

Investigating the function of RBMXL2 in spermatogenesis in humans and using mouse models.

Saad Khalid Aldalaqan

Newcastle University Biosciences Institute



A thesis submitted for the degree of Doctor

of Philosophy

(June 2023)

Table of Contents

Chapter 1:	Introduction:	1
1.1	DNA to protein	1
1.2	RNA transcription	2
1.3	mRNA capping and polyadenylation	2
1.4	Splicing of Pre-mRNA and alternative splicing	4
1.5	The splicing reaction and spliceosome machinery	5
1.6	Recognition of splice site	8
1.7	Cryptic Splice sites	8
1.8	Male Germ Cell Development	10
1.9	RNA binding proteins	12
1.10	RNA binding proteins (RBPs) throughout spermatogenesis	13
1.10.1	Nuclear RNA binding proteins important in spermatogenesis	14
1.10.2	The RBMX family of RNA binding proteins	15
1.10.2.1	RBMX	15
1.10.2.2	RBMX	16
1.10.2.3	RBMXL2	17
1.10.3	Other nuclear RNA binding proteins important in spermatogenesis	20
1.10.3.1	Transformer-2 (Tra-2 α and Tra-2 β)	20
1.10.3.2	Sam68 (KHDRBS1)	21
1.10.3.3	PTB protein family	21
1.10.3.4	RNA binding motif-5 (RBM-5)	22
1.10.3.5	DEAD-box helicase-5 (DDX-5)	22
1.10.3.6	Polyadenylation Factors (Testis cleavage stimulation factor-2 τ)	23

1.11	Examples of important Cytoplasmic-RBPs in spermatogenesis	24
1.11.1	Translational Factors important during spermatogenesis	24
1.11.1.1	DAZ protein family	24
1.11.1.2	Y-Box proteins	25
1.11.1.3	Pumilio and Nanos	26
1.11.1.4	Cytoplasmic polyadenylation element-binding protein (CPEB)	27
1.12	Hypotheses and project aims.....	28
1.12.1	Background to this thesis.....	28
1.12.2	Project aims:	29
1.12.2.1	Investigating whether Knockout of <i>Rbmxl2</i> will show different phenotypes in different mice strains?	29
1.12.2.2	Identification of RNA targets of RBMXL2 in mouse testis using RNA-seq and iCLIP: 29	
1.12.2.3	Is RBMXL2/RBMY provides a direct or specialist replacement for RBMX during Meiosis?.....	29
Chapter 2:	Material and methods:.....	30
2.1	Mouse lines.....	30
2.2	Animal care	30
2.3	Genotyping	30
2.4	Histological sections:	30
2.4.1	Paraffin wax embedding of mouse testes	31
2.4.2	Testis sectioning.....	32
2.5	Staining:	32
2.5.1	H&E staining.....	32
2.5.2	Immunolocalisation using Fluorescent staining	33
2.6	Minigene assays	33
2.6.1	Cloning of mouse <i>Esco1</i> exon 3 into an exon trap vector and analysis in	

	<i>HEK293 cells</i>	33
2.6.2	<i>Cloning of human Esco1 exon 4 into an exon trap vector and analysis in HEK293 cells.</i>	36
2.6.3	<i>Primer design</i>	39
2.6.4	<i>Purification of DNA fragments from agarose-TAE gels</i>	40
2.6.5	<i>Preparation of plasmid DNA and bacterial cloning</i>	41
2.6.6	<i>Sanger sequencing</i>	41
2.6.7	<i>Cell culture</i>	41
2.6.8	<i>Cell passages and maintenance</i>	42
2.6.9	<i>Cell counting</i>	42
2.6.10	<i>Lipofectamine 3000 transfection</i>	42
2.6.11	<i>RNA extraction</i>	44
2.6.12	<i>RT-PCR</i>	45
2.6.13	<i>Calculation of percentage splicing inclusion (PSI %):</i>	45
2.6.14	<i>Western blotting</i>	46
2.7	<i>iCLIP</i>	47
2.7.1	<i>Definition</i>	47
2.7.2	<i>Steps:</i>	47
2.7.2.1	<i>UV cross-linking of mice testes</i>	47
2.7.2.2	<i>Preparation of magnetic beads</i>	47
2.7.2.3	<i>Cell lysis, partial RNA digestion and immunoprecipitation</i>	48
2.7.2.4	<i>De-phosphorylation of RNA 3' end and linker ligation</i>	48
2.7.2.5	<i>SDS-PAGE and membrane transfer</i>	49
2.7.2.6	<i>RNA isolation</i>	49
2.7.2.7	<i>Reverse transcription</i>	50
2.7.2.8	<i>cDNA purification</i>	50
2.7.2.9	<i>Ligation of primer to the 5' end of the cDNA</i>	51

2.7.2.10	Circularised cDNA amplification	51
2.7.2.11	Purification with AMPure	51
2.7.2.12	Size selection of cDNA library by gel purification	52
2.7.2.13	High-throughput sequencing	52
2.7.2.14	Bioinformatics analysis of iCLIP sequencing data	55
2.7.2.15	Gene Ontology (GO) enrichment analysis	55
2.7.2.16	Analysis of long mouse exons	55
Chapter 3:	Do different mouse strains have a different Rbmxl2 knockout	
phenotype?	61	
3.1	Introduction:	61
3.2	Hypothesis:	63
3.3	Aims:	63
3.4	Results.....	64
3.4.1	<i>Rbmxl2</i> knockout mice do not have different testis weight between the C57BL/6 and Sv/129 strains.....	64
3.4.2	<i>Rbmxl2</i> knockout mice have an increased content of “Sertoli Cell Only” tubules in the Sv/129 strain compared to C57BL/6	66
3.4.3	<i>Rbmxl2</i> deletion causes a Stage IV arrest in the Sv/129 background compared to C57BL/6.....	68
3.4.4	C57BL/6 and Sv/129 strains have similar patterns of expression of RBMXL2 protein over germ cell development	72
3.4.5	Staining of wild type testes with γ H2AX as a marker for the sex body .	75
3.4.6	Investigating important genes regulated by RBMXL2 over the first wave of spermatogenesis	77
3.4.6.1	<i>Kdm4d</i> has an earlier cryptic splicing between in Sv/129 over the first wave of spermatogenesis.	77
3.4.6.2	<i>Meioc</i> has an earlier cryptic splicing in Sv/129 over the first wave of spermatogenesis.....	82

3.4.7	<i>Rbmxl2</i> knockout has additional DNA breakage in the first wave of spermatogenesis of Sv/129 mice strain.	86
3.5	Discussion:	89
3.5.1	<i>Rbmxl2</i> knockout mice have different phenotypes between the Sv/129 and C57BL/6	89
3.5.2	<i>RBMXL2</i> protein expression pattern is similar in both mouse strains	90
3.5.3	Defective splice patterns appear earlier in Sv/129 in the first wave of spermatogenesis	90
3.5.4	Investigating the first wave of spermatogenesis of Sv/129 <i>Rbmxl2</i> knockout	91
3.6	Chapter summary:	92
Chapter 4:	Identification of novel targets of RBMXL2-RNA protein contact using iCLIP	93
4.1	Introduction:	93
4.2	Aims:	95
4.3	Results:	95
4.3.1	Optimisation of the RBMXL2 iCLIP experiment.....	96
4.3.2	RBMXL2 iCLIP analysis of the genome-wide RNA targets	99
4.3.3	Location of RBMXL2 binding within the genome.....	101
4.3.4	Analysis of known regulated targets of RBMXL2 protein for iCLIP RNA binding site	104
4.3.5	Identification of RBMXL2 binding site sequences (this analysis was done with collaborators mentioned below).	114
4.3.6	RBMXL2 Binds the 3' end of the transcript	116
4.3.7	Investigation of RBMXL2, RBMX mechanism to regulate mouse <i>Esco1</i> gene using a Minigene assay.....	117
4.3.8	Investigation of RBMXL2, RBMX mechanism to regulate human <i>ESCO1</i> gene using Minigene assay	125
4.4	Discussion.....	130
4.4.1	RBMXL2 iCLIP in mouse testes	130

