,728,963-139,737,915	Included exon	Yes	Repressed by RBMX	-	105	-	-
ATRX	chrX:77,681,520-77,684,593	Exitron	Yes	Repressed by RBMX	-	3074	Yes	High
AXIN1	chr16:289,608-293,487	Skipped exon	Yes	Activated by RBMX	-	108	-	-
B3GLCT	chr13:31,286,820-31,315,046	Cryptic upstream terminal exon	No	Repressed by RBMX	No	-	-	-
B3GNTL1	chr17:83,011,428-83,014,453	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	-	-
BDNF	chr11:27,654,893-27,658,585	Upstream termination site	Yes	Repressed by RBMX	Yes	3693	-	-

Gene	Coordinates	Event type		RBMX effect				
BICD1	chr12:32,106,773-32,120,014	Cryptic upstream terminal exon	No	Repressed by RBMX	No	-	-	-
BICD1	chr12:32,377,540-32,383,637	Upstream termination site	Yes	Repressed by RBMX	Yes	6098	-	-
BRCA2	chr13:32,336,265-32,341,196	Upstream termination site	No	Repressed by RBMX	Yes	4932	Yes	High
BRCA2	chr13:32,336,265-32,341,196	Upstream termination site	No	Repressed by RBMX	Yes	4932	Yes	High
BRD1	chr22:49,787,390-49,794,033	Alternative 3' splice site	Yes	Activated by RBMX	-	498	-	-
C12orf73	chr12:103,951,631-103,956,630	Included exon	Yes	Repressed by RBMX	-	121	-	-
C4orf3	chr4:119,297,084-119,298,779	Upstream termination site	No	Repressed by RBMX	No	1696	-	-
C5orf22	chr5:31,545,713-31,551,292	Included exon	Yes	Repressed by RBMX	-	198	around the exon	Medium
C6orf89	chr6:36,886,029-36,899,425	Included exon	Yes	Repressed by RBMX	-	100	-	-
C6orf89	chr6:36,923,347-36,928,964	Exitron	No	Repressed by RBMX	-	5618	-	-
CALCOCO1	chr12:53,712,122-53,713,099	Skipped exon	Yes	Activated by RBMX	-	72	-	-
CAMK2G	chr10:73,817,555-73,821,681	Skipped exon	Yes	Activated by RBMX	-	114	-	-
CASK	chrX:41,555,636-41,559,778	Included exon	Yes	Repressed by RBMX	-	69	-	-
CBX5	chr12:54,271,841-54,280,007	Upstream terminal exon	Yes	Repressed by RBMX	Yes	2501	-	-
CBY1	chr22:38,656,751-38,668,016	Tandem included exon	Yes	Repressed by RBMX	-	188	-	-
CBY1	chr22:38,656,751-38,668,016	Tandem included exon	Yes	Repressed by RBMX	-	140	-	-
CC2D2A	chr4:15,510,241-15,511,675	Upstream terminal exon	Yes	Repressed by RBMX	No	429	-	-
CC2D2A	chr4:15,480,828-15,482,317	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1121	-	-
CCDC149	chr4:24,806,117-24,808,819	Upstream termination site	Yes	Repressed by RBMX	Yes	2703	-	-
CCDC93	chr2:117,949,504-117,952,372	Upstream terminal exon	Yes	Repressed by RBMX	No	2198	-	-
CCNL2	chr1:1,390,766-1,393,395	Alternative 3' splice site	Yes	Activated by RBMX	-	2038	Yes	Medium
CD40	chr20:46,122,757-46,126,639	Tandem included exon	Yes	Repressed by RBMX	-	94	-	-
CD40	chr20:46,123,220-46,128,137	Tandem included exon	Yes	Repressed by RBMX	-	62	-	-
CD44	chr11:35,198,247-35,201,195	Alternative 3' splice site	No	Repressed by RBMX	-	114	-	-
CD44	chr11:35,201,082-35,201,787	Intron retention	No	Repressed by RBMX	-	-	-	-
CD44	chr11:35,204,512-35,206,243	Intron retention	No	Repressed by RBMX	-	-	-	-
CD44	chr11:35,206,244-35,209,964	Tandem included exon	Yes	Repressed by RBMX	-	102	-	-
CD44	chr11:35,198,247-35,201,667	Tandem included exon	Yes	Repressed by RBMX	-	114	-	-
CD44	chr11:35,201,196-35,204,511	Tandem included exon	Yes	Repressed by RBMX	-	120	-	-
CD44	chr11:35,210,055-35,214,851	Tandem included exon	Yes	Repressed by RBMX	-	204	-	-
CD44	chr11:35,229,129-35,232,402	Upstream termination site	Yes	Repressed by RBMX	Yes	3274	-	-

CD44	chr11:35,190,066-35,198,120	Tandem included exon	Yes	Repressed by RBMX	-	129	-	-
CD44	chr11:35,196,875-35,201,081	Tandem included exon	Yes	Repressed by RBMX	-	126	-	-
CD44	chr11:35,201,788-35,206,111	Tandem included exon	Yes	Repressed by RBMX	-	129	-	-
CD44	chr11:35,208,207-35,211,245	Tandem included exon	Yes	Repressed by RBMX	-	90	-	-
CD44	chr11:35,204,641-35,208,104	Tandem included exon	Yes	Repressed by RBMX	-	132	-	-
CDH24	chr14:23,049,944-23,052,472	Skipped exon	Yes	Activated by RBMX	-	114	-	-
CDK11A	chr1:1,721,712-1,722,707	Skipped exon	Yes	Activated by RBMX	-	47	-	-
CDV3	chr3:133,586,723-133,590,261	Alternative 3' splice site	Yes	Repressed by RBMX	-	3103	Yes	Medium
CDV3	chr3:133,586,723-133,590,261	Alternative 3' splice site	Yes	Repressed by RBMX	-	3103	Yes	Medium
CEP126	chr11:101,961,741-101,963,880	Upstream termination site	No	Repressed by RBMX	Yes	2140	Yes but outside TW	Low
CEP57L1	chr6:109,095,576-109,145,218	Included exon	Yes	Repressed by RBMX	-	255	-	-
CEP63	chr3:134,546,289-134,549,061	Included exon	Yes	Repressed by RBMX	-	138	-	-
CFH	chr1:196,690,240-196,701,566	Upstream terminal exon	Yes	Repressed by RBMX	No	249	-	-
CFLAR	chr2:201,116,104-201,129,728	Alternative first exon	Yes	Repressed by RBMX	-	238	-	-
CHD3	chr17:7,907,489-7,907,893	Skipped exon	Yes	Activated by RBMX	-	102	-	-
CHKB	chr22:50,580,190-50,580,416	Intron retention	Yes	Repressed by RBMX	-	-	-	-
CHKB	chr22:50,581,749-50,582,357	Intron retention	Yes	Repressed by RBMX	-	-	-	-
CLK1	chr2:200,859,747-200,861,237	Included exon	Yes	Repressed by RBMX	-	91	-	-
COL8A1	chr3:99,675,722-99,703,306	Included exon	Yes	Repressed by RBMX	-	170	-	-
CPQ	chr8:96,880,006-96,969,077	Upstream terminal exon	No	Repressed by RBMX	Yes	3143	-	-
CREBZF	chr11:85,657,985-85,662,457	Upstream termination site	No	Repressed by RBMX	Yes	4473	Yes	Medium
CREBZF	chr11:85,663,524-85,665,129	Intron retention	Yes	Repressed by RBMX	-	-	-	-
CTNNBIP1	chr1:9,877,990-9,910,122	Alternative first exon	Yes	Repressed by RBMX	-	229	-	-
CTSC	chr11:88,326,086-88,328,111	Upstream terminal exon	Yes	Repressed by RBMX	Yes	358	-	-
CTSC	chr11:88,326,444-88,334,936	Included exon	Yes	Repressed by RBMX	-	85	-	-
CUL4A	chr13:113,229,520-113,230,240	Upstream terminal exon	Yes	Repressed by RBMX	No	599	Slightly downstream, one tag	Low
CYFIP2	chr5:157,392,833-157,395,594	Upstream termination site	Yes	Repressed by RBMX	No	2762	-	-
DEF8	chr16:89,957,661-89,959,366	Upstream terminal exon	Yes	Repressed by RBMX	No	353	-	-
DENND1A	chr9:123,401,585-123,403,401	Upstream terminal exon	Yes	Repressed by RBMX	Yes	328	Slightly downstream, few tags	Low

Gene	Coordinates	Event type		Regulation				
DENND4A	chr15:65,680,635-65,690,414	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	upstream intron	High
DEPDC1	chr1:68,481,613-68,483,949	Skipped exon	Yes	Activated by RBMX	-	852	Yes	Medium
DEPDC1B	chr5:60,629,287-60,638,749	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	Yes	Medium
DGKA	chr12:55,939,530-55,940,313	Skipped exon	Yes	Activated by RBMX	-	89	-	-
DGKA	chr12:55,931,162-55,936,422	Alternative first exon	Yes	Repressed by RBMX	-	61	-	-
DLC1	chr8:13,204,576-13,276,261	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
DMKN	chr19:35,510,184-35,510,476	Intron retention	Yes	Activated by RBMX	-	-	-	-
DMKN	chr19:35,510,253-35,513,678	Alternative first exon	Yes	Repressed by RBMX	-	629	-	-
DMTN	chr8:22,080,626-22,081,112	Skipped exon	Yes	Activated by RBMX	-	66	-	-
DNAJB6	chr7:157,385,009-157,386,318	Upstream terminal exon	Yes	Repressed by RBMX	Yes	778	exon not the TW	Low
DNASE1L1	chrX:154,409,197-154,411,890	Skipped exon	Yes	Activated by RBMX	-	258	-	-
DOCK5	chr8:25,268,886-25,271,221	Upstream terminal exon	Yes	Repressed by RBMX	No	432	-	-
DOCK5	chr8:25,411,194-25,415,711	Upstream termination site	Yes	Repressed by RBMX	No	4518	Yes but outside TW	Low
DOCK9	chr13:98,955,552-99,086,393	Alternative first exon	Yes	Activated by RBMX	-	207	-	-
DPF1	chr19:38,216,140-38,216,403	Intron retention	No	Activated by RBMX	-	-	-	-
DPYD	chr1:97,719,758-97,721,509	Upstream terminal exon	Yes	Repressed by RBMX	No	1169	-	-
DTNB	chr2:25,455,405-25,531,597	Upstream termination site	No	Repressed by RBMX	No	78	-	-
DUXAP10	chr14:19,292,526-19,296,351	Upstream termination site	Yes	Repressed by RBMX	Yes	3826	Yes	Low
E2F8	chr11:19,240,231-19,241,620	Alternative first exon	Yes	Repressed by RBMX	-	415	-	-
ECHDC2	chr1:52,912,022-52,921,721	Alternative first exon	Yes	Repressed by RBMX	-	169	-	-
ECHDC2	chr1:52,905,091-52,907,867	Included exon	Yes	Repressed by RBMX	-	93	-	-
EEF1AKMT3	chr12:57,773,129-57,780,254	Included exon	Yes	Repressed by RBMX	-	139	-	-
EFCAB2	chr1:245,083,689-245,091,918	Upstream terminal exon	Yes	Repressed by RBMX	No	790	-	-
EHMT1	chr9:137,790,971-137,794,173	Upstream terminal exon	Yes	Repressed by RBMX	No	2120	yes but in the exon and maybe branchpoint,	Medium
EIF2S1	chr14:67,383,315-67,386,516	Upstream termination site	Yes	Repressed by RBMX	Yes	3202	-	-
EIF4E2	chr2:232,567,215-232,569,210	Upstream terminal exon	Yes	Repressed by RBMX	Yes	266	-	-
ENDOV	chr17:80,428,661-80,430,371	Upstream terminal exon	Yes	Repressed by RBMX	No	571	-	-