4.4.2	RBMXL2 binding within the genome.....	131
4.4.3	Identifying RBMXL2 binding motifs.....	131
4.4.4	RBMXL2-bound genes that have a role in spermatogenesis and meiosis.....	131
4.4.5	Mouse and human ESCO1 minigene assays did not follow the expected pattern of splicing regulation by RBMX and RBMXL2	132
4.5	Chapter summary:.....	133
Chapter 5 : RBMXL2 is functionally interchangeable with RBMX in somatic cells....		134
5.1	Introduction:.....	134
5.2	Hypothesis:	137
5.3	Aims:	138
5.4	Results:.....	138
5.4.1	<i>Establishment of a new strategy to investigate RBMXL2 using an inducible stable cell line approach</i>	<i>138</i>
5.4.2	<i>RT-PCR analysis of some initial targets of RBMX to establish if the rescue experiment was working:.....</i>	<i>140</i>
5.4.3	<i>RNA-seq data from knockdown of RBMX with and without overexpression of RBMXL2:</i>	<i>146</i>
5.4.4	<i>Experiment to test if RBMY is able to replace RBMX function in somatic cells:</i>	<i>153</i>
5.4.5	<i>Investigating the importance of the RBMXL2 (RRM) in RNA binding:.....</i>	<i>159</i>
5.5	Discussion:	165
5.5.1	<i>RBMXL2 is replacing RBMX in somatic cells:.....</i>	<i>165</i>
5.5.2	<i>RBMY another member of the family is also replacing the function of RBMX:.....</i>	<i>166</i>
5.5.3	<i>The disordered domain of RBMXL2 is responsible for RNA-protein binding:.....</i>	<i>166</i>
5.6	Chapter summary:	167

Chapter 6 : Discussion and future work:	168
6.1 <i>Rbmxl2</i> knockout has a severe phenotype on the Sv/129 background.	168
6.2 Using iCLIP to identify direct targets of RBMXL2:.....	169
6.3 RBMXL2 can substitute for RBMX and RBMY in cell lines to ensure accurate splicing of ultra-long exons.	171
6.4 Conclusion:.....	172

Table of figures:

FIGURE 1. 1: INVOLVEMENT OF THE CAP BINDING COMPLEX IN RNA PROCESSING.....	3
FIGURE 1. 2: A SCHEMATIC SHOWING DIFFERENT MODES OF NORMAL ALTERNATIVE SPLICING EVENTS.....	4
FIGURE 1. 3: SPLICING COMPRISES TWO TRANSESTERIFICATION REACTIONS.	6
FIGURE 1. 4: A SCHEMATIC OF THE SPLICEOSOME CYCLE. REPRESENTING THE INTERACTIONS OF THE MAJOR SNRNPs.	7
FIGURE 1. 5: A SCHEMATIC SHOWING SOME CRYPTIC SPLICE EVENTS SUCH AS THE CRYPTIC 5' AND 3' SPLICE SITE AND THE EXITRON.....	9
FIGURE 1. 6: MOUSE SPERMATOGENESIS.....	11
FIGURE 1. 7: THE ORIGIN OF THE RBMX FAMILY OF GENES. <i>RBMY</i> , A PARALOG OF <i>RBMX</i> , SHARES 60% OF <i>RBMX</i> 'S DNA.	16
FIGURE 1. 8: THE PROTEIN STRUCTURES OF THE RBMX FAMILY. IN BLUE THE RRM (RNA RECOGNITION MOTIF) THAT BINDS TO RNA.....	17
FIGURE 2. 1 : CLONING AN ESCO1 MINIGENE: A) THE ESCO1 EXON 3 TRACK FROM UCSC BROWSER (HTTP://GENOME.UCSC.EDU).	35
FIGURE 2. 2: CLONING A HUMAN ESCO1 EXON 4 MINIGENE. A) HUMAN ESCO1 EXON 4 SHOWN IN THE GENCODE TRACK FROM THE UCSC GENOME BROWSER (HTTP://GENOME.UCSC.EDU)	37
FIGURE 2. 3: ILLUSTRATING MINIGENE EXPERIMENT.	38
FIGURE 2. 4: VISUAL ILLUSTRATION OF THE ICLIP PROTOCOL.	56
FIGURE 2. 5: THE EXPERIMENT DESIGN OF RESCUE EXPERIMENT	60
FIGURE 3. 1: MOUSE TESTES STAGING CHART FROM STAGE 1 TO 12. TAKEN WITH MODIFICATION FROM (RUTHIG & LAMB, 2022).....	63
FIGURE 3. 2: TESTIS WEIGHT COMPARISON BETWEEN THE RBMXL2 KNOCKOUTS IN THE C57BL/6 AND Sv/129 STRAINS (N=5).....	65
FIGURE 3. 3: HISTOLOGICAL COMPARISON OF TESTIS DEVELOPMENT FOR RBMXL2 KNOCKOUT MICE BETWEEN THE Sv/129 AND C57BL/6 MOUSE STRAINS AT X20 MAGNIFICATION (N=1)	66
FIGURE 3. 4: HISTOLOGICAL SECTION OF WT AND RBMXL2 KNOCKOUT MICE TESTIS (N=3).	67
FIGURE 3. 5: HISTOLOGICAL SECTION OF WT AND RBMXL2 KNOCKOUT (C57BL/6).	69
FIGURE 3. 6: HISTOLOGICAL SECTION OF RBMXL2 KNOCKOUT (Sv/129) VS. WILD TYPE.....	71
FIGURE 3. 7: FLUORESCENT STAINING OF RBMXL2 PROTEIN IN WILD TYPE TESTIS (Sv/129).	73
FIGURE 3. 8: FLUORESCENT STAINING OF WILD TYPE TESTIS (Sv/129).	73
FIGURE 3. 9: FLUORESCENT STAINING WITHOUT SECONDARY ANTIBODY OF RBMXL2 ON WILD TYPE TESTIS (Sv/129)	74
FIGURE 3. 10: FLUORESCENT STAINING OF WILD TYPE TESTIS (Sv/129). γH2AX STAINED PRELEPTOTENE AND SEX BODY OF PACHYTENE IN GREEN.	75