Gene	Location	Event type		RBMX regulation				
EPB41L5	chr2:120,100,815-120,105,115	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1056	-	-
ERCC1	chr19:45,413,435-45,413,962	Upstream terminal exon	Yes	Repressed by RBMX	Yes	310	-	-
ERCC4	chr16:13,934,303-13,952,348	Alternative terminal exon	No	Activated by RBMX	Yes	2297	-	-
ESCO1	chr18:21,568,095-21,575,671	Skipped exon	No	Activated by RBMX	-	2117	Yes	High
ETAA1	chr2:67,402,975-67,405,335	Alternative 3' splice site	Yes	Repressed by RBMX	-	2111	Yes	Medium
EXO5	chr1:40,508,775-40,509,450	Alternative 5' splice site	Yes	Repressed by RBMX	-	134	-	-
EXO5	chr1:40,508,909-40,509,732	Included exon	Yes	Repressed by RBMX	-	175	-	-
EXOC7	chr17:76,090,398-76,094,413	Tandem included exon	Yes	Repressed by RBMX	-	93	-	-
EXOC7	chr17:76,089,321-76,091,142	Tandem included exon	Yes	Repressed by RBMX	-	69	-	-
FAM135A	chr6:70,524,122-70,526,698	Alternative 3' splice site	Yes	Activated by RBMX	-	1768	Yes	High
FAM135A	chr6:70,524,931-70,526,698	Intron retention	Yes	Activated by RBMX	-	1768	Yes	High
FAM172A	chr5:93,834,388-93,881,469	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	Yes but at the splice site	Low
FAM200A	chr7:99,546,295-99,548,506	Upstream termination site	No	Repressed by RBMX	Yes	2212	Yes but upstream/downstream	Low
FAM81A	chr15:59,458,647-59,460,986	Upstream terminal exon	Yes	Repressed by RBMX	No	1054	-	-
FAM86B1	chr8:12,185,526-12,186,699	Included exon	Yes	Repressed by RBMX	-	266	-	-
FAM86B1	chr8:12,186,834-12,188,310	Included exon	Yes	Repressed by RBMX	-	102	-	-
FANCB	chrX:14,864,560-14,865,580	Upstream termination site	No	Repressed by RBMX	Yes	1021	Yes	Medium
FANCM	chr14:45,175,071-45,176,976	Upstream termination site	No	Repressed by RBMX	Yes	1906	Yes	High
FASTKD3	chr5:7,865,984-7,868,978	Skipped exon	Yes	Activated by RBMX	-	1551	Yes	Medium
FASTKD3	chr5:7,865,984-7,869,037	Alternative first exon	Yes	Repressed by RBMX	-	59	-	-
FBXL20	chr17:39,252,663-39,261,567	Upstream termination site	Yes	Repressed by RBMX	Yes	8905	Yes	Low
FBXO31	chr16:87,367,373-87,383,404	Cryptic upstream terminal exon	Yes	Repressed by RBMX	No	-	-	-
FBXO31	chr16:87,355,633-87,360,294	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
FBXO38	chr5:148,425,702-148,433,423	Included exon	Yes	Repressed by RBMX	-	735	-	-
FBXO38	chr5:148,427,213-148,433,423	Alternative 5' splice site	Yes	Repressed by RBMX	-	735	-	-
FEM1B	chr15:68,289,607-68,295,862	Upstream termination site	No	Repressed by RBMX	Yes	6256	Yes but upstream/downstream	Low
FMNL3	chr12:49,657,191-49,661,965	Skipped exon	Yes	Activated by RBMX	-	153	-	-
FN1	chr2:215,391,815-215,394,527	Included exon	Yes	Repressed by RBMX	-	273	-	-

Gene	Location	Event	Col1	RBMX	Col2	Num	Notes	Level
FNIP2	chr4:158,870,470-158,881,229	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	-	-
FNIP2	chr4:158,833,629-158,835,404	Skipped exon	Yes	Activated by RBMX	-	90	-	-
FRMD5	chr15:43,870,767-43,874,462	Upstream termination site	Yes	Repressed by RBMX	No	3696	-	-
FUT10	chr8:33,451,728-33,461,371	Upstream terminal exon	Yes	Repressed by RBMX	No	1783	Yes but near the splice site	Low
FUT10	chr8:33,367,644-33,372,804	Upstream termination site	Yes	Repressed by RBMX	Yes	5161	upstream TW	Low
GALE	chr1:23,798,731-23,799,365	Skipped exon	Yes	Activated by RBMX	-	126	-	-
GALE	chr1:23,799,366-23,800,754	Alternative first exon	Yes	Repressed by RBMX	-	192	-	-
GANC	chr15:42,310,847-42,314,833	Upstream terminal exon	Yes	Repressed by RBMX	Yes	754	-	-
GATA3	chr10:8,073,739-8,074,894	Alternative termination site	Yes	Activated by RBMX	No	1156	-	-
GDPD5	chr11:75,477,796-75,490,236	Included exon	Yes	Repressed by RBMX	-	245	-	-
GEN1	chr2:17,780,621-17,784,719	Upstream termination site	No	Repressed by RBMX	No	4099	Yes	High
GFOD1	chr6:13,463,695-13,486,637	Upstream terminal exon	Yes	Repressed by RBMX	Yes	6936	Yes	Low
GFOD1	chr6:13,382,652-13,486,637	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
GGT1	chr22:24,610,034-24,610,531	Alternative 3' splice site	Yes	Repressed by RBMX	-	282	-	-
GLIS3	chr9:4,118,882-4,286,037	Included exon	Yes	Repressed by RBMX	-	208	-	-
GLYCTK	chr3:52,292,260-52,293,471	Intron retention	Yes	Repressed by RBMX	-	-	-	-
GLYCTK	chr3:52,290,960-52,291,922	Intron retention	Yes	Repressed by RBMX	-	-	-	-
GOLGA8B	chr15:34,565,890-34,582,749	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	-	-
GOLGB1	chr3:121,663,199-121,664,614	Upstream termination site	Yes	Repressed by RBMX	Yes	1416	upstream TW	Low
GOSR1	chr17:30,522,254-30,527,592	Upstream termination site	No	Repressed by RBMX	Yes	5339	-	-
GPR108	chr19:6,736,712-6,737,603	Alternative first exon	Yes	Activated by RBMX	-	332	-	-
GPR157	chr1:9,104,635-9,111,275	Skipped exon	Yes	Activated by RBMX	-	195	-	-
GPR161	chr1:168,079,543-168,085,796	Upstream termination site	Yes	Repressed by RBMX	No	6254	-	-
GRAMD1A	chr19:35,000,426-35,009,118	Alternative first exon	Yes	Repressed by RBMX	-	247	Yes (more in the introns downstream the second first exon	Low
GXYLT1	chr12:42,119,000-42,144,874	Included exon	Yes	Repressed by RBMX	-	93	-	-
H2AFY	chr5:135,346,058-135,360,496	Mutually exclusive exons	Yes	Activated by RBMX	-	91	-	-
HDAC8	chrX:72,473,416-72,488,932	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1314	-	-

HECTD2	chr10:91,461,357-91,463,664	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1570	-	-
HECW1	chr7:43,249,480-43,258,709	Alternative terminal exon	Yes	Activated by RBMX	No	175	-	-
HELQ	chr4:83,451,866-83,455,268	Upstream terminal exon	Yes	Repressed by RBMX	Yes	2080	-	-
HEXIM2	chr17:45,160,700-45,162,487	Alternative first exon	Yes	Repressed by RBMX	-	244	-	-
HINT2	chr9:35,812,960-35,813,338	Intron retention	Yes	Repressed by RBMX	-	-	-	-
HIVEP3	chr1:41,549,068-41,575,543	Cryptic upstream terminal exon	No	Repressed by RBMX	No	8088	-	-
HMMR	chr5:163,464,803-163,469,640	Skipped exon	Yes	Activated by RBMX	-	48	-	-
HPS1	chr10:98,429,643-98,430,570	Skipped exon	Yes	Activated by RBMX	-	99	-	-
HPS1	chr10:98,429,149-98,429,790	Upstream terminal exon	Yes	Repressed by RBMX	Yes	494	-	-
ICE1	chr5:5,436,477-5,438,735	Upstream terminal exon	No	Repressed by RBMX	Yes	1656	-	-
IFNAR2	chr21:33,260,728-33,262,756	Upstream terminal exon	Yes	Repressed by RBMX	Yes	2882	-	-
IGF1R	chr15:98,957,061-98,964,530	Upstream termination site	No	Repressed by RBMX	No	7470	TW	Low
ING3	chr7:120,955,625-120,956,980	Alternative terminal exon	Yes	Activated by RBMX	Yes	821	-	-
ING3	chr7:120,955,625-120,966,625	Included exon	Yes	Repressed by RBMX	-	97	-	-
ING4	chr12:6,653,397-6,656,726	Included exon	Yes	Repressed by RBMX	-	276	Slightly downstream the exon	Low
INSR	chr19:7,112,257-7,117,410	Upstream termination site	No	Repressed by RBMX	Yes	5154	-	-
IQCB1	chr3:121,781,875-121,790,072	Included exon	No	Repressed by RBMX	-	149	Yes	Low
IQCB1	chr3:121,826,181-121,828,860	Skipped exon	Yes	Activated by RBMX	-	163	-	-
IQSEC2	chrX:53,266,329-53,279,563	Upstream terminal exon	Yes	Repressed by RBMX	Yes	736	-	-
ISOC2	chr19:55,455,379-55,456,348	Skipped exon	Yes	Activated by RBMX	-	210	-	-
ITSN2	chr2:24,315,225-24,328,051	Included exon	Yes	Repressed by RBMX	-	67	-	-
JADE1	chr4:128,857,455-128,864,632	Upstream terminal exon	Yes	Repressed by RBMX	Yes	2929	-	-
JPT2	chr16:1,678,357-1,685,438	Included exon	Yes	Repressed by RBMX	-	84	-	-
KANSL1L	chr2:210,151,405-210,171,148	Upstream terminal exon	Yes	Repressed by RBMX	Yes	3207	-	-
KAT2A	chr17:42,118,404-42,119,436	Alternative 5' splice site	Yes	Activated by RBMX	-	383	-	-
KAT6A	chr8:42,049,303-42,051,900	Included exon	Yes	Repressed by RBMX	-	132	-	-
KAT6B	chr10:74,975,399-74,977,315	Alternative 5' splice site	Yes	Repressed by RBMX	-	932	-	-
KAT6B	chr10:74,975,399-74,977,315	Alternative 5' splice site	Yes	Repressed by RBMX	-	932	-	-
KCNQ3	chr8:132,180,157-132,184,240	Alternative 3' splice site	No	Activated by RBMX	-	2657	-	-
KCNQ3	chr8:132,283,277-132,321,444	Cryptic upstream terminal exon	No	Repressed by RBMX	No	-	-	-

Gene	Coordinates	Event type						
KDSR	chr18:63,327,726-63,331,901	Upstream termination site	Yes	Repressed by RBMX	No	4176	-	-
KHDC1	chr6:73,274,136-73,291,997	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1421	-	-
KIAA0586	chr14:58,512,628-58,522,186	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	Yes	Medium
KIAA1217	chr10:24,536,894-24,544,980	Mutually exclusive exons	Yes	Activated by RBMX	-	1599	-	-
KIF1B	chr1:10,297,247-10,308,597	Upstream terminal exon	Yes	Repressed by RBMX	Yes	5435	Yes	Medium
KLHL2	chr4:165,317,970-165,322,031	Included exon	Yes	Repressed by RBMX	-	64	-	-
KLHL36	chr16:84,661,578-84,667,686	Upstream termination site	Yes	Repressed by RBMX	No	6109	-	-
KLHL5	chr4:39,045,097-39,075,964	Included exon	Yes	Repressed by RBMX	-	477	Yes	Low
KLHL7	chr7:23,125,173-23,126,050	Upstream terminal exon	Yes	Repressed by RBMX	Yes	290	-	-
KMT5B	chr11:68,165,717-68,171,014	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1462	Yes but at the splice site	Low
KNL1	chr15:40,620,640-40,625,640	Upstream termination site	No	Repressed by RBMX	Yes	5001	Yes	High
KNL1	chr15:40,620,640-40,625,640	Upstream termination site	No	Repressed by RBMX	Yes	5001	Yes	High
KRBA2	chr17:8,370,387-8,375,899	Included exon	Yes	Repressed by RBMX	-	773	-	-
LACTB	chr15:63,127,690-63,131,467	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1983	-	-
LAMA1	chr18:6,959,341-6,961,759	Alternative 5' splice site	No	Repressed by RBMX	-	542	-	-
LATS1	chr6:149,662,239-149,676,259	Included exon	Yes	Repressed by RBMX	-	99	-	-
LCORL	chr4:17,873,388-17,878,213	Upstream termination site	No	Repressed by RBMX	Yes	4826	Yes	High
LETMD1	chr12:51,049,186-51,056,502	Tandem included exon	Yes	Repressed by RBMX	-	116	-	-
LETMD1	chr12:51,049,186-51,056,502	Tandem included exon	Yes	Repressed by RBMX	-	102	-	-
LETMD1	chr12:51,049,186-51,056,502	Tandem included exon	Yes	Repressed by RBMX	-	83	-	-
LETMD1	chr12:51,049,186-51,056,502	Tandem included exon	Yes	Repressed by RBMX	-	187	-	-
LGALSL	chr2:64,458,285-64,461,381	Upstream termination site	No	Repressed by RBMX	Yes	3097	-	-
LIMD1	chr3:45,594,105-45,636,149	Alternative first exon	No	Activated by RBMX	-	189	-	-
LIMS2	chr2:127,654,424-127,654,896	Intron retention	Yes	Repressed by RBMX	-	-	-	-
LMO7	chr13:75,777,766-75,795,400	Alternative first exon	No	Repressed by RBMX	-	325	-	-
LPIN3	chr20:41,357,040-41,357,447	Intron retention	No	Activated by RBMX	-	-	-	-
LRIF1	chr1:110,950,124-110,963,620	Skipped exon	Yes	Activated by RBMX	-	1528	Yes	Medium
LRP10	chr14:22,876,940-22,881,580	Upstream termination site	No	Repressed by RBMX	Yes	4641	-	-
LRP6	chr12:12,198,973-12,203,202	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	Yes	Medium
LRR1	chr14:49,602,469-49,614,255	Skipped exon	Yes	Activated by RBMX	-	722	Yes	Low
LRRC8E	chr19:7,888,601-7,895,598	Included exon	Yes	Repressed by RBMX	-	287	-	-

LRRFIP1	chr2:237,708,631-237,748,363	Tandem included exon	Yes	Repressed by RBMX	-	45	-	-
LRRFIP1	chr2:237,708,631-237,748,363	Tandem included exon	Yes	Repressed by RBMX	-	51	-	-
LRRFIP1	chr2:237,708,631-237,748,363	Tandem included exon	Yes	Repressed by RBMX	-	18	-	-
LRRFIP1	chr2:237,708,631-237,748,363	Tandem included exon	Yes	Repressed by RBMX	-	48	-	-
LRRFIP1	chr2:237,708,631-237,748,363	Tandem included exon	Yes	Repressed by RBMX	-	60	-	-
LYRM2	chr6:89,634,014-89,637,353	Upstream termination site	No	Repressed by RBMX	Yes	3340	Yes	Medium
LYST	chr1:235,826,323-235,830,225	Upstream terminal exon	Yes	Repressed by RBMX	Yes	2714	-	-
MANBA	chr4:102,690,772-102,723,861	Skipped exon	No	Activated by RBMX	-	124	-	-
MANBA	chr4:102,690,772-102,723,861	Skipped exon	Yes	Activated by RBMX	-	171	-	-
MANBAL	chr20:37,289,687-37,298,762	Included exon	Yes	Repressed by RBMX	-	239	-	-
MAP3K20	chr2:173,217,251-173,227,146	Upstream terminal exon	Yes	Repressed by RBMX	Yes	5997	-	-
MAP3K8	chr10:30,437,407-30,439,481	Upstream terminal exon	Yes	Repressed by RBMX	Yes	566	Yes but near the splice site	Low
MAP4K4	chr2:101,870,416-101,873,646	Included exon	Yes	Repressed by RBMX	-	192	-	-
MAP4K4	chr2:101,859,865-101,864,929	Tandem included exon	Yes	Repressed by RBMX	-	162	-	-
MAP4K4	chr2:101,859,865-101,864,929	Tandem included exon	Yes	Repressed by RBMX	-	231	-	-
MAP4K4	chr2:101,867,310-101,869,621	Included exon	Yes	Repressed by RBMX	-	9	-	-
MAPK13	chr6:36,139,294-36,144,521	Upstream termination site	Yes	Repressed by RBMX	Yes	5228	-	-
MAPK9	chr5:180,246,028-180,248,972	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1483	-	-
MARK1	chr1:220,579,558-220,582,898	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1834	-	-
MAZ	chr16:29,806,648-29,808,569	Alternative first exon	Yes	Activated by RBMX	-	246	-	-
MBD1	chr18:50,268,846-50,271,540	Intron retention	Yes	Activated by RBMX	-	-	Yes	Low
MBD5	chr2:148,463,920-148,472,455	Upstream terminal exon	Yes	Repressed by RBMX	Yes	4115	-	-
MBNL3	chrX:132,379,678-132,382,177	Cryptic exon	Yes	Repressed by RBMX	-	-	-	-
MCTP1	chr5:94,819,119-94,868,332	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	-	-
MECP2	chrX:154,021,573-154,031,450	Upstream termination site	Yes	Repressed by RBMX	Yes	9878	-	-
MELTF	chr3:197,018,953-197,021,403	Upstream terminal exon	Yes	Repressed by RBMX	No	849	-	-
METTL6	chr3:15,414,163-15,415,771	Skipped exon	Yes	Activated by RBMX	-	86	-	-
METTL6	chr3:15,409,760-15,414,020	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1678	Yes a bit upstream TW	Low
MFSD8	chr4:127,917,889-127,920,836	Alternative termination site	Yes	Activated by RBMX	Yes	2948	-	-
MIOS	chr7:7,572,436-7,573,769	Upstream termination site	Yes	Repressed by RBMX	Yes	1334	Yes	Low
MIPOL1	chr14:37,198,105-37,247,102	Tandem included exon	Yes	Repressed by RBMX	-	137	-	-

MIPOL1	chr14:37,197,950-37,247,102	Alternative first exon	Yes	Repressed by RBMX	-	137	-	-
MIPOL1	chr14:37,198,105-37,247,102	Tandem included exon	Yes	Repressed by RBMX	-	113	-	-
MIPOL1	chr14:37,198,105-37,247,102	Tandem included exon	Yes	Repressed by RBMX	-	139	-	-
MKI67	chr10:128,113,603-128,116,490	Skipped exon	Yes	Activated by RBMX	-		Yes	Medium
MKKS	chr20:10,412,530-10,413,931	Upstream termination site	No	Repressed by RBMX	Yes	1402	-	-
MLEC	chr12:120,696,316-120,701,859	Upstream termination site	Yes	Repressed by RBMX	No	5544	-	-
MMAA	chr4:145,655,147-145,660,033	Upstream termination site	Yes	Repressed by RBMX	Yes	4887	-	-
MOB1A	chr2:74,152,528-74,156,645	Upstream termination site	No	Repressed by RBMX	Yes	4118	Yes	Low
MOSPD1	chrX:134,893,292-134,899,089	Upstream terminal exon	No	Repressed by RBMX	No	3743	-	-
MPPE1	chr18:11,897,357-11,906,202	Tandem included exon	Yes	Repressed by RBMX	-	95	-	-
MPPE1	chr18:11,897,357-11,906,202	Tandem included exon	Yes	Repressed by RBMX	-	216	-	-
MROH1	chr8:144,200,542-144,213,465	Upstream terminal exon	Yes	Repressed by RBMX	Yes	472	-	-
MSLN	chr16:760,734-762,671	Alternative first exon	Yes	Activated by RBMX	-	88	-	-
MTA1	chr14:105,445,512-105,450,057	Skipped exon	Yes	Activated by RBMX	-	51	Yes	Low
MTBP	chr8:120,470,938-120,478,228	Cryptic upstream terminal exon	No	Repressed by RBMX	No	-	Yes possibly branchpoint and	Low
MTERF1	chr7:91,874,765-91,880,054	Included exon	Yes	Repressed by RBMX	-	101	-	-
MTERF4	chr2:241,095,566-241,097,242	Upstream terminal exon	Yes	Repressed by RBMX	Yes	873	Yes	Low
MYO18A	chr17:29,122,254-29,180,412	Alternative first exon	Yes	Activated by RBMX	-	86	-	-
N4BP2L2	chr13:32,527,533-32,537,027	Alternative 5' splice site	Yes	Repressed by RBMX	-	1259	Yes	High
N4BP2L2	chr13:32,527,533-32,538,617	Skipped exon	Yes	Activated by RBMX	-	1259	Yes	High
NASP	chr1:45,606,592-45,613,168	Skipped exon	Yes	Activated by RBMX	-	1017	Yes	High
NAT1	chr8:18,210,181-18,219,410	Tandem included exon	Yes	Repressed by RBMX	-	150	-	-
NAT1	chr8:18,210,181-18,219,410	Tandem included exon	Yes	Repressed by RBMX	-	384	-	-
NAT1	chr8:18,210,181-18,219,410	Tandem included exon	Yes	Repressed by RBMX	-	118	-	-
NAV1	chr1:201,819,837-201,826,969	Upstream termination site	No	Repressed by RBMX	Yes	7133	Yes	Low
NBPF3	chr1:21,445,220-21,457,147	Upstream terminal exon	No	Repressed by RBMX	Yes	3775	Yes	Low
NCOR1	chr17:16,086,443-16,091,862	Included exon	Yes	Repressed by RBMX	-	174	-	-
NEDD4L	chr18:58,396,167-58,401,540	Upstream termination site	No	Repressed by RBMX	Yes	5374	-	-
NFAT5	chr16:69,691,749-69,694,239	Exitron	No	Repressed by RBMX	-	2491	Yes	High
NFIA	chr1:61,455,303-61,462,788	Upstream termination site	Yes	Repressed by RBMX	Yes	7486	-	-
NFX1	chr9:33,347,118-33,348,731	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1069	Yes	Low

NIN	chr14:50,754,868-50,759,856	Skipped exon	Yes	Activated by RBMX	-	2139	Yes	Medium
NIPA1	chr15:22,823,728-22,829,789	Upstream termination site	No	Repressed by RBMX	No	6062	-	-
NIPAL3	chr1:24,463,825-24,467,461	Upstream terminal exon	Yes	Repressed by RBMX	Yes	703	-	-
NLK	chr17:28,043,332-28,048,926	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1010	-	-
NLRX1	chr11:119,182,094-119,184,016	Intron retention	No	Activated by RBMX	-	-	-	-
NME6	chr3:48,290,279-48,294,803	Upstream termination site	Yes	Repressed by RBMX	Yes	4525	Yes	Low
NPC1	chr18:23,562,424-23,568,822	Cryptic upstream terminal exon	No	Repressed by RBMX	No	-	-	-
NPC1	chr18:23,567,087-23,568,822	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
NPEPL1	chr20:58,701,159-58,705,171	Upstream terminal exon	No	Repressed by RBMX	Yes	2249	Yes	Low
NPR2	chr9:35,800,842-35,801,642	Skipped exon	No	Activated by RBMX	-	85	-	-
NPR2	chr9:35,808,509-35,808,853	Intron retention	Yes	Repressed by RBMX	-	-	-	-
NSD1	chr5:177,135,087-177,136,030	Intron retention	Yes	Activated by RBMX	-	-	Yes	Medium
NTNG2	chr9:132,162,240-132,170,319	Upstream terminal exon	No	Repressed by RBMX	Yes	3971	-	-
NUDT16	chr3:131,383,162-131,388,830	Upstream termination site	Yes	Repressed by RBMX	No	5669	-	-
NUDT7	chr16:77,725,585-77,736,649	Tandem included exon	Yes	Repressed by RBMX	-	422	-	-
NUDT7	chr16:77,735,987-77,741,581	Tandem included exon	Yes	Repressed by RBMX	-	94	-	-
NUP98	chr11:3,711,835-3,713,817	Upstream terminal exon	Yes	Repressed by RBMX	No	894	-	-
NXPE3	chr3:101,821,404-101,827,392	Upstream termination site	Yes	Repressed by RBMX	No	5989	-	-
NXT2	chrX:109,536,844-109,538,044	Alternative first exon	Yes	Repressed by RBMX	-	178	-	-
OCEL1	chr19:17,226,693-17,227,199	Intron retention	Yes	Repressed by RBMX	-	-	-	-
ODF2L	chr1:86,349,327-86,352,858	Alternative terminal exon	Yes	Activated by RBMX	No	2882	Yes	Medium
ODF2L	chr1:86,354,860-86,358,786	Included exon	Yes	Repressed by RBMX	-	159	-	-
OGFOD3	chr17:82,392,535-82,398,195	Included exon	Yes	Repressed by RBMX	-	213	-	-
PACS2	chr14:105,368,540-105,373,008	Upstream terminal exon	No	Repressed by RBMX	No	457	Yes	High
PBX1	chr1:164,846,584-164,851,831	Upstream termination site	Yes	Repressed by RBMX	Yes	5248	-	-
PBX1	chr1:164,559,184-164,792,493	Alternative first exon	Yes	Activated by RBMX	-	379	-	-
PCDHGC3	chr5:141,350,116-141,494,806	Alternative first exon	Yes	Repressed by RBMX	-	2581	-	-
PDE10A	chr6:165,327,485-165,333,127	Upstream termination site	No	Repressed by RBMX	Yes	5643	-	-
PDIK1L	chr1:26,111,213-26,114,291	Alternative first exon	Yes	Repressed by RBMX	-	112	-	-
PDIK1L	chr1:26,111,916-26,114,291	Alternative first exon	No	Repressed by RBMX	-	191	-	-
PEMT	chr17:17,541,096-17,576,919	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-

Gene	Location	Event type		RBMX				
PEX13	chr2:61,017,852-61,019,099	Upstream terminal exon	Yes	Repressed by RBMX	Yes	938	-	-
PFN2	chr3:149,964,904-149,968,357	Alternative 3' splice site	Yes	Repressed by RBMX	-	1683	Yes	Low
PGGHG	chr11:293,163-293,727	Intron retention	Yes	Repressed by RBMX	-	-	-	-
PHF20	chr20:35,772,080-35,775,058	Upstream terminal exon	Yes	Repressed by RBMX	Yes	571	-	-
PI4KB	chr1:151,316,510-151,327,270	Tandem included exon	Yes	Repressed by RBMX	-	202	-	-
PI4KB	chr1:151,316,510-151,327,270	Tandem included exon	Yes	Repressed by RBMX	-	173	-	-
PICALM	chr11:85,983,974-85,996,825	Included exon	Yes	Repressed by RBMX	-	150	-	-
PIGX	chr3:196,722,557-196,727,922	Skipped exon	Yes	Activated by RBMX	-	117	-	-
PITPNM3	chr17:6,460,569-6,461,372	Upstream terminal exon	Yes	Repressed by RBMX	Yes	400	-	-
PLPP5	chr8:38,264,739-38,266,140	Upstream terminal exon	Yes	Repressed by RBMX	Yes	211	-	-
PML	chr15:74,024,928-74,028,198	Upstream terminal exon	No	Repressed by RBMX	Yes	2613	-	-
PMS1	chr2:189,805,755-189,807,054	Cryptic upstream terminal exon	Yes	Repressed by RBMX	No	-	-	-
PMS1	chr2:189,805,755-189,818,016	Cryptic exon	Yes	Repressed by RBMX	-	-	-	-
POLR1D	chr13:27,622,010-27,623,449	Upstream terminal exon	Yes	Repressed by RBMX	Yes	575	Yes	Low
POLR2D	chr2:127,843,553-127,848,185	Upstream termination site	Yes	Repressed by RBMX	Yes	4633	Yes	Medium
POLR2J3	chr7:102,566,997-102,568,009	Upstream terminal exon	No	Repressed by RBMX	No	87	Yes	Low
PPARD	chr6:35,347,151-35,358,403	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
PPDPF	chr20:63,520,895-63,521,511	Skipped exon	Yes	Activated by RBMX	-	80	-	-
PPHLN1	chr12:42,441,315-42,448,623	Alternative terminal exon	Yes	Activated by RBMX	Yes	633	-	-
PPM1K	chr4:88,257,620-88,262,726	Upstream termination site	Yes	Repressed by RBMX	Yes	5107	-	-
PPM1M	chr3:52,246,813-52,247,681	Skipped exon	Yes	Activated by RBMX	-	255	-	-
PPP1R35	chr7:100,435,534-100,435,847	Skipped exon	Yes	Activated by RBMX	-	137	-	-
PPP1R3F	chrX:49,282,064-49,288,090	Upstream terminal exon	Yes	Repressed by RBMX	Yes	2257	-	-
PRKAG1	chr12:49,005,853-49,013,061	Included exon	Yes	Repressed by RBMX	-	174	Yes but branchpoint not exon	Low
PRKX	chrX:3,609,875-3,612,176	Upstream terminal exon	No	Repressed by RBMX	Yes	1460	-	-
PRMT2	chr21:46,649,740-46,654,297	Upstream terminal exon	No	Repressed by RBMX	No	1453	Yes	Medium
PRRC2B	chr9:131,495,740-131,500,197	Upstream termination site	No	Repressed by RBMX	Yes	4458	Yes	Medium
PSPH	chr7:56,032,055-56,051,448	Alternative first exon	Yes	Repressed by RBMX	-	311	-	-
PTPN13	chr4:86,693,675-86,716,529	Included exon	Yes	Repressed by RBMX	-	561	-	-
PTPRK	chr6:128,235,457-128,240,034	Upstream terminal exon	Yes	Repressed by RBMX	No	148	-	-