FIGURE 3. 11: FLUORESCENT STAINING OF TWO WILD-TYPE MOUSE TESTES (Sv/129 AND C57BL/6), STAGE X. RBMXL2 PROTEIN IS PSEUDOCOLOURED RED, AND γ H2AX IS PSEUDOCOLOURED GREEN. THE RBMXL2 PROTEIN CELL TYPE EXPRESSION WAS IDENTICAL IN BOTH TUBULES.	76
FIGURE 3. 12: RNASEQ TRACK OF THE KDM4D GENE ON THE UCSC WEBSITE. RNASEQ DATA WAS GENERATED FROM RBMXL2 KNOCKOUT TESTES AND WILD TYPE (EHRMANN ET AL., 2019).	78
FIGURE 3. 13: DIAGRAM DEPICTS THE MRNA PRODUCT WITH AND WITHOUT RBMXL2.	79
FIGURE 3. 14: CAPILLARY GEL ELECTROPHORESIS SHOWED THE SPLICING PATTERNS OF KDM4D BETWEEN THE WILD TYPE AND RBMXL2 KNOCKOUT TESTES IN BOTH MOUSE STRAINS (N=1)	80
FIGURE 3. 15: KDM4D EXPRESSION WAS INVESTIGATED IN THIS TIME COURSE STUDY COMPARING C57BL/6 AND SV/129 BACKGROUNDS (N=1).	81
FIGURE 3. 16: RNASEQ TRACK OF THE MEIOC GENE ON THE UCSC WEBSITE INCLUDING RNASEQ DATA GENERATED FROM RBMXL2 KNOCKOUT TESTES AND WILD TYPE.	82
FIGURE 3. 17: GEL ELECTROPHORESIS SHOWED THE SPLICING OF MEIOC BETWEEN TESTES FROM WILD TYPE AND RBMXL2 KNOCKOUT TESTES IN BOTH MOUSE STRAINS (N=1)..	83
FIGURE 3. 18: MEIOC EXPRESSION WAS INVESTIGATED IN THIS TIME COURSE STUDY COMPARING C57BL/6 AND SV/129 BACKGROUNDS (N=1).	85
FIGURE 3. 19: ANALYSIS OF DNA DAMAGE IN SV/129 WILD TYPE VS. RBMXL2 KNOCKOUT TESTIS (N=3).	87
FIGURE 3. 20: A DIAGRAM ILLUSTRATES THE MEIOTIC ARREST BETWEEN THE TWO MOUSE STRAINS, Sv/129 AND C57BL/6, AFTER KNOCKOUT OF THE <i>RBMXL2</i> GENE.	90
FIGURE 4. 1: VISUAL ILLUSTRATION OF THE ICLIP PROTOCOL.	94
FIGURE 4. 2: WESTERN BLOT TO SHOW THE SPECIFICITY OF THE RBMXL2 ANTIBODY.	95
FIGURE 4. 3: ICLIP EARLY OPTIMISATION OF IMMUNOPRECIPITATIONS STEP.	97
FIGURE 4. 4: TBE AGAROSE GEL SHOWING EXPERIMENTS TO OPTIMISE PRODUCTION OF cDNA LIBRARIES AT 15, 19, AND 30 CYCLES WITH A DNA SIZE MARKER (LADDER).	98
FIGURE 4. 5: TAPESTATION RESULT FOR THE cDNA LIBRARY OF ONE OF ICLIP SAMPLES. THE AMPLIFICATION PRODUCT WAS AROUND 143 BP (REGION DESIGNATED IN THE DOTTED RED BOX).	98
FIGURE 4. 6: A CORRELATION ANALYSIS BETWEEN THE 4 BIOLOGICAL REPLICATES SHOWING THE POSITIVE CORRELATION. CORRELATION MATRIX MADE WITH DR. CHILE SIACHISUMO'S ASSISTANCE ON (HTTPS://USEGALAXY.ORG/).	100
FIGURE 4. 7: BAR CHART SHOWING WHERE RBMXL2 PROTEIN BINDS TO DIFFERENT REGIONS OF THE GENOME. RBMXL2 BINDS STRONGLY IN INTERGENIC AND INTRONIC REGIONS.	101
FIGURE 4. 8: BAR CHART SHOWING THE SUBTYPES OF GENE THAT ARE BOUND BY RBMXL2 PROTEIN.	102
FIGURE 4. 9: A CHART REPRESENTING GENE ONTOLOGY ENRICHMENT ANALYSIS OF RBMXL2 BOUND GENES (TOP 500 GENES) IDENTIFIED BY ICLIP.	103
FIGURE 4. 10: A SCREENSHOT FROM IGV SHOWS THE EFFECT ON PROCESSING OF MEIOC EXON 5 AFTER KNOCKOUT OF THE <i>RBMXL2</i> GENE VERSUS THE WILD TYPE.	104
FIGURE 4. 11: CRYPTIC SPLICING OF <i>ESCO1</i>	105
FIGURE 4. 12: RBMXL2 BINDING TO <i>ESCO1</i> . SCREENSHOTS FROM IGV SHOWED SEVERAL ICLIP TAGS BOUND TO EXON 3 OF <i>ESCO1</i>	106
FIGURE 4. 13: AN EXITRON WITHIN THE <i>MEIOC</i> GENE.	107

FIGURE 4. 14: ICLIP MAPPING OF RBMXL2 RNA PROTEIN INTERACTIONS WITHIN MEIOC EXON 5.	108
FIGURE 4. 15: CRYPTIC SPLICING WITHIN <i>KDM4D</i>	109
FIGURE 4. 16: MAPPING OF ICLIP TAGS WITHIN THE <i>KDM4D</i> GENE	110
FIGURE 4. 17: CRYPTIC SPLICING IN <i>ALMS1</i> . 5' CRYPTIC SS.	111
FIGURE 4. 18: MAPPING OF ICLIP TAGS NEAR TO THE CRYPTIC SPLICE SITE USED IN <i>ALMS1</i>	112
FIGURE 4. 19: CRYPTIC SPLICING IN <i>BRCA2</i>	113
FIGURE 4. 20: CRYPTIC SPLICING OF <i>BRCA2</i>	113
FIGURE 4. 21: TOP 50 K-MER ENRICHMENT DETERMINED FROM ANALYSIS OF MY ICLIP DATA. PINK IS THE CONTROL (ICLIP DONE WITH IGG ANTIBODY) WHILE GREY IS RBMXL2 PROTEIN BINDING TO THE MOTIFS. THE MOST ENRICHED MOTIF WAS TACATT.....	114
FIGURE 4. 22: RBMXL2 BINDS AND REGULATES RNA PROCESSING EVENTS IN ULTRA-LONG EXONS.	115
FIGURE 4. 23: METAGENE ANALYSIS OF RBMXL2 ICLIP TRACKS FROM MOUSE TESTES SHOWN FULL TRANSCRIPT REGIONS NORMALIZED PER GENE. HIGHEST RBMXL2 BINDING DENSITY WERE SEEN TOWARDS THE 3' END OF THE GENE.	116
FIGURE 4. 24: A SCREENSHOT OF A SASHIMI PLOT SHOWS RNA-SEQ DATA FROM DAY 18 MOUSE TESTIS, THREE WILD TYPE SAMPLES THEN THREE <i>RBMXL2</i> KNOCKOUT SAMPLES.....	118
FIGURE 4. 25: A SCHEMATIC OF THE <i>ESCO1</i> EXON 3 MINIGENE DESIGN.....	119
FIGURE 4. 26: : A SCHEMATIC FIGURE SHOWING DOMAIN STRUCTURES OF RBMXL2 AND RBMX PROTEINS AND PROTEINS CONSTRUCT MISSING DOMAINS OF RBMXL2 AND RBMX PROTEINS TAGGED WITH GFP.	120
FIGURE 4. 27: QIAXCEL ELECTROPHORETOGRAM DISPLAYING SPLICING PATTERNS OF THE <i>ESCO1</i> MINIGENE.	120
FIGURE 4. 28: RBMX WEAKLY SUPPRESSES THE USE OF THE CRYPTIC 3' SPLICE SITE IN <i>ESCO1</i> EXON 3 (N=3).....	121
FIGURE 4. 29: WESTERN BLOT SHOWS PLASMIDS WERE EXPRESSING SIMILAR LEVELS OF RNA BINDING PROTEINS IN TRIPLICATE TRANSFECTIONS. WESTERN BLOTS WERE PROBED WITH ANTIBODIES SPECIFIC FOR GFP. NO LOADING CONTROL WAS PERFORMED FOR THIS EXPERIMENT.....	121
FIGURE 4. 30: SCHEMATIC FIGURE OF <i>ESCO1</i> GENE EXON 3 WITH INTERNAL PRIMER TO DETECT THE FULL LENGTH, INSERTED WITHIN PXJ41 PLASMID.....	123
FIGURE 4. 31: QIAXCEL ELECTROPHORETOGRAM DISPLAYING THE FULL-LENGTH SPLICING OF EXON 3 OF THE <i>ESCO1</i> GENE (N=3).....	124
FIGURE 4. 32: EXON 4 OF THE <i>ESCO1</i> GENE IS ACTIVATED BY RBMX..	126
FIGURE 4. 33: A SCHEMATIC OF THE <i>ESCO1</i> EXON 4 MINIGENE DESIGN.	126
FIGURE 4. 34: RBMXL2 AND RBMX SUPPRESS THE SPLICING OF EXON 4 OF HUMAN <i>ESCO1</i> GENE (N=3).....	127
FIGURE 4. 35 :RBMXL2 AND RBMX SUPPRESSES THE ACTIVATION OF EXON 4 IN HUMAN <i>ESCO1</i> GENE (N=3).	129
FIGURE 5. 1: CHORD PLOT SHOWING GENE ONTOLOGY OF ENRICHED GENES INVOLVED IN CELL CYCLE AND DNA REPAIR WITHIN RBMX REGULATED EXONS.	136
FIGURE 5. 2: A WESTERN BLOT SHOWING THE SUCCESSFUL KNOCKDOWN OF RBMX COMPARED TO THE siRNA CONTROL AND TUBULIN AS A CONTROL.....	139
FIGURE 5. 3: A SNAPSHOT OF PRIMERS DESIGNED FROM THE UCSC WEBSITE (HTTP://GENOME.UCSC.EDU/).....	141
FIGURE 5. 4: CAPILLARY GEL ELECTROPHORETOGRAMS SHOW RBMXL2'S ABILITY TO SUBSTITUTE RBMX IN SOMATIC CELLS AND ENSURE THE PROPER SPLICING INCLUSION OF ULTRA-LONG EXONS ON DIFFERENT GENES (SIACHISUMO ET AL., 2023).....	142

FIGURE 5. 5: RNASEQ DATA USED TO ANALYSE THE EFFECTS OF THE RESCUE EXPERIMENT ON ESCO1 GENE. THE RNASEQ DATA WAS VISUALISED USING A SASHIMI PLOT. PICTURE TAKEN AS A BROWSER SNAPSHOT FROM THE INTEGRATIVE GENOMICS VIEWER (ROBINSON ET AL., 2011).....	146
FIGURE 5. 6:RBMXL2 CAN SUBSTITUTE THE FUNCTION OF RBMX IN ENSURING APPROPRIATE SPLICING AND INCLUSION OF ETAA1 EXON 5.....	147
FIGURE 5. 7: RBMXL2 CAN SUBSTITUTE THE FUNCTION OF RBMX IN ENSURING APPROPRIATE SPLICING AND INCLUSION OF REV3L.....	149
FIGURE 5. 8: RBMXL2 CAN SUBSTITUTE THE FUNCTION OF RBMX IN ENSURING APPROPRIATE SPLICING AND INCLUSION OF ATRX EXON 9.....	151
FIGURE 5. 9: RNASEQ DATA USED TO ANALYSE THE EFFECTS OF THE RESCUE EXPERIMENT ON RBM41 GENE.	153
FIGURE 5. 10: WESTERN ANALYSIS OF RBMY RESCUE EXPERIMENT.	154
FIGURE 5. 11: CAPILLARY GEL ELECTROPHORETOGRAMS SHOW RBMY'S ABILITY TO SUBSTITUTE FOR RBMX IN SOMATIC CELLS AND ENSURE THE PROPER SPLICING INCLUSION OF ULTRA-LONG EXONS ON DIFFERENT GENES (SIACHISUMO ET AL., 2023).....	155
FIGURE 5. 12: COMPARISON OF THE FULL-LENGTH RBMXL2 PROTEIN AND RBMXL2 Δ RRM THAT WAS USED FOR THE STABLE CELL LINE.....	160
FIGURE 5. 13: WESTERN ANALYSIS OF THE RBMXL2 Δ RRM RESCUE EXPERIMENT.....	160
FIGURE 5. 14: CAPILLARY GEL ELECTROPHORETOGRAMS SHOW RBMXL2 Δ RRM'S ABILITY TO SUBSTITUTE FOR RBMX IN SOMATIC CELLS AND ENSURE THE PROPER SPLICING INCLUSION OF ULTRA-LONG EXONS ON DIFFERENT GENES (SIACHISUMO ET AL., 2023).....	161

Abstract

RBMX, RBMXL2, and RBMY belong to a family of RNA-binding proteins. How important these proteins are in whole animal biology, what their global RNA targets and mechanisms of action are, and to what extent these proteins overlap in function has been poorly understood.

Using histology I found that *Rbmxl2* knockout phenotypes differed between mouse strains (C57BL/6 and Sv/129). Both strains, however, were infertile, with no sperm found, which indicates the importance of RBMXL2 in spermatogenesis. I performed an iCLIP analysis to investigate RBMXL2-RNA binding, which revealed that RBMXL2 directly binds to exons and introns of protein-coding genes. The strongest RBMXL2 binding was observed in genes associated with spermatogenesis and reproduction, specifically meiosis, such as *Meioc* and *Esco1*. Binding was also enriched within ultra-long exons to prevent the selection of cryptic splice sites. I attempted minigene experiments to try and model RBMXL2 function in transfected cells, but these experiments did not replicate RBMXL2 function in vivo. As an alternative I developed a stable cell line “rescue” approach to over-express RBMXL2 and test if this could replace endogenous RBMX. Using this approach, I demonstrated that RBMXL2 is likely to replace RBMX's function in meiosis when the X and Y are inactive. RBMXL2's ability to restore splicing control is dependent on its disordered domain not the RRM. RBMY, which is distantly related to RBMX, also replaces RBMX function in somatic cells.

These experiments show that RBMXL2 is crucial in male fertility, but there are differences in how different mouse strains respond to deletion of the *Rbmxl2* gene. RBMXL2 binds to RNAs from genes important in spermatogenesis. Despite the divergence between RBMX and RBMXL2 and RBMY these proteins can still substitute for each other suggesting they fulfil the same functions in their respective cell types.