Gene	Location	Event type		RBMX effect		Count		Level
PYGL	chr14:50,923,237-50,931,672	Upstream terminal exon	Yes	Repressed by RBMX	Yes	864	-	-
PYROXD2	chr10:98,402,392-98,407,581	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
RAD51B	chr14:68,468,251-68,470,717	Upstream terminal exon	Yes	Repressed by RBMX	No	195	Yes but in the intron before	Low
RAD51B	chr14:68,468,251-68,497,881	Upstream terminal exon	Yes	Repressed by RBMX	No	758	site / branchpoint	Low
RAET1L	chr6:150,018,334-150,020,904	Alternative terminal exon	Yes	Activated by RBMX	No	522	-	-
RAPH1	chr2:203,461,926-203,489,583	Tandem included exon	Yes	Repressed by RBMX	-	81	-	-
RAPH1	chr2:203,461,926-203,489,583	Tandem included exon	Yes	Repressed by RBMX	-	75	-	-
RBM41	chrX:107,061,885-107,067,693	Exitron	Yes	Repressed by RBMX	-	5809	Yes	High
RBMX	chrX:136,872,346-136,875,257	Included exon	Yes	Repressed by RBMX	-	1007	Yes	High
REEP5	chr5:112,876,379-112,878,835	Alternative termination site	Yes	Activated by RBMX	No	2457	-	-
RESF1	chr12:31,980,878-31,985,957	Upstream termination site	Yes	Repressed by RBMX	Yes	5080	Yes	High
REV3L	chr6:111,372,596-111,377,700	Alternative 3' splice site	No	Repressed by RBMX	-	4162	Yes	High
REV3L	chr6:111,363,979-111,365,264	Included exon	No	Repressed by RBMX	-	99	-	-
RFX3	chr9:3,293,259-3,330,258	Included exon	Yes	Repressed by RBMX	-	75	-	-
RFX7	chr15:56,120,094-56,142,777	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	Yes	High
RGS5	chr1:163,244,505-163,306,232	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	-	-
RIF1	chr2:151,462,467-151,466,120	Alternative 3' splice site	Yes	Repressed by RBMX	-	3237	Yes	High
RNASE4	chr14:20,684,759-20,694,186	Upstream terminal exon	Yes	Repressed by RBMX	Yes	640	-	-
RNFT1	chr17:59,958,445-59,960,067	Included exon	Yes	Repressed by RBMX	-	78	-	-
RPAP3	chr12:47,699,469-47,701,463	Upstream terminal exon	No	Repressed by RBMX	Yes	696	-	-
RPL13P5	chr12:6,883,919-6,884,371	Intron retention	Yes	Repressed by RBMX	-	-	-	-
RPS6KB1	chr17:59,946,551-59,950,420	Alternative termination site	Yes	Activated by RBMX	Yes	3870	-	-
S100A4	chr1:153,544,654-153,545,063	Intron retention	Yes	Repressed by RBMX	-	-	-	-
SAMD12	chr8:118,189,468-118,197,732	Upstream termination site	Yes	Repressed by RBMX	No	8265	-	-
SAMD9L	chr7:93,130,056-93,135,991	Upstream termination site	No	Repressed by RBMX	Yes	5936	-	-
SAP130	chr2:127,955,345-127,986,784	Included exon	Yes	Repressed by RBMX	-	105	-	-
SAT1	chrX:23,783,884-23,785,327	Skipped exon	Yes	Activated by RBMX	-	110	Yes	High
SAT2	chr17:7,626,794-7,627,142	Skipped exon	Yes	Activated by RBMX	-	102	-	-
SAYSD1	chr6:39,105,777-39,114,882	Skipped exon	Yes	Activated by RBMX	-	4407	Yes	Low

Gene	Coordinates	Type		Regulation		Number		Level
SCAF11	chr12:45,926,142-45,928,859	Exitron	No	Repressed by RBMX	-	2718	Yes	High
SCLY	chr2:238,083,355-238,092,083	Upstream terminal exon	Yes	Repressed by RBMX	No	866	Yes	High
SCRIB	chr8:143,791,441-143,791,875	Skipped exon	Yes	Activated by RBMX	-	75	-	-
SEC16A	chr9:136,474,049-136,477,684	Exitron	Yes	Repressed by RBMX		3636	upstream the exitron	Low
SEC22C	chr3:42,563,687-42,568,864	Included exon	Yes	Repressed by RBMX	-	216	-	-
SERTAD4	chr1:210,241,558-210,246,863	Upstream termination site	Yes	Repressed by RBMX	Yes	5306	-	-
SGO1	chr3:20,171,233-20,175,055	Alternative 5' splice site	Yes	Repressed by RBMX	-	807	only	Low
SGO2	chr2:200,542,665-200,570,332	Upstream terminal exon	Yes	Repressed by RBMX	No	670	Yes but mainly in the exon	Low
SGO2	chr2:200,542,665-200,543,600	Cryptic upstream terminal exon	Yes	Repressed by RBMX	No	-	-	-
SHLD2	chr10:87,151,350-87,158,047	Alternative 5' splice site	Yes	Repressed by RBMX	-	1530	Yes	Medium
SHROOM2	chrX:9,786,456-9,932,174	Alternative first exon	Yes	Activated by RBMX	-	255	-	-
SIKE1	chr1:114,769,479-114,774,372	Alternative termination site	Yes	Activated by RBMX	No	4894	-	-
SIN3B	chr19:16,853,159-16,857,107	Upstream terminal exon	Yes	Repressed by RBMX	Yes	2965	-	-
SIX5	chr19:45,766,112-45,768,041	Skipped exon	Yes	Activated by RBMX	-	806	-	-
SLC12A2	chr5:128,114,682-128,117,468	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	Yes	Medium
SLC20A2	chr8:42,529,468-42,540,914	Upstream terminal exon	No	Repressed by RBMX	Yes	5868	mainly splice site	Low
SLC25A19	chr17:75,286,803-75,289,353	Included exon	Yes	Repressed by RBMX	-	90	-	-
SLC25A19	chr17:75,286,803-75,289,353	Included exon	Yes	Repressed by RBMX	-	74	-	-
SLC25A22	chr11:795,170-798,271	Alternative first exon	Yes	Repressed by RBMX	-	221	-	-
SLC25A26	chr3:66,221,128-66,243,202	Included exon	Yes	Repressed by RBMX	-	157	-	-
SLC25A44	chr1:156,194,248-156,199,834	Included exon	No	Repressed by RBMX	-	161	-	-
SLC29A1	chr6:44,219,615-44,226,060	Alternative first exon	Yes	Activated by RBMX	-	60	-	-
SLC2A13	chr12:40,090,020-40,105,252	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
SLC2A8	chr9:127,397,169-127,399,906	Alternative first exon	No	Activated by RBMX	-	207	-	-
SLC2A8	chr9:127,400,007-127,403,659	Skipped exon	Yes	Activated by RBMX	-	197	-	-
SLC35A3	chr1:100,015,421-100,022,385	Included exon	Yes	Repressed by RBMX	-	134	-	-
SLC35E2B	chr1:1,676,847-1,692,448	Included exon	Yes	Repressed by RBMX	-	148	-	-
SLC35G1	chr10:93,900,752-93,903,824	Upstream termination site	Yes	Repressed by RBMX	Yes	3073	-	-

Gene	Location	Type		RBMX				
SLC38A1	chr12:46,183,063-46,189,071	Upstream termination site	Yes	Repressed by RBMX	No	6009	Yes	Medium
SLC38A1	chr12:46,189,897-46,197,719	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
SLC38A6	chr14:60,984,804-60,999,944	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
SLC39A11	chr17:72,817,855-72,849,633	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
SLC39A9	chr14:69,424,203-69,429,457	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	Yes	Medium
SLC9A3R2	chr16:2,026,928-2,036,323	Alternative first exon	Yes	Activated by RBMX	-	291	-	-
SLFN12	chr17:35,409,465-35,411,927	Upstream termination site	Yes	Repressed by RBMX	Yes	2463	-	-
SP110	chr2:230,189,474-230,200,884	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
SP140	chr2:230,238,382-230,239,078	Upstream terminal exon	Yes	Repressed by RBMX	Yes	318	-	-
SPATA33	chr16:89,658,422-89,661,508	Upstream terminal exon	Yes	Repressed by RBMX	No	969	TW	Low
SPDL1	chr5:169,588,576-169,591,047	Included exon	Yes	Repressed by RBMX	-	77	-	-
SPIDR	chr8:47,294,031-47,316,812	Cryptic upstream terminal exon	No	Repressed by RBMX	No	-	Yes	High
SPIN3	chrX:56,990,831-56,995,215	Upstream terminal exon	Yes	Repressed by RBMX	No	4119	-	-
SPRED1	chr15:38,324,810-38,327,816	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
SPTAN1	chr9:128,591,626-128,594,174	Included exon	Yes	Repressed by RBMX	-	60	-	-
SPTLC3	chr20:13,009,385-13,056,256	Upstream terminal exon	No	Repressed by RBMX	Yes	7312	-	-
SRD5A1	chr5:6,652,009-6,656,077	Included exon	Yes	Repressed by RBMX	-	190	Yes but only 2 in introns around the	Low
SRD5A1	chr5:6,656,180-6,660,673	Cryptic upstream terminal exon	No	Repressed by RBMX	No	-	-	-
SRD5A1	chr5:6,668,202-6,674,386	Upstream termination site	Yes	Repressed by RBMX	Yes	6185	-	-
SRGAP2	chr1:206,453,381-206,456,161	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1282	-	-
SS18	chr18:26,035,128-26,038,554	Included exon	Yes	Repressed by RBMX	-	93	-	-
ST3GAL2	chr16:70,399,534-70,438,948	Cryptic exon	Yes	Repressed by RBMX	-	-	-	-
ST7	chr7:117,222,063-117,223,907	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1047	-	-
STAT6	chr12:57,108,300-57,111,128	Skipped exon	Yes	Activated by RBMX	-	170	-	-
STEAP3	chr2:119,223,889-119,239,362	Included exon	Yes	Repressed by RBMX	-	415	-	-
STK36	chr2:218,694,636-218,696,601	Alternative 3' splice site	Yes	Activated by RBMX	-	75	-	-

STK4	chr20:44,973,434-44,978,442	Cryptic exon	Yes	Repressed by RBMX	-	87	Yes	High
STK4	chr20:44,972,159-44,974,099	Cryptic exon	Yes	Repressed by RBMX	-	133	Yes	Low
STK4	chr20:44,973,434-44,975,684	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	1585	Yes	High
STX3	chr11:59,795,483-59,800,854	Included exon	Yes	Repressed by RBMX	-	114	-	-
SUPT5H	chr19:39,457,741-39,458,817	Skipped exon	Yes	Activated by RBMX	-	12	-	-
SWT1	chr1:185,168,399-185,172,426	Alternative terminal exon	No	Activated by RBMX	No	2627	-	-
SYTL2	chr11:85,709,501-85,714,412	Included exon	Yes	Repressed by RBMX	-	120	-	-
SZT2	chr1:43,404,551-43,408,760	Cryptic upstream terminal exon	No	Repressed by RBMX	No	-	Yes	Low
TAF1C	chr16:84,184,851-84,186,900	Alternative 3' splice site	Yes	Activated by RBMX	-	-	-	-
TANGO2	chr22:20,064,542-20,067,929	Upstream termination site	Yes	Repressed by RBMX	Yes	3388	-	-
TANGO2	chr22:20,063,338-20,064,541	Alternative 5' splice site	Yes	Repressed by RBMX	-	261	-	-
TAPT1	chr4:16,166,794-16,174,203	Skipped exon	Yes	Activated by RBMX	-	77	-	-
TAPT1	chr4:16,202,581-16,226,258	Tandem included exon	Yes	Repressed by RBMX	-	131	-	-
TAPT1	chr4:16,191,524-16,213,767	Tandem included exon	Yes	Repressed by RBMX	-	119	-	-
TARBP2	chr12:53,504,470-53,504,815	Alternative 3' splice site	Yes	Repressed by RBMX	-	204	-	-
TBC1D16	chr17:79,932,343-79,941,107	Upstream termination site	Yes	Repressed by RBMX	Yes	8765	Yes	Low
TBC1D24	chr16:2,496,034-2,497,113	Upstream termination site	No	Repressed by RBMX	No	1080	-	-
TBC1D31	chr8:123,101,008-123,104,201	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1955	-	-
TBRG1	chr11:124,626,610-124,626,903	Skipped exon	Yes	Activated by RBMX	-	59	-	-
TCAIM	chr3:44,361,519-44,363,037	Upstream terminal exon	Yes	Repressed by RBMX	Yes	918	-	-
TCF25	chr16:89,907,323-89,910,590	Included exon	Yes	Repressed by RBMX	-	142	Near the second 5'ss	Low
TCFL5	chr20:62,843,598-62,854,015	Upstream terminal exon	Yes	Repressed by RBMX	No	2520	Yes	High
TEP1	chr14:20,381,402-20,381,912	Skipped exon	Yes	Activated by RBMX	-	134	-	-
TFDP2	chr3:141,995,142-142,000,353	Alternative first exon	Yes	Repressed by RBMX	-	102	-	-
TFPI	chr2:187,520,619-187,544,503	Included exon	Yes	Repressed by RBMX	-	122	-	-
TGFB2	chr1:218,441,204-218,444,619	Upstream termination site	Yes	Repressed by RBMX	No	3416	-	-
TGFB2	chr1:218,347,048-218,360,660	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
THAP5	chr7:108,565,106-108,566,022	Alternative 5' splice site	Yes	Repressed by RBMX	-	193	-	-
THG1L	chr5:157,739,321-157,741,449	Upstream termination site	Yes	Repressed by RBMX	No	2129	Yes	Low
TIRAP	chr11:126,283,154-126,290,461	Included exon	Yes	Repressed by RBMX	-	169	-	-

TIRAP	chr11:126,289,665-126,290,461	Alternative 5' splice site	Yes	Activated by RBMX	-	169	-	-
TLR6	chr4:38,829,538-38,856,760	Included exon	Yes	Repressed by RBMX	-	153	-	-
TMEM128	chr4:4,235,542-4,237,826	Alternative 3' splice site	Yes	Activated by RBMX	-	1536	-	-
TMEM184C	chr4:147,634,169-147,637,646	Upstream termination site	No	Repressed by RBMX	Yes	3478	Yes	High
TMEM241	chr18:23,373,764-23,377,645	Included exon	Yes	Repressed by RBMX	-	115	-	-
TMUB2	chr17:44,188,010-44,189,588	Alternative 3' splice site	Yes	Repressed by RBMX	-	654	-	-
TMUB2	chr17:44,187,744-44,188,934	Included exon	Yes	Repressed by RBMX	-	103	-	-
TNC	chr9:115,063,190-115,064,646	Tandem included exon	Yes	Repressed by RBMX	-	273	-	-
TNC	chr9:115,060,003-115,063,795	Tandem included exon	Yes	Repressed by RBMX	-	273	-	-
TNC	chr9:115,042,342-115,048,259	Tandem included exon	Yes	Repressed by RBMX	-	273	-	-
TNC	chr9:115,064,069-115,073,602	Tandem included exon	Yes	Repressed by RBMX	-	273	-	-
TNC	chr9:115,057,426-115,062,916	Tandem included exon	Yes	Repressed by RBMX	-	273	-	-
TNC	chr9:115,046,683-115,052,739	Tandem included exon	Yes	Repressed by RBMX	-	273	-	-
TNC	chr9:115,048,533-115,057,152	Tandem included exon	Yes	Repressed by RBMX	-	276	-	-
TNC	chr9:115,053,016-115,059,729	Tandem included exon	Yes	Repressed by RBMX	-	273	-	-
TNFRSF11B	chr8:118,924,016-118,924,762	Upstream termination site	No	Repressed by RBMX	No	747	-	-
TNFSF12	chr17:7,550,979-7,556,777	Included exon	Yes	Repressed by RBMX	-	239	-	-
TNRC6B	chr22:40,322,854-40,335,807	Upstream termination site	Yes	Repressed by RBMX	No	12954	Yes	Low
TOM1L2	chr17:17,911,624-17,972,261	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	-	-
TRA2A	chr7:23,521,841-23,531,788	Skipped exon	Yes	Activated by RBMX	-	301	-	-
TRAK1	chr3:42,202,753-42,212,615	Upstream terminal exon	Yes	Repressed by RBMX	No	2849	-	-
TRAPPC10	chr21:44,059,215-44,060,262	Upstream terminal exon	Yes	Repressed by RBMX	Yes	798	-	-
TRAPPC12	chr2:3,443,892-3,453,934	Cryptic upstream terminal exon	Yes	Repressed by RBMX	No	-	mostly splice site	Low
TRAPPC12	chr2:3,460,337-3,462,533	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	Yes but not TW	Low
TRAPPC13	chr5:65,625,107-65,630,891	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1250	-	-
TRAPPC2	chrX:13,719,983-13,734,524	Skipped exon	Yes	Activated by RBMX	-	142	-	-
TRIQK	chr8:92,917,011-92,954,405	Included exon	Yes	Repressed by RBMX	-	158	-	-
TUG1	chr22:30,975,170-30,979,395	Upstream termination site	No	Repressed by RBMX	Yes	4226	Yes	High
U2AF1L4	chr19:35,745,125-35,745,347	Alternative 3' splice site	Yes	Activated by RBMX	-	88	-	-
U2AF1L4	chr19:35,744,739-35,745,347	Skipped exon	Yes	Activated by RBMX	-	88	-	-
UBE2A	chrX:119,575,401-119,582,587	Included exon	Yes	Repressed by RBMX	-	90	-	-

UBE2D4	chr7:43,952,650-43,956,136	Upstream termination site	Yes	Repressed by RBMX	Yes	3487	-	-
UBE2J2	chr1:1,263,387-1,267,861	Included exon	Yes	Repressed by RBMX	-	193	-	-
UBXN11	chr1:26,298,063-26,301,693	Included exon	Yes	Repressed by RBMX	-	99	-	-
UHRF1BP1L	chr12:100,058,029-100,059,533	Upstream termination site	No	Repressed by RBMX	Yes	1505	Yes	Low
ULK3	chr15:74,840,234-74,840,641	Intron retention	Yes	Repressed by RBMX	-	-	-	-
UMAD1	chr7:7,676,209-7,727,891	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	-	-
UNK	chr17:75,792,479-75,794,198	Upstream terminal exon	Yes	Repressed by RBMX	No	756	-	-
UPP1	chr7:48,094,828-48,106,872	Tandem included exon	Yes	Repressed by RBMX	-	118	-	-
UPP1	chr7:48,094,828-48,106,872	Tandem included exon	Yes	Repressed by RBMX	-	159	-	-
UPP1	chr7:48,094,828-48,106,872	Tandem included exon	Yes	Repressed by RBMX	-	115	-	-
USB1	chr16:58,010,113-58,013,558	Upstream terminal exon	Yes	Repressed by RBMX	Yes	2541	-	-
USF3	chr3:113,648,385-113,661,425	Upstream termination site	No	Repressed by RBMX	No	13041	-	-
VPS26C	chr21:37,261,082-37,267,237	Cryptic upstream terminal exon	Yes	Repressed by RBMX	No	-	Maybe at branchpoint	Low
VPS26C	chr21:37,230,518-37,232,376	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	-	-
VTI1A	chr10:112,447,468-112,450,726	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	-	-
WDR19	chr4:39,272,980-39,274,958	Alternative 3' splice site	Yes	Repressed by RBMX	-	-	Yes	Low
WDR90	chr16:662,673-665,801	Intron retention	No	Repressed by RBMX	-	-	Yes	Low
WHAMM	chr15:82,819,489-82,823,099	Included exon	Yes	Repressed by RBMX	-	129	-	-
XRN1	chr3:142,324,275-142,329,433	Cryptic upstream terminal exon	No	Repressed by RBMX	Yes	-	that whole intron	High
XYLB	chr3:38,412,936-38,417,722	Upstream termination site	No	Repressed by XBMX	Yes	4787	Yes	High
ZBTB10	chr8:80,519,223-80,526,264	Upstream termination site	Yes	Repressed by RBMX	Yes	7042	-	-
ZBTB21	chr21:41,986,831-41,994,108	Exitron	Yes	Repressed by RBMX	-	7278	Yes	High
ZBTB21	chr21:41,994,109-42,002,896	Cryptic exon	Yes	Repressed by RBMX	-	-	-	-
ZBTB44	chr11:130,226,677-130,234,273	Upstream termination site	No	Repressed by RBMX	Yes	7597	-	-
ZC3HAV1	chr7:139,060,338-139,064,878	Upstream terminal exon	Yes	Repressed by RBMX	No	799	-	-
ZDBF2	chr2:206,304,717-206,314,427	Upstream termination site	Yes	Repressed by RBMX	Yes	9711	Yes	High
ZDHHC3	chr3:44,925,171-44,929,305	Upstream terminal exon	Yes	Repressed by RBMX	Yes	1677	-	-
ZDHHC3	chr3:44,929,437-44,933,117	Included exon	Yes	Repressed by RBMX	-	84	-	-
ZFP64	chr20:52,151,278-52,160,122	Upstream terminal exon	Yes	Repressed by RBMX	Yes	2451	-	-

ZFP82	chr19:36,383,120-36,394,110	Upstream termination site	Yes	Repressed by RBMX	Yes	5481	-	-
ZFYVE28	chr4:2,327,861-2,335,704	Upstream terminal exon	No	Repressed by RBMX	No	1282	Yes	Low
ZMYM1	chr1:35,112,977-35,116,458	Upstream termination site	No	Repressed by RBMX	Yes	3482	Yes	High
ZMYM1	chr1:35,112,977-35,116,458	Upstream termination site	No	Repressed by RBMX	Yes	3482	Yes	High
ZMYM6	chr1:34,986,980-34,988,935	Upstream termination site	No	Repressed by RBMX	Yes	1956	Yes	High
ZMYM6	chr1:35,017,250-35,020,382	Upstream terminal exon	Yes	Repressed by RBMX	No	3133	-	-
ZNF107	chr7:64,679,337-64,687,828	Upstream terminal exon	Yes	Repressed by RBMX	Yes	3494	Yes	Low
ZNF12	chr7:6,703,065-6,705,158	Cryptic upstream terminal exon	No	Repressed by RBMX	No	-	-	-
ZNF121	chr19:9,560,373-9,567,109	Upstream termination site	Yes	Repressed by RBMX	Yes	6737	Yes	High
ZNF133	chr20:18,298,063-18,305,680	Tandem included exon	Yes	Repressed by RBMX	-	176	Yes but near splice sites and across	Medium
ZNF133	chr20:18,298,063-18,305,680	Tandem included exon	Yes	Repressed by RBMX	-	171	Yes but near splice sites and across	Medium
ZNF138	chr7:64,831,451-64,833,664	Upstream termination site	Yes	Repressed by RBMX	Yes	2214	Yes	High
ZNF142	chr2:218,650,527-218,656,149	Included exon	Yes	Repressed by RBMX	-	600	-	-
ZNF195	chr11:3,371,704-3,379,037	Included exon	Yes	Repressed by RBMX	-	80	-	-
ZNF197	chr3:44,641,900-44,648,471	Upstream termination site	No	Repressed by RBMX	Yes	6572	Yes	Medium
ZNF211	chr19:57,634,756-57,640,703	Included exon	Yes	Repressed by RBMX	-	156	-	-
ZNF226	chr19:44,175,498-44,177,685	Upstream termination site	Yes	Repressed by RBMX	Yes	2188	Yes	High
ZNF226	chr19:44,172,953-44,175,429	Upstream terminal exon	Yes	Repressed by RBMX	Yes	446	-	-
ZNF248	chr10:37,829,035-37,833,116	Exitron	Yes	Repressed by RBMX	-	4082	Yes	High
ZNF260	chr19:36,515,700-36,528,218	Included exon	No	Repressed by RBMX	-	219	-	-
ZNF268	chr12:133,202,144-133,211,184	Upstream termination site	Yes	Repressed by RBMX	Yes	9041	-	-
ZNF273	chr7:64,918,293-64,924,431	Upstream terminal exon	Yes	Repressed by RBMX	No	1046	-	-
ZNF292	chr6:87,254,650-87,264,196	Upstream termination site	No	Repressed by RBMX	Yes	9547	Yes	High
ZNF322	chr6:26,634,393-26,638,728	Upstream termination site	Yes	Repressed by RBMX	No	4336	-	-
ZNF33A	chr10:38,054,375-38,060,063	Upstream termination site	Yes	Repressed by RBMX	Yes	5689	-	-
ZNF347	chr19:53,148,681-53,149,367	Intron retention	No	Activated by RBMX	-	-	-	-
ZNF347	chr19:53,134,935-53,142,556	Upstream termination site	No	Repressed by RBMX	No	7622	-	-
ZNF37A	chr10:38,117,390-38,124,621	Upstream termination site	No	Repressed by RBMX	No	7232	Yes	High
ZNF37BP	chr10:42,513,510-42,521,127	Upstream termination site	Yes	Repressed by RBMX	No	7618	Yes	High