Covid-19 Impact Statement

The university was closed in response to the Covid-19 pandemic, and access to the laboratory was restricted. This had a significant effect on the work I included in this thesis. During the first and second years, the pandemic occurred. Prior to the March 2020 lockdown, I had just begun to optimise my staining.

Due to the uncertainty, the Elliott group eliminated most of the Sv/129 and C57BL/6 mouse backgrounds, significantly slowing down my project. I was unable to collect the necessary mouse testis sooner to determine the histological differences between the two mouse strains or to start optimising the iCLIP experiment.

Despite this, I returned to Saudi Arabia under a total curfew during the UK's lockdown. As a result, I spent a total of six months in my home country, caring for my children and monitoring their home- schooling. Nonetheless, I was able to publish a review article in the journal cell cycle in 2021.

I returned to the United Kingdom after restrictions were lifted. However, social distancing restricted the amount of time and space I could spend in the lab. We were operating at a limited capacity. This made it extremely challenging for me to make up for lost time. Due to social distancing restrictions, only 8 daily lab spaces were available for our group of approximately 13 researchers. To allow others to complete their own work, we had to work half-days and come in only a few days per week. This has forced me to reduce the amount of work I can complete daily and weekly, delaying my ability to collect data for my project.

Declaration

I, Saad Aldalaqan, declare that this thesis is my original work, unless otherwise noted. I confirm that the thesis was not submitted for any other degree or qualification. I have properly acknowledged all sources of information and included all relevant citations.

Saad Khalid Aldalaqan

(2023)

Dedication

I want to dedicate this academic achievement to the unwavering support and love of my dear mother Jamelah, whose belief in me has been an endless source of inspiration. Your encouragement and sacrifices have paved the way for my success, and I am forever grateful for your staunch faith in my abilities. This accomplishment is a testament to your endless support and guidance throughout my educational journey.

To my wife Mai Aldalaqan and my beloved children, Abdulaziz and Talal, this achievement is dedicated to you. You have been my greatest motivation and driving force, reminding me of the importance of perseverance and hard work. Your unconditional love and understanding during the countless hours spent studying and researching have been a constant source of strength. I dedicate this milestone to you, hoping it serves as a reminder that anything is possible with determination and dedication.

Thank you, Mom, Mai and my dear children, for your endless support, patience, and belief in my dreams. This achievement is as much yours as it is mine, and I am profoundly grateful to have you by my side on this journey.

Acknowledgment

I sincerely thank everyone who has supported me during this project. While I would like to express gratitude to many, I can only acknowledge a few named individuals. I would not have been able to complete this work without the patience and guidance of all those involved.

First and foremost, I'd like to express my gratitude to my supervisor and mentor, Professor David Elliott. His belief in me and allowing me to undertake this PhD has been a tremendous privilege. I cannot imagine going through this journey with any other supervisor. Professor Elliott has displayed incredible patience and has provided immense knowledge and experience that have encouraged and supported me throughout my research. Without his expertise and guidance, this achievement would not have been possible. I am fortunate to have had such a dedicated and exceptional supervisor.

My second supervisor, Dr. Louise Reynard, provided invaluable guidance, support, and expertise throughout my thesis. Her dedication, insightful feedback, and commitment to my academic development have significantly impacted the quality and direction of my research. In addition, the time she spent with me under the microscope was beneficial to this thesis.

Secondly, I extend my gratitude to all the members of the Elliott and Munkley labs for their unwavering support and invaluable advice throughout my years of study. I want to give special mention to Mrs Caroline Dalgliesh, who has been a great friend and a patient listener to all my questions, offering valuable advice. I am particularly thankful to Dr Chile Siachisumo, who has provided significant help and support in my project. She has been a mentor and friend, being patient through all the optimization steps of the iCLIP experiment. I also want to acknowledge Dr Ingrid Ehrmann for her help and guidance, which have instilled confidence in me and provided helpful advice. I also extend my gratitude to Dr Sara Luzzi for her help and guidance, especially in the field of bioinformatics. Finally, I want to extend my gratitude to all members of Professor Elliott's Group, Dr Alice Coomer, Rahul Advani, Kayla Bastian and Farimah Ghorbani. I am deeply grateful to Professor Ian Adams of the MRC Human Genetics Unit at the University of Edinburgh for his invaluable guidance and patience during my learning journey. His expertise in staging has significantly enhanced and positively impacted my research

Mr Rafiqul Hussain, your expertise on library purification during the iCLIP experiment troubleshooting was invaluable. I want to thank collaborators: Dr Graham R Smith, Dr Ivaylo D Yonchev (Sheffield University), whose contributions helped with this project.

Thirdly, I would like to acknowledge my progression panel, Professor Helen Arthur and Dr Michael Jackson, for their support and helpful advice throughout my project. Finally, I want to thank my brothers and sisters: Noor, Roaa and Ali. And all my friends in Saudi Arabia and the UK.

Without the help and encouragement of the aforementioned people, as well as numerous more, I would not have been able to complete my PhD. I would want to express my gratitude to everyone involved.

List of Abbreviations

ASO	antisense oligonucleotide
APA	Alternative polyadenylation
ASE	alternatively spliced exon
BP	branchpoint
BSA	bovine serum albumin
CBC	cap-binding complex
cDNA	complementary DNA
CLIP	UV cross-linking immunoprecipitation
CTD	carboxyl terminal domain
DAPI	4',6-diamidino-2-phenylindole
DEPC	Diethyl pyrocarbonate
dH ₂ O	distilled water
DMSO	dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	double strand breaks
DTT	Dithiothreitol
eCLIP-seq	enhanced CLIP sequencing
EDTA	Ethylenediaminetetraacetic acid
EtoH	Ethanol
FBS	Fetal Bovine Serum
FFPE	Formalin-Fixed Paraffin-Embedded
GFP	Green Fluorescent Protein
GO	gene ontology
H&E	Hematoxylin and eosin
HCE	human capping enzyme
HEK293 cells	human embryonic kidney cells
HITS-CLIP	High-throughput sequencing CLIP
hnRNP	heterogeneous nuclear ribonucleoproteins
HR	Homologous recombination

iCLIP	Individual-nucleotide resolution CLIP
IF	immunofluorescence
IGV	Integrative Genomics Viewer
IP	immunoprecipitation
IR	infrared
irCLIP-seq	infrared CLIP sequencing
lncRNA	long intergenic non-coding RNA
KO	Knockout
NEB	New England Biolabs
NMD	Nonsense-mediated RNA decay
MDA-MB-231	MD Anderson human invasive breast cancer cell line
mRNA	messenger RNA
ml	Millilitre
MT	methyl transferase
M7G	N7-methylguanosine
mins	Minutes
nt	Nucleotides
O/E	Overexpression
PAR-CLIP-seq	photoactivatable ribonucleoside-enhanced CLIP-sequencing
PBS	phosphate-buffered saline
PBS-T	phosphate-buffered saline- Tween20
PFA	4% Paraformaldehyde solution
PCR	polymerase chain reaction
pre-mRNA	precursor mRNA
PSI	percentage splicing inclusion
PNK	Polynucleotide Kinase
RBMX	RNA Binding Motif Protein X-Linked
RBMXL2	RBMX Like 2
RBMX	RNA Binding Motif Protein Y-Linked
RBD	RNA binding domain

RBP	RNA binding protein
RGG domain	glycine-arginine-rich (GAR) domain
RIP-seq	RNA Immunoprecipitation sequencing
RNA	ribonucleic acid
Rnase	Ribonuclease
RNA-seq	RNA-Sequencing
RNA Pol II	RNA polymerase II
RPM	Revolutions per minute
RPMI	Roswell Park Memorial Institute 1640 medium
RRM	RNA recognition motif
RS domain	arginine/serine domain
RT-PCR	reverse transcriptase-PCR
RT	Room temperature
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
siRNA	small interfering RNA
SnRNPs	small nuclear ribonucleoproteins
SR protein	serine/arginine-rich protein
TBE	Tris-Borate-EDTA
TBS	Tris Buffered Saline
T-BST	Tris-Buffered Saline with Tween-20
TE	Tris/EDTA
Tra2 β	Transformer-2 protein homolog beta
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
WT	Wild type
UCSC	University of California, Santa Cruz
μ M	Micro molar
μ l	Microliter
UMI	Unique molecular identifier

UTR	Untranslated region
UV	Ultraviolet

Chapter 1: Introduction:

1.1 DNA to protein

In 2003, the draft human genome sequence was completed. This achievement allows scientists to navigate the whole human genome. Molecular biology is based on the idea that genetic information transcribed from DNA to make RNA then translated to produce a protein. This code simplifies the nature of these complex processes, from DNA to proteins. Bacterial DNA, for example, is transcriptionally converted into RNA before being translated into proteins. Nevertheless, the situation is different in complex organisms, such as humans and animals.

While transcription and translation occur individually, DNA is transcribed to the precursor messenger RNA “pre-mRNA” that, before translation, has to go through a chains of post-transcriptional alterations, starting with 5' capping, splicing events, and polyadenylation. The pre-mRNA is made up of sequences that code for proteins (called exons) and sequences that do not code for proteins (called introns). After 5' capping, pre-mRNA splicing takes place, in which exons are joined together to be translated into protein and introns are left behind, forming a lariat that later will be degraded.

Alternative splicing is the process by which a single pre-mRNA can make more than one mRNA transcript, producing different types of proteins (Chow et al., 1977). Different protein isoforms often have different properties, such as cellular localisation, transcriptional efficiency, and protein stability (Wang et al., 2015). Almost 95% of all human genes can be spliced in more than one way, resulting in multiple pre-mRNAs from the same gene (Pan et al., 2008). With the help of alternative splicing, eukaryotic organisms can change how their mRNA and proteins are expressed without making their genomes bigger. Alternative splicing is a primary mechanism that drives the evolution of eukaryotic organisms and is crucial for development (Kalsotra & Cooper, 2011). Alternative splicing and alternative polyadenylation allow eukaryotic cells to make more than 90,000 proteins (Lander, 2011; Harrison et al., 2002) from around 21,000 genes (Pertea et al., 2018).

1.2 RNA transcription

Transcription is a complex process which occurs in three steps: initiation, elongation, and finally termination. RNA polymerase II (RNA-Pol-II) is the enzyme and master regulator which cooperates with certain processing factors to facilitate transcription of mRNAs (Glover-Cutter et al., 2008). RNA Pol-II initiates transcription by binding to gene's promoter region leading to the unwinding of the DNA (Mischo & Proudfoot, 2013). In addition, it reads the complementary strand of the DNA and adds nucleotides to the transcribed RNA. RNA Pol-II eventually detects the signal that indicates the end of transcription. It activates the polyadenylation complex when recognised, releasing the RNA transcript through the C-Terminal Domain (CTD).

1.3 mRNA capping and polyadenylation

Recent studies highlight the fact that post-transcriptional modifications occur during transcription rather than afterwards (as indicated by the prefix “post-” which gives the wrong impression). The first mRNA modification is a 5' cap addition. In humans, adding a 5' cap is done by a capping enzyme and RNA guanine-7 methyl transferase (Furuichi & Shatkin, 2000). Basically, RNA Pol-II helps recruit both capping enzymes by its own phosphorylated CTD domain (Shuman, 1997).

This 5' cap consists of guanine nucleotide that is methylated on the carbon 7 of the guanosine base (m7G) and links to RNA via a triphosphate bridge. The importance of the m7G cap in mRNA stability and many other mRNA modifications comes from the cap-binding complex (CBC), which contains the nuclear CBP20 and CBP80 proteins that are replaced in the cytoplasm by eIF4F as seen in the figure 1.1 below (Gonatopoulos-Pournatzis & Cowling, 2014; Svitkin et al., 2001). Polyadenylation is essential for mRNA export from the nucleus. After transcript 3' end formation, certain factors interact with the mRNA 3' end to attach the poly(A) tail. In eukaryotes, almost all mRNAs have a poly(A) tail. The cleavage and polyadenylation specificity factor (CPSF) is an important recruited protein complex that identifies the conserved polyadenylation site (AAUAAA) and helps cleave the mRNA about 10 to 35 nucleotides downstream (Proudfoot & Brownlee, 1976; Lutz, 2008). The conserved

polyadenylation site (AAUAAA) is short and potentially found often within genes, sometimes resulting in alternative polyadenylation.

This can potentially lead to different mRNA isoforms. The polyadenylation site is determined by the RNA signal's strength and the AAUAAA sequence's proximity to the GU-rich region (Proudfoot, 1991). Also, the concentration of various regulatory factors such as cleavage stimulation factor (CSF) influences alternative polyadenylation (Colgan & Manley, 1997). However, some studies indicate that mRNA splicing of the last intron could be linked to polyadenylation, suggesting that splicing factors are involved in poly (A) site selection (Niwa et al., 1990; Wassarman & Steitz, 1993).