ZNF451	chr6:57,099,142-57,109,716	Upstream terminal exon	Yes	Repressed by RBMX	Yes	9114	Yes	High
ZNF480	chr19:52,321,579-52,325,922	Upstream termination site	No	Repressed by RBMX	Yes	4344	Yes	High
ZNF518A	chr10:96,156,197-96,165,512	Upstream termination site	Yes	Repressed by RBMX	No	9316	Yes	High
ZNF557	chr19:7,082,878-7,087,968	Upstream termination site	No	Repressed by RBMX	No	5091	Yes	Low
ZNF615	chr19:52,003,901-52,008,140	Skipped exon	Yes	Activated by RBMX	-	38	-	-
ZNF638	chr2:71,331,866-71,355,718	Skipped exon	Yes	Activated by RBMX	-	1519	Yes	High
ZNF641	chr12:48,337,180-48,343,727	Upstream termination site	Yes	Repressed by RBMX	Yes	6548	-	-
ZNF641	chr12:48,345,475-48,348,115	Alternative first exon	No	Activated by RBMX	-	209	-	-
ZNF655	chr7:99,560,696-99,563,862	Skipped exon	Yes	Activated by RBMX	-	86	-	-
ZNF655	chr7:99,562,495-99,564,643	Upstream terminal exon	Yes	Repressed by RBMX	No	781	-	-
ZNF655	chr7:99,560,696-99,563,862	Included exon	Yes	Repressed by RBMX	-	127	-	-
ZNF655	chr7:99,558,595-99,560,532	Alternative first exon	Yes	Activated by RBMX	-	156	-	-
ZNF69	chr19:11,904,649-11,914,556	Upstream termination site	Yes	Repressed by RBMX	No	9908	-	-
ZNF701	chr19:52,582,202-52,587,174	Upstream termination site	Yes	Repressed by RBMX	No	4973	Yes	Low
ZNF720	chr16:31,753,901-31,759,886	Alternative terminal exon	Yes	Activated by RBMX	Yes	5432	-	-
ZNF738	chr19:21,375,965-21,379,066	Upstream terminal exon	Yes	Repressed by RBMX	No	438	-	-
ZNF763	chr19:11,978,116-11,980,497	Upstream termination site	No	Repressed by RBMX	No	2382	-	-
ZNF780A	chr19:40,069,152-40,076,209	Upstream termination site	Yes	Repressed by RBMX	No	7058	Yes	High
ZNF780B	chr19:40,031,507-40,036,626	Upstream termination site	No	Repressed by RBMX	No	5120	Yes	High
ZNF782	chr9:96,816,472-96,819,778	Upstream termination site	Yes	Repressed by RBMX	No	3307	Yes	High
ZNF782	chr9:96,852,006-96,854,087	Included exon	Yes	Repressed by RBMX	-	217	-	-
ZNF800	chr7:127,373,342-127,377,185	Alternative 3' splice site	No	Repressed by RBMX	-	1693	Yes	High
ZNF84	chr12:133,037,503-133,041,277	Alternative first exon	Yes	Activated by RBMX	-	52	-	-
ZNF84	chr12:133,048,849-133,063,304	Cryptic downstream terminal exon	Yes	Repressed by RBMX	Yes	-	Yes but at the repressed last exon	Low
ZNF85	chr19:20,948,744-20,950,697	Upstream termination site	Yes	Repressed by RBMX	No	1954	-	-
ZNF860	chr3:31,981,903-31,988,659	Cryptic exon	Yes	Repressed by RBMX	-	-	-	-
ZNF888	chr19:52,904,417-52,908,179	Upstream termination site	No	Repressed by RBMX	No	3763	-	-
ZRANB3	chr2:135,302,972-135,315,358	Upstream terminal exon	No	Repressed by RBMX	No	10634	Yes	Low
ZRANB3	chr2:135,345,136-135,345,549	Cryptic upstream terminal exon	Yes	Repressed by RBMX	Yes	-	-	-
ZSCAN2	chr15:84,604,334-84,616,611	Upstream terminal exon	Yes	Repressed by RBMX	Yes	303	-	-

| ZSCAN26 | chr6:28,276,195-28,278,201 | Upstream termination site | No | Repressed by RBMX | Yes | 2007 | Yes | Medium |
| ZSCAN26 | chr6:28,271,854-28,272,669 | Alternative 5' splice site | Yes | Repressed by RBMX | - | 486 | - | - |

Taylor & Francis
Taylor & Francis Group

REVIEW

OPEN ACCESS Check for updates

Cryptic splicing: common pathological mechanisms involved in male infertility and neuronal diseases

Saad Aldalaqan, Caroline Dalgliesh, Sara Luzzi, Chileleko Siachisumo, Louise N Reynard, Ingrid Ehrmann, and David J. Elliott

Newcastle University Bioscience Institute, Newcastle University, Central Parkway Newcastle, UK

ABSTRACT
High levels of transcription and alternative splicing are recognized hallmarks of gene expression in the testis and largely driven by cells in meiosis. Because of this, the male meiosis stage of the cell cycle is often viewed as having a relatively permissive environment for gene expression. In this review, we highlight recent findings that identify the RNA binding protein RBMXL2 as essential for male meiosis. RBMXL2 functions as a "guardian of the transcriptome" that protects against the use of aberrant (or "cryptic") splice sites that would disrupt gene expression. This newly discovered protective role during meiosis links with a wider field investigating mechanisms of cryptic splicing control that protect neurons from amyotrophic lateral sclerosis and Alzheimer's disease. We discuss how the mechanism repressing cryptic splicing patterns during meiosis evolved, and why it may be essential for sperm production and male fertility.

Pre-mRNA RNA splicing is a crucial mechanism in eukaryotes and is required to enable expression of protein-coding RNAs (mRNAs) from most mammalian genes. Splicing joins together exons within nascent RNA transcripts, thus creating open reading frames from split genes. Splicing is carried out by a molecular machine called the spliceosome [1]. For accurate pre-mRNA splicing the spliceosome needs to precisely identify short consensus sequences called splice sites at exon-intron junctions and join these together. Because of their short length, sequences similar to splice sites (but not selected by the spliceosome) can occur somewhat frequently within genes. Such infrequently used splice sites have the potential to be selected by the spliceosome but are generally not used, so are referred to as "cryptic" in this review (Figure 1). Cryptic splice site sequences are only weakly recognized by the spliceosome and may be located within repetitive sequences and repressed by nuclear RNA binding proteins [2–4]. However, cryptic splice sites can become activated under certain conditions, including some neurological diseases, and their selection can disrupt production of full-length proteins [5].

Some nuclear RNA binding proteins play key roles in repressing the selection of cryptic splicing patterns within the nervous system. These include TDP43 protein that represses cryptic splicing patterns in neurons but becomes disrupted in amyotrophic lateral sclerosis (ALS) leading to the death of motor neurons [6–9]. Cryptic exons are also included in the hippocampus of patients with Alzheimer's disease [10]. Through its role in cryptic splicing repression TDP43 has been identified as a "guardian of the transcriptome" that is essential for neuron survival [7]. Whether repression of cryptic splicing is important outside of the nervous system has been less well understood. Here, we highlight recent research that reveal a male germ cell-specific nuclear RNA binding protein that operates as a newly discovered guardian of the transcriptome during meiosis.

The testis is considered a relatively permissive site for gene expression patterns. Most human genes produce multiple different mRNAs by using alternative splice sites or by using different combinations of exons. Such alternative splicing permits single genes to produce multiple mRNA isoforms to help amplify the information embedded in the genome.
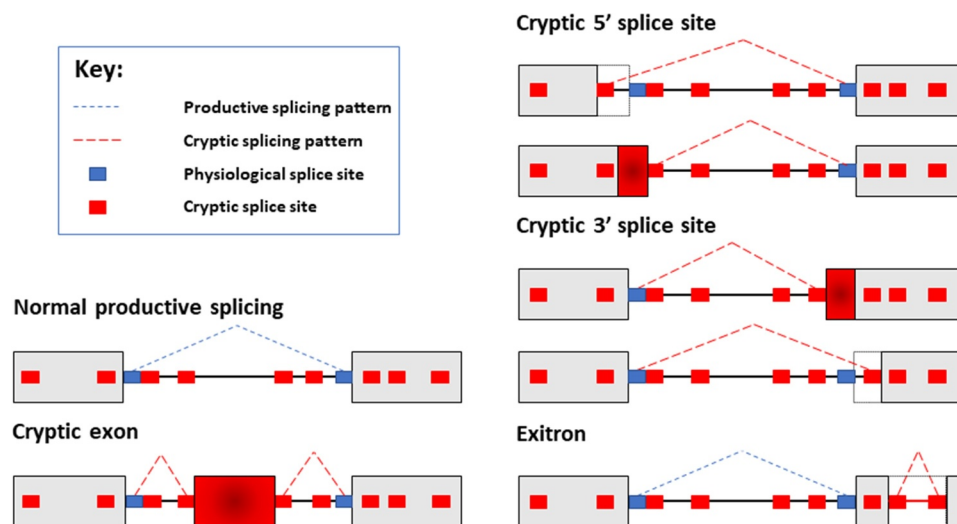
Figure 1. Schematic diagram of cryptic splicing patterns. Most genes are split between exons (shown as gray boxes here) and introns (shown as connecting lines between the boxes). Normal patterns of splice site selection will involve the spliceosome recognizing bona fide splice sites, and joining exons together to create mRNAs. In this example, normal productive splicing is indicated with dashed blue lines. Cryptic splice sites (smaller red boxes) resemble physiological splice sites (smaller blue boxes), and are found within both introns and exons. While normally these cryptic splice sites are ignored by the spliceosome, potentially they could act as decoy sites splice sites for spliceosome selection. Use of cryptic splice sites would produce different mRNAs from genes. Here the normal splicing patterns is shown as a broken blue line joining the physiological splice sites. Examples of cryptic splicing are indicated with dashed red lines. These cryptic splicing events are inclusion of a cryptic exon embedded deep within an intron; selection of cryptic 5′ and 3′ splice sites; and aberrant recognition of cryptic splice sites within an exon, leading to the interior of this exon being aberrantly recognized as an intron (in a cryptic splicing event known as an exitron).

Particularly high levels of alternative splicing have been detected in the testis and in the brain compared to other tissues [11–13]. Alternative splicing patterns can evolve rapidly between species and early analyses detected higher levels of evolutionary divergent splicing in the testis compared to other tissues. This includes the brain, where alternative mRNA isoforms were more likely to be frequently conserved between species than alternative splice isoforms in the testis [11]. More recent comparative transcriptomic analyses confirm some newly evolved exons are exclusively expressed within the testis but suggest there may also be broadly similar levels of conserved mRNA splice isoforms in the testes compared to other tissues [14–16]. The more recently evolved splicing events within the testis are less likely to play a fundamental biological role than more evolutionarily ancient alternative splicing events that have been maintained under selective pressure. However, some recently evolved exons within the testis might later evolve into more generally useful mRNA isoforms via an evolutionary model, which is called the "testis-first" hypothesis. This hypothesis suggests that splicing permissiveness in the testis enables genes to "try out" new exon combinations before they can later be placed under selective pressure [17]. As well as high levels of alternative splicing, there are also particularly high levels of transcription within the testis compared with most other tissues – both in amounts of RNA produced and numbers of genes transcribed [12,18,19].

The human testis produces between 45 and 207 million sperm a day, making it one of the most active developmental pathways still operating in adults [20,21]. The testis contains populations of germ cells (in the developmental pathway leading to sperm) and somatic cells (including Sertoli cells that support germ cell development and Leydig cells that produce testosterone). A population of mitotically active cells called spermatogonia that are early in the germ cell developmental pathway differentiate into cells called spermatocytes. Spermatocytes undergo meiosis, a special form of cell division that produces haploid daughter cells *via* two sequential divisions. The first meiotic division is preceded by a long prophase that lasts around 2 weeks in mice, referred to as meiotic prophase I. This is divided up into five sequential sub-stages called leptotene, zygotene, pachytene, diplotene and diakinesis – all characterized by distinct chromosomal behaviors.

During meiotic prophase I chromosomes condense, and non-sister chromatids form crossovers and undergo genetic recombination. Subsequently, cells separate sister chromatids through a second cell division called meiosis II. This produces haploid spermatids that after meiosis differentiate into spermatozoa (Figure 2).