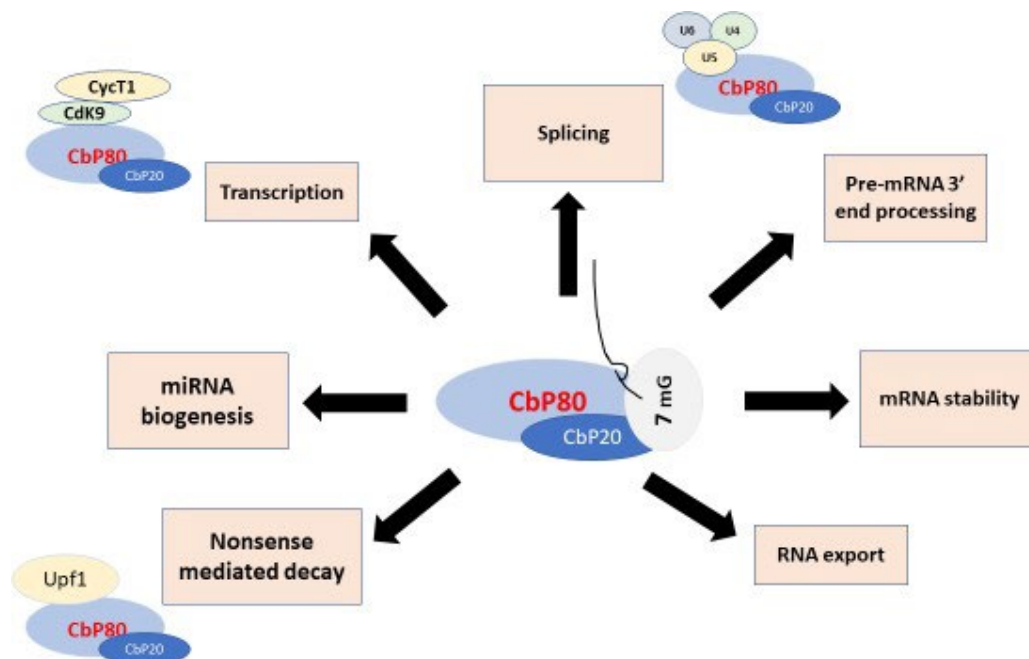


Figure 1. 1: Involvement of the Cap Binding Complex in RNA processing. The CBC consists of CbP80 and CbP20 protein subunits that bind to the 7-methylguanosine cap to facilitate associations with essential RNA processing mechanisms. Spliceosome snRNPs are U4, U5 and U6. CycT1 stands for cyclin T1, and Cdk9 stands for cyclin-dependent kinase 9. Image adapted from (Gonatopoulos-Pournatzis & Cowling, 2014).

1.4 Splicing of Pre-mRNA and alternative splicing

Splicing is the process of removing the non-coding regions (introns) and joining of the protein coding exons, to form a mature RNA (Aldalaqan et al., 2022). Alternative splicing has been identified as the process that allows limited gene sets to produce numerous proteins. Splicing requires the recognition of exon-intron junctions, named 5'- and 3'-splice sites. Introns typically have a GU at the 5' splice site and an AG at the 3' splice site. Failures of the splicing machinery to identify exon-intron junctions result in unproductive mRNAs which impact levels of proper gene expression (Wang et al., 2008). Various mRNA isoforms can be produced through a combination of one or more of five events: exon skipping, intron retention, mutually exclusive exons, alternative promoters, and the alternate recognition of 5' or 3' intron splice sites as shown in figure 1.2.

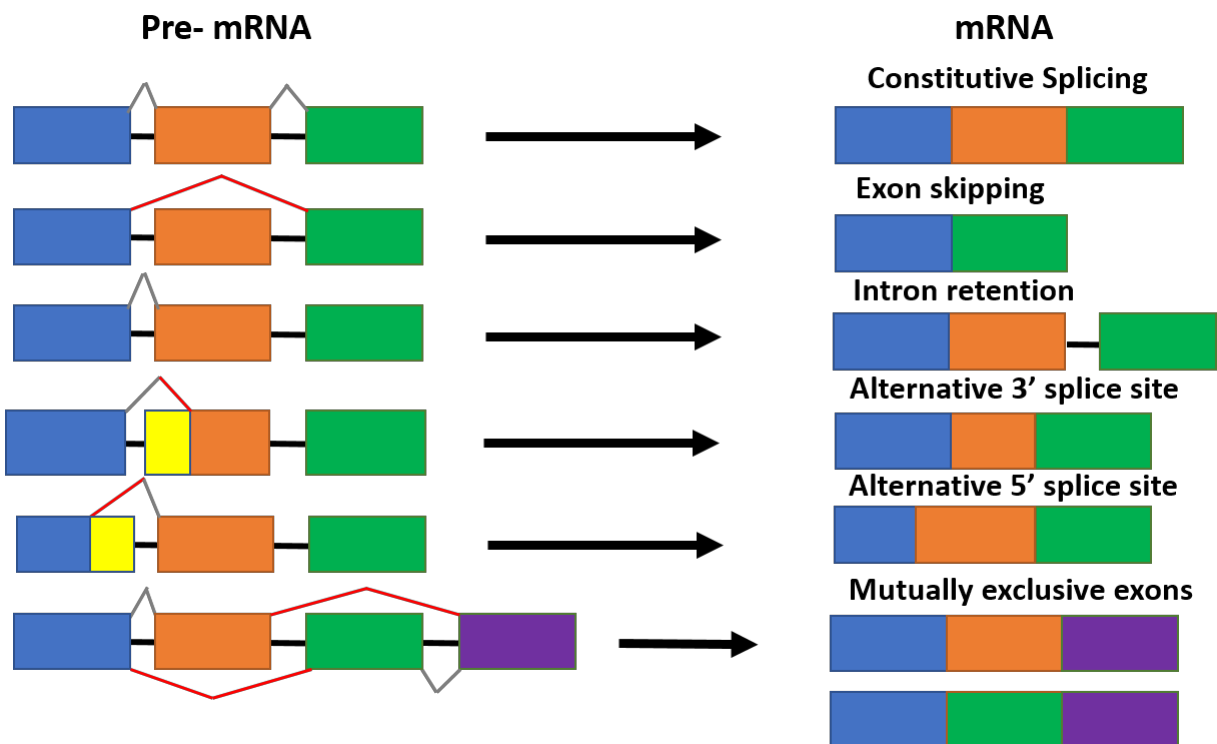


Figure 1. 2: A schematic showing different modes of normal alternative splicing events.

Exons are represented by boxes, while introns are represented by lines. Exon inclusion is represented by grey lines, while exon exclusion is characterised by red lines. Different pre-mRNA can be formed using exon skipping or intron retention, alternate 3' or 5' sites, and mutually exclusive exons.

1.5 The splicing reaction and spliceosome machinery

Splicing consists of two transesterification reactions as seen in the figure below 1.3 (Rogalska et al., 2022). These splicing transesterification reactions are: (1) the first reaction occurs via nucleophilic attack, where the phosphodiester bond of the 5' splice site is attacked by the branch-point adenosine nucleotide, leading to 5' splice site cleavage and formation of lariat. (2) The second reaction involves the 3' OH-group of the cleaved upstream exon attacking the 3' splice site. Resulting in the fusion of exons, and introns are freely formed a lariat. This lariat degrades and the snRNPs are recycled.

Spliceosome complexes recognise and excise introns from pre-mRNA. In humans, there are major and minor spliceosomes (Papasaikas et al., 2016). These consist of more than 150 proteins and several snRNPs.

The major spliceosome is made up of five small nuclear RNPs called U1, U2, U4, U5 and U6. Their function is mainly to recognise and then catalyse the removal of the U2-type introns as seen in the figure 1.4 below. The spliceosome cycle can be broken down into several stages. First, the U1 snRNP binds to the 5' splice site of the intron. After that, splicing factor 1 (SF1) recognizes the branch point, which is later displaced by the U2 snRNP. Then, the U2AF65 and U2AF35 proteins recognize the branch point and 3' splice site, forming the pre-spliceosome A complex. Next, the U4, U6, and U5 snRNPs interact with the A complex, forming the pre-catalytic spliceosome B complex. Next, U1 and U4 are dissociated, resulting in formation of the catalytically active spliceosome B complex, and then the C complex. Finally, ATPases, helicases, and the GTPase Snu114 assist in conformational changes during splicing. The minor spliceosome is responsible for the removal of the U12-type introns by utilising the snRNPs group known as: U4atac, U6atac, U11, U12 and U5 (Patel & Steitz, 2003; Wahl et al., 2009; Rogalska et al., 2022).