The cell types that are responsible for the high levels of splicing and gene transcription in the testis have been identified as spermatocytes [12,16,22,23]. Recent transcriptomic analyses of purified testicular cell types reveal that alternative splicing and gene expression levels peak during mid to late pachytene and diplotene stages of meiosis (Figure 2) [24]. In contrast, leptotene, zygotene and early pachytene are transcriptionally quiescent [24–27]. Further RNA sequencing analyses of purified mouse germ cell types detected extensive transcription of both genes and intergenic regions during pachytene and diplotene and in round spermatids, pinpointing these particular cell types as being major contributors to the high levels of testis gene expression and transcriptome complexity [28]. The more "permissive" gene expression environment in the testis may perhaps occur as a result of relaxed chromatin folding. High levels of autosomal transcription during pachytene and diplotene are driven by patterns of open chromatin, including increased levels of the epigenetic mark H3K4me2 (a marker of active promoters)

and decreased CpG methylation (a modification normally associated with patterns of gene repression) [12,29]. Furthermore, bursts of meiotic gene expression are driven by activation of super enhancers bound by the MYBL1 and SCML2 transcription factors [30].

A permissive gene environment during meiosis would be consistent with some relaxation of splicing fidelity being tolerated. Despite this, recent data suggest that the interesting parallels between gene expression programs in the brain and testis [13] also extend to a requirement to repress cryptic splicing patterns that would cause cell death. Humans and mice (and likely all placental mammals) express a testis-specific RNA binding protein called RBMXL2 (also known as heterogeneous nuclear ribonucleoproteins G-testis or hnRNP-GT) [31,32]. Mutations have been detected within infertile men for the human *RBMXL2* gene [33]. Mouse RBMXL2 protein is expressed in pachytene and diplotene spermatocytes, the stages of meiosis that have the highest levels of transcription and alternative splicing (Figure 2) [34]. Genetic deletion of the mouse *Rbmxl2* gene causes cell death during meiotic diplotene, thereby reducing testis size and preventing sperm production [34]. Detailed molecular analysis of this mouse model show that RBMXL2 protein prevents the spliceosome selecting cryptic splice sites during the pachytene and diplotene stages of
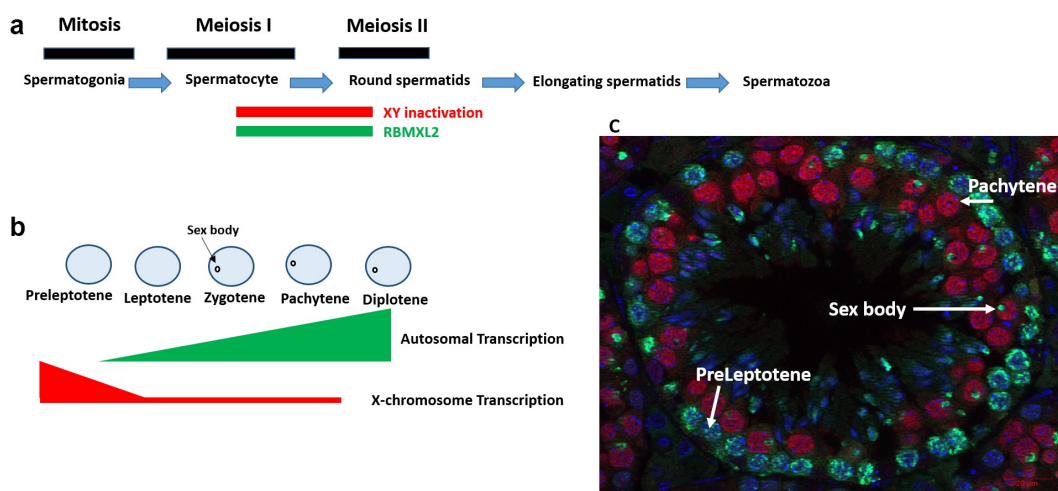
Figure 2. RBMXL2 is expressed during diplotene and pachytene of male meiotic prophase. a. Mouse germ cell development, showing the expression window of RBMXL2 and time period of XY inactivation. b. Transcription patterns of the X chromosome and autosomes during meiotic prophase. c. Seminiferous tubule counterstained with antibodies specific to RBMXL2 protein (pseudocoloured red, detected within nuclei of cells in pachytene in this tubule) and γH2AX (green color, detected within nuclei of preleptotene cells and within the sex bodies of pachytene cells).

male meiosis [34] – thus performing a similar molecular role to TDP43 in neurons. Both RBMXL2 and TDP43 are members of a group of proteins called hnRNPs (heterogeneous nuclear ribonucleoproteins) that bind to nuclear RNAs as they are transcribed to control their splicing patterns. Although they both operate as "guardians of transcriptomes", RBMXL2 protein is only expressed in spermatocytes and spermatids. In contrast, TDP43 protein is expressed more ubiquitously. Despite this, point mutations affecting TDP43 specifically cause neuronal cell death [35]. Interestingly, while complete genetic knockout of TDP43 causes embryonic death in mice [36,37], conditional genetic knockout of TDP43 in the testis causes male infertility [38].

Genes encoding mRNAs that are incorrectly spliced in the absence of RBMXL2 are enriched in functions associated with meiosis, chromosome segregation and spermatogenesis (Sara Luzzi unpublished). The inappropriate selection of cryptic splice sites in spermatocytes in the absence of RBMXL2 protein might therefore cause spermatocyte cell death by preventing proper expression of key genes needed for meiosis. RBMXL2 regulates splicing patterns of over a hundred genes during meiotic prophase. Important genes that contain cryptic splice sites that are repressed by RBMXL2 protein include *Brca2* (encoding a DNA repair protein involved in genetic recombination) and *Meioc* (which encodes a cytoplasmic protein that is critical during meiotic prophase) [39–41]. Exactly which RBMXL2 target genes are most important for spermatocyte survival is unknown. Complicating this prediction, the phenotypes caused by splicing errors within a narrow window of meiosis might be different from traditional genetic knockouts that assess when a gene is first needed in a developmental pathway. For example, *Meioc* genetic knockout spermatocytes have an unusually short meiotic prophase and do not reach pachytene or diplotene – the developmental window in which RBMXL2 protein is expressed (Figure 2). Similarly, genetic knockout of the *Brca2* gene prevents germ cell development at an early stage before germ cells enter meiosis [41]. The effects of changing *Meioc* and *Brca2* RNA processing pathways during the narrow window of meiotic prophase when RBMXL2 is normally expressed are not well understood and difficult to

predict. An alternative cause of spermatocyte cell death in the absence of RBMXL2 may be through genotoxic damage via formation of RNA:DNA hybrids (R-loops). The loss of splicing factors can cause normally intronic regions to be included within incorrectly spliced mRNAs (rather than being removed by splicing). R-loops form as a result of transcription involving local melting of DNA close to the elongating RNA polymerase, and intronic regions remaining within the pre-mRNA being able to base pair with the melted DNA duplex (forming R loops), leading to DNA damage [42]. In principle it should be possible to potentially correct aberrant splicing patterns caused by loss of RBMXL2 during meiosis. However, while individual cryptic splice sites can be therapeutically targeted using antisense oligonucleotides [43], it would be difficult to use this approach to correct the hundreds of targets that are normally controlled by RBMXL2 during meiosis.

Understanding RBMXL2 function fits into a bigger picture involving the evolution of new genes and regulation of gene expression patterns in the body. The *RBMXL2* gene originated via retro-transposition of an mRNA encoded by the X-linked gene *RBMX* approximately 65 million years ago. As a result of this, the *RBMXL2* gene does not contain introns [32]. RBMX is an RNA binding protein important for controlling splicing, transcription and genome stability [44,45]. Unpaired regions of the X and Y chromosomes (including also *RBMX* gene) become transcriptionally silent during pachytene, within a heterochromatic structure called sex body (or XY body) (Figure 2). This process is termed meiotic sex chromosome inactivation (MSCI) [46], and leads to either reduced or complete loss of sex-linked gene expression for the reminder of meiosis (over ~9 days in mice and even longer in humans [47]). Most retrogenes decay rapidly after they are formed. However, many essential X-linked genes that are transcriptionally silenced by MSCI are functionally replaced by retrogenes that are only expressed in the testis [48].

RBMXL2 is only expressed during and immediately after meiosis, so how are the genes controlled by RBMXL2 in the testis normally spliced in other tissues that do not express RBMXL2? A possible

answer is that the splice events that are controlled by RBMXL2 during meiosis might be controlled instead by RBMX in other cell types within the body. In this scenario *RBMXL2* may functionally replace *RBMX* function during meiotic prophase, either as a direct "like for like" replacement or as a more specialized replacement that has evolved to control specific gene expression pathways needed for meiosis [49–51]. However, whether RBMX and RBMXL2 proteins have similar functional activity is not yet fully answered. RBMXL2 mainly operates as a splicing repressor during meiotic prophase in mice. In contrast, global data from human cells has characterized the properties of RBMX protein mainly as a splicing activator that binds to specifically methylated pre-mRNA to slow progression of RNA polymerase II transcription thus facilitating spliceosome function and activating exon inclusion [52]. RBMX is also mutated in the X-linked intellectual disability syndrome Shashi syndrome, where it leads to increased p53 activity and neuronal defects via splicing activation of *MDM4* exon 6 [53].

Why is it important to repress cryptic splice sites during meiosis? Conventional splice sites have evolved to enable exons to be precisely joined together and maintain protein-coding open reading frames. Since cryptic splice sites are usually not used they are not under the same selective pressure as *bona fide* splice sites. Cryptic splice site inclusion into mRNAs often disrupts protein-coding reading frames by introducing premature termination codons (PTCs). In cells that are not dividing by meiosis, PTC-containing mRNAs are degraded by an RNA stability pathway called Nonsense Mediated Decay (NMD) that prevents translation into truncated proteins that could be harmful to the cell. However, uniquely in the testes PTC-containing transcripts can become stabilized. This stabilization occurs during meiosis, thus increasing the likelihood of mRNAs originating from cryptic splicing being translated into potentially toxic proteins. The reason for this stabilization is because of meiotic-associated changes in the NMD pathway. One of the core protein components of the NMD pathway is a protein called UPF3B that is encoded by a gene on the X chromosome that is turned off during meiosis by MCSI. As a consequence its autosomal

paralogue gene *UPF3A* becomes active when germ cells enter pachytene [54]. Genetic deletion of *UPF3A* induces meiotic defects in a mouse model showing *UPF3A* expression is critical for meiosis [54]. However, UPF3A protein has only weak activity in the NMD pathway because of an amino acid substitution compared to UPF3B. In fact, while UPF3B promotes mRNA degradation via NMD, UPF3A may operate as an NMD repressor [54]. Hence meiotic expression of UPF3A may lead to translation of some PTC-containing mRNAs and represent a possible reason why it is particularly important to repress cryptic splicing events during meiosis.

Other RNA binding proteins are also essential for splicing control during meiosis and have been recently reviewed [55–57]. However, the molecular defects that appear during mouse meiotic prophase in the absence of RBMXL2 protein, involving a high frequency of aberrantly selected cryptic splicing events, are largely distinct from those that have been identified for other splicing regulators during meiosis. One example is the splicing regulator PTBP2 that is expressed at high levels in spermatocytes [58]. Although PTBP2 protein represses cryptic splice sites in other cell types, its role during germ cell development seems more consistent as a master regulator of developmentally regulated splicing [58,59]. Genetic knockout of *Ptbp2* causes male germ cells to be prematurely sloughed off into the lumen of seminiferous tubules and defects to accumulate in the Sertoli cells cytoskeleton, suggesting impaired interactions between somatic and germ cells without PTBP2 protein [58]. This testicular phenotype correlates with disrupted splicing patterns detected for ~200 genes normally controlled by PTBP2, mainly with roles in Sertoli-germ cell communication. In contrast, more than 60% of the splicing events controlled by RBMXL2 during meiosis involve repression of cryptic splice sites rather than regulation of already known alternative splice events [34].

How does cryptic splicing repression by RBMXL2 integrate with other recently discovered aspects of splicing control during meiosis? High-throughput RNA sequencing analysis of purified meiotic spermatocytes and spermatids show that 10% of the alternative splicing events during meiosis occur via intron retention [23,24,60]. These intron-retained

mRNAs play a key role in developmental gene expression. Stable mRNAs containing retained introns are transcribed in spermatocytes and remain nuclear for a few days before being spliced and translated in the post-meiotic stages of spermatogenesis to encode crucial proteins in sperm development [23]. Mechanistically, intron retention involves the repression of splice sites, leading to whole introns being retained within mRNAs. The retained introns detected in spermatocytes and round spermatids have weak splice sites, suggesting a model where the extremely high levels of transcription during meiosis may overload the splicing machinery thus leading to intron retention. Despite RBMXL2's established role in repressing the selection of cryptic splice sites during meiosis, global changes in intron retention were not detected in the *Rbmxl2* knockout mouse model [34]. However, slower patterns of intron removal during meiosis may make some pre-mRNAs more vulnerable to cryptic exons being mistakenly selected by the spliceosome in mice that do not express RBMXL2 protein.

A key question for the future is why do spermatocytes die without RBMXL2 protein, and to what extent does this resemble neuronal cell death in cells depleted for TDP43 activity? Is cell death caused by the loss of specific important proteins as a result of cryptic splicing events in protein coding mRNAs or does cell death result from genotoxic damage caused by accumulation of R-loops from incorrectly spliced mRNAs (Figure 3)? What is the mechanism by which RBMXL2 represses cryptic splicing patterns? Does RBMXL2 bind to sequences in pre-mRNAs near to cryptic splice sites to sterically occlude the spliceosome? Or does RBMXL2 protein bind and antagonize the function of splicing activator proteins [61] preventing them from activating selection of cryptic splice sites that are otherwise poised for selection by the spliceosome? Could RBMXL2 repress cryptic exons by stabilizing formation of stalled spliceosomes on nascent RNA [62,63] (Figure 3)? Does RBMXL2 only repress cryptic splicing patterns or is it also involved in other aspects of meiosis (Figure 3)? Finally, do RBMX and RBMXL2 perform similar functions in cryptic splicing repression? This last question is of wide importance: RBMX protein

Figure 3. Mechanistic models to explain the impact of cryptic splicing on meiosis, and possible additional roles of RBMXL2. a. RBMXL2 could repress cryptic splicing by either counteracting the function of splicing activators; sterically blocking access of the spliceosome to cryptic splice sites; or promoting stalling of spliceosome assembly at cryptic splice sites. In the absence of RBMXL2 protein cryptic splicing could hinder production of proteins important for meiosis (Model 1) or alternatively impair transcription elongation and promote formation of R-loops (Model 2). Both scenarios would lead to meiotic arrest. b. Both RBMX and RBMXL2 are known to regulate splicing in somatic and germ cells respectively. Could RBMXL2 also have a function in other pathways known to be regulated by RBMX such as RNA polymerase II transcription and DNA repair/cell division?.

has been implicated in controlling chromosome biology and DNA repair as well as splicing and transcription and is also mutated in an X-linked intellectual disability syndrome [44,53,64]. Supporting a wider biological role, RBMX also controls transcription of the *CBX5* gene within leukemia cells [65,66]. Further mechanistic investigations of RBMXL2 and RBMX functions will help to address these issues and should reveal further gene expression pathways that operate during human development and disease.

Disclosure statement

Funding

References

[1] Papasaikas P, Valcárcel J, Eu JV. Valcárcel (J. Special Issue: 40 Years of TiBS the spliceosome: the Ultimate RNA Chaperone and Sculptor. Trends Biochem Sci. 2016;41(1):33–45.

[2] Zarnack K, König J, Tajnik M, et al. Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. Cell. 2013;152(3):453–466.

[3] McClory SP, Lynch KW, Ling JP. HnRNP L represses cryptic exons. RNA. 2018;24(6):761–768.

[4] Attig J, Agostini F, Gooding C, et al. Heteromeric RNP Assembly at LINEs Controls Lineage-Specific RNA processing. Cell. 2018;174(5):1067–1081.e17.

[5] Sibley CR, Blazquez L, Ule J. Lessons from non-canonical splicing. Nat Rev Genet. 2016;17(7):407–421.

[6] Jp L, Pletnikova O, Jc T, et al. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. Science. 2015;349(6248):978–988.

[7] Donde A, Sun M, Ling JP, et al. Splicing repression is a major function of TDP-43 in motor neurons. Acta Neuropathol Internet]. 2019; 138:813–826.

[8] Tan Q, Yalamanchili HK, Park J, et al. Extensive cryptic splicing upon loss of RBM17 and TDP43 in neurodegeneration models. Hum Mol Genet. 2016;25(23):70–81.e9.

[9] Fratta P, Sivakumar P, Humphrey J, et al. Mice with endogenous TDP-43 mutations exhibit gain of splicing function and characteristics of amyotrophic lateral sclerosis. EMBO J. 2018;37(11):e98684.

[10] Sun M, Bell W, Kd L, et al. Cryptic exon incorporation occurs in Alzheimer's brain lacking TDP-43 inclusion but exhibiting nuclear clearance of TDP-43. Acta Neuropathol. 2017;133(6):923–931.

[11] Kan Z, Garrett-Engele PW, Johnson JM, et al. Evolutionarily conserved and diverged alternative splicing events show different expression and functional profiles. Nucleic Acids Res. 2005;33(17):5659–5666.

[12] Soumillon M, Necsulea A, Weier M, et al. Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. Cell Rep. 2013;3(6):2179–2190.

[13] Naro C, Cesari E, Sette C. Splicing regulation in brain and testis: common themes for highly specialized organs [Internet]. Cell Cycle [cited 2021 Jun 17]; 20 (5–6):480–489.

[14] Merkin J, Russell C, Chen P, et al. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. Science. 2012;338(6114):1593–1599.

[15] Mazin PV, Khaitovich P, Cardoso-Moreira M, et al. Alternative splicing during mammalian organ development. Nat Genet. 2021;133(6):923–931.

[16] Schmid R, Grellscheid SN, Ehrmann I, et al. The splicing landscape is globally reprogrammed during male meiosis. Nucleic Acids Res. 2013;41(22):10170–10184.

[17] Carelli FN, Hayakawa T, Go Y, et al. The life history of retrocopies illuminates the evolution of new mammalian genes. Genome Res. 2016;26(3):301–314.

[18] Ramsköld D, Wang ET, Burge CB, et al. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol. 2009;5(12):e1000598.

[19] Cardoso-Moreira M, Halbert J, Valloton D, et al. Gene expression across mammalian organ development. Nature. 2019;571(7766):505–509.

[20] Amann RP, Howards SS. Daily spermatozoal production and epididymal spermatozoal reserves of the human male. J Urol. 1980;124(2):211–215.

[21] Griswold MD. Spermatogenesis: the Commitment to Meiosis. Physiol Rev. 2016;96(1):1–17.

[22] Gamble J, Chick J, Seltzer K, et al. An expanded mouse testis transcriptome and mass spectrometry defines novel proteins. Reproduction. 2020;159(1):15–26.

[23] Naro C, Jolly A, Di Persio S, et al. An Orchestrated Intron Retention Program in Meiosis Controls Timely Usage of Transcripts during Germ Cell Differentiation. Dev Cell. 2017;41(1):82–93.e4.

[24] Chen Y, Zheng Y, Gao Y, et al. Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. Cell Res. 2018;28(9):879–896.

[25] Monesi V. Ribonucleic acid synthesis during mitosis and meiosis in the mouse testis. J Cell Biol. 1964;22(3):521–532.

[26] Turner JMA, Mahadevaiah SK, Fernandez-Capetillo O, et al. Silencing of unsynapsed meiotic chromosomes in the mouse. Nat Genet. 2005;37(1):41–47.

[27] Paronetto MP, Messina V, Barchi M, et al. Sam68 marks the transcriptionally active stages of spermatogenesis and modulates alternative splicing in male germ cells. Nucleic Acids Res. 2011;39(12):4961.

[28] Ernst C, Eling N, Martinez-Jimenez CP, et al. Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. Nat Commun. 2019;10(1):1–20.

[29] Maezawa S, Yukawa M, Alavattam KG, et al. Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. Nucleic Acids Res. 2018;46(2):593–608.

[30] Maezawa S, Sakashita A, Yukawa M, et al. Super-enhancer switching drives a burst in gene expression at the mitosis-to-meiosis transition. Nat Struct Mol Biol. 2020;27(10):978–988.

[31] Venables JP. RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing. Hum Mol Genet. 2000;9 (5):685–94 .

[32] Elliott DJ, Venables JP, Newton CS, et al. An evolutionarily conserved germ cell-specific hnRNP is encoded by a retrotransposed gene. Hum Mol Genet. 2000;9 (14):2117–2124.

[33] Gh W, Gianotten J, Nj L, et al. Heterogeneous nuclear ribonucleoprotein G-T (HNRNP G-T) mutations in men with impaired spermatogenesis. Mol Hum Reprod. 2004;10(4):265–269.

[34] Ehrmann I, Crichton JH, Gazzara MR, et al. An ancient germ cell-specific RNA-binding protein protects the germline from cryptic splice site poisoning. Elife Internet]. [cited 2019 8 Aug 27]; e39304.

[35] Buratti E, Fe B. TDP-43: gumming up neurons through protein-protein and protein-RNA interactions. Trends Biochem Sci. 2012;37(6):237–247.

[36] Sephton CF, Good SK, Atkin S, et al. TDP-43 is a developmentally regulated protein essential for early embryonic development. J Biol Chem. 2010;285 (9):6826–6834.

[37] Wu L-S, Cheng W-C, Hou S-C, et al. TDP-43, a neuro-pathosignature factor, is essential for early mouse embryogenesis. Genesis. 2010;48(1):56–62.

[38] Reddi PP. Transcription and Splicing Factor TDP-43: role in regulation of gene expression in testis. Semin. Reprod. Med. 2017;35(2):167–172

[39] Soh YQS, Mikedis MM, Kojima M, et al. Meioc maintains an extended meiotic prophase I in mice. PLoS Genet. 2017;13(4):e1006704.

[40] Abby E, Tourpin S, Ribeiro J, et al. Implementation of meiosis prophase i programme requires a conserved retinoid-independent stabilizer of meiotic transcripts. Nat Commun. 2016;7:10324.

[41] Connor F, Bertwistle D, Mee PJ, et al. Tumorigenesis and a DNA repair defect in mice with a truncating Brca2 mutation. Nat Genet. 1997;17(4):423–430.

[42] Tam AS, Stirling PC. Splicing, genome stability and disease: splice like your genome depends on it! Curr. Genet. 2019;65:905–912.

[43] Erdos MR, Cabral WA, Tavarez UL, et al. A targeted antisense therapeutic approach for Hutchinson–Gilford progeria syndrome. Nat Med. 2021;27 (3):536–545.

[44] Elliott DJ, Dalgliesh C, Hysenaj G, et al. RBMX family proteins connect the fields of nuclear RNA processing, disease and sex chromosome biology. Int J Biochem Cell Biol. 2019;108:1–6.

[45] Adamson B, Smogorzewska A, Fd S, et al. A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. Nat Cell Biol. 2012;14 (3):318–328.

[46] Turner JMA. Meiotic sex chromosome inactivation. Development. 2007;134(10):1823–1831.

[47] MacDonald CC. Tissue-specific mechanisms of alternative polyadenylation: testis, brain, and beyond (2018 update). Wiley Interdiscip. Rev. RNA. 2019;:10(4): e1526.

[48] Emerson JJ, Kaessmann H, Betrán E, et al. Extensive gene traffic on the mammalian X Chromosome. Science. 2004;303(5657):537–540.

[49] Jiang L, Li T, Zhang X, et al. RPL10L Is Required for male meiotic division by compensating for rpl10 during meiotic sex chromosome inactivation in mice. Curr Biol. 2017;27(10):1498–1505.e6.

[50] Wang PJ. X chromosomes, retrogenes and their role in male reproduction. Trends Endocrinol Metab. 2004;15 (2):79–83.

[51] Long M, Emerson JJ. Meiotic Sex Chromosome Inactivation: compensation by Gene Traffic. Curr Biol. 2017;27(13):R659–61.

[52] Zhou KI, Shi H, Lyu R, et al. Regulation of Co-transcriptional Pre-mRNA Splicing by m6A through the Low-Complexity Protein hnRNPG. Mol Cell. 2019;76(1):70–81.e9.

[53] Cai T, Cinkornpumin JK, Yu Z, et al. Deletion of RBMX RGG/RG motif in Shashi-XLID syndrome leads to aberrant p53 activation and neuronal differentiation defects. Cell Rep. 2021;36(2):109337.

[54] Shum EY, Jones SH, Shao A, et al. The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay. Cell. 2016;165 (2):382–395.

[55] Licatalosi DD Roles of RNA-binding proteins and post-transcriptional regulation in driving male germ cell development in the mouse. In: Advances in

Experimental Medicine and Biology . ; 2016;907:123–151.

[56] Legrand JMD, Hobbs RM. RNA processing in the male germline: mechanisms and implications for fertility. Semin Cell Dev Biol. 2018;79:80–91.

[57] Song H, Wang L, Chen D, et al. The function of pre-mRNA alternative splicing in mammal spermatogenesis. Int J Biol Sci. 2020;16(1):38–48.

[58] Hannigan MM, Zagore LL, Licatalosi DD. Ptbp2 Controls an Alternative Splicing Network Required for Cell Communication during Spermatogenesis. Cell Rep. 2017;19(12):2598–2612.

[59] Jp L, Chhabra R, Jd M, et al. PTBP1 and PTBP2 Repress Nonconserved Cryptic Exons. Cell Rep. 2016;17(1):104–113.

[60] Naro C, Sette C. Timely-regulated intron retention as device to fine-tune protein expression. Cell Cycle. 2017;16(14):1321–1322.

[61] Nasim MT, Chernova TK, Chowdhury HM, et al. HnRNP G and Tra2β: opposite effects on splicing matched by antagonism in RNA binding. Hum Mol Genet. 2003;12(11):1337–1348.

[62] Vaquero-Garcia J, Barrera A, Gazzara MR, et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. Elife. 2016;5:e11752.

[63] Chiou N-T, Shankarling G, Lynch KW. hnRNP L and hnRNP A1 induce extended U1 snRNA interactions with an exon to repress spliceosome assembly. Mol Cell. 2013;49(5):972–982.

[64] Shashi V, Xie P, Schoch K, et al. The RBMX gene as a candidate for the Shashi X-linked intellectual disability syndrome. Clin Genet. 2015;88(4):1067–1081.e17.

[65] Prieto C, Nguyen DTT, Liu Z, et al. Transcriptional control of CBX5 by the RNA-binding proteins RBMX and RBMXL1 maintains chromatin state in myeloid leukemia. Nat Cancer. 2021;2(7):318–328.

[66] Prieto C, Nguyen D, Vu LP, et al. RNA binding protein rbmx is required in acute myeloid leukemia by regulating the transcriptional activity of the Heterochromatin Protein HP1α. Blood. 2018;132(Supplement 1):883.