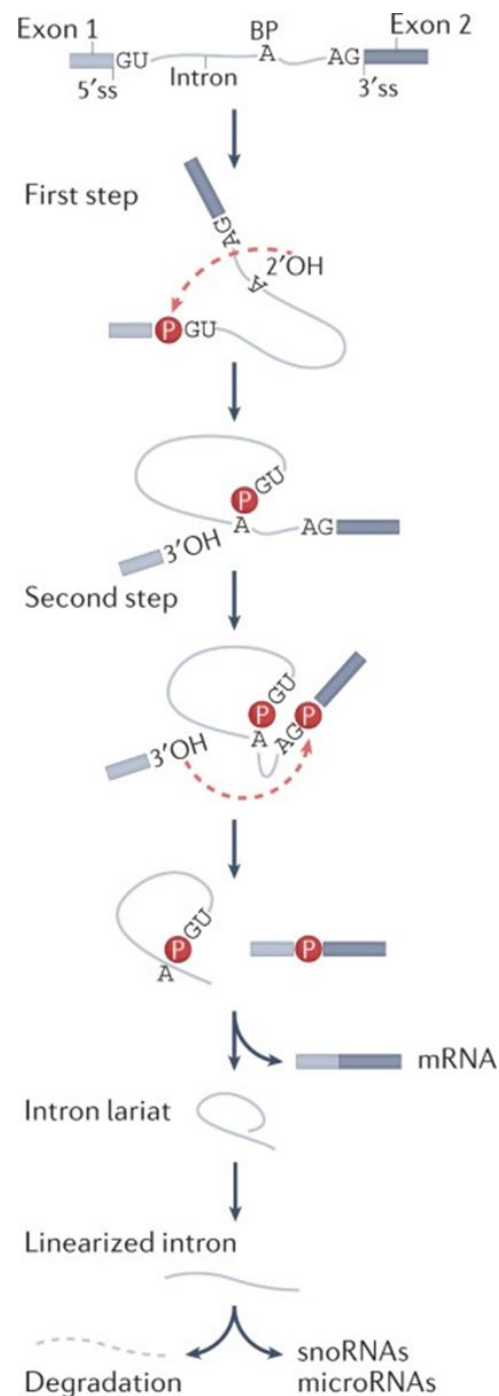


Figure 1. 3: Splicing comprises two transesterification reactions. The phosphate bond is represented by P in a red circle, a red dotted arrow shows the nucleophilic attacks. A lariat (looped RNA) structure is formed after the first attack as a result of a new phosphodiester bond. Introns are generally discarded after splicing, but can include other genes. Introns can often contain genes for noncoding RNAs such as snoRNAs and microRNAs. MicroRNAs (miRNAs) are tiny RNA molecules that oversee gene expression by attaching to messenger RNAs (mRNAs), affecting how they are translated or how stable they remain and managing a range of cellular functions. SnoRNAs are minuscule RNA molecules situated in the nucleolus, where they aid in the alteration and refinement of ribosomal RNA (rRNA) and other non-coding RNAs, guaranteeing their correct role in activities such as ribosome formation. Figure taken from (Rogalska et al., 2022).

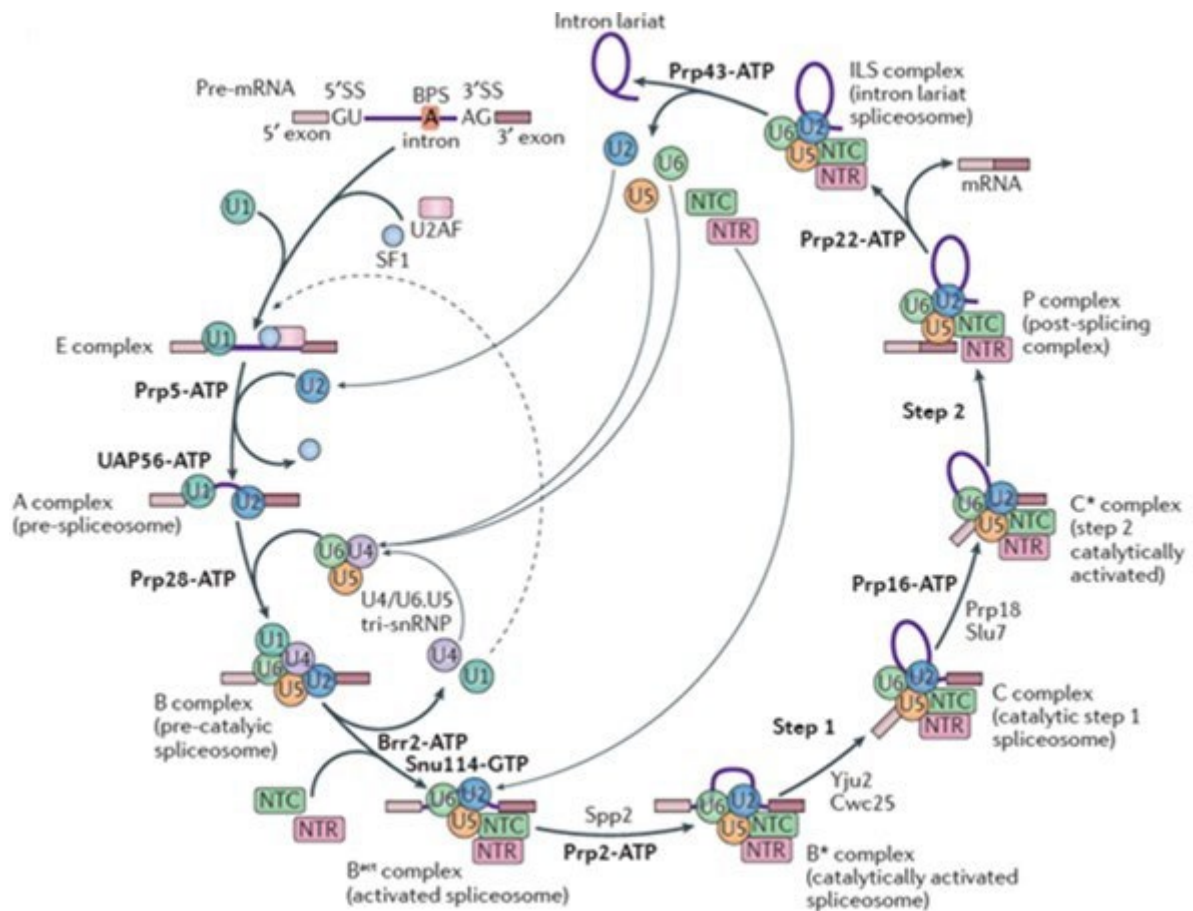


Figure 1. 4: A schematic of the Spliceosome cycle. Representing the interactions of the major snRNPs. Several steps are required to create and activate the spliceosome to complete RNA splicing. First, specific proteins called U1 small nuclear ribonucleoprotein (U1 snRNP), splicing factor 1 (SF1), and U2AF recognise particular RNA sites called 5' splice site (5'SS), branchpoint sequence (BP), and 3' splice site (3'SS). This recognition results in the formation of the E complex, the first spliceosome. After that, SF1 is replaced by U2 snRNP, forming the A complex, a pre-spliceosome. The A complex then combines with U4/U6 and U5 tri-snRNP to create the pre-catalytic spliceosome called the B complex. There are other spliceosome complexes including B act, B*, C, C*, P, and the intron lariat spliceosome complex (ILS), each with its unique composition. Transitions between these types are managed by special RNA-dependent ATPase/helicases (highlighted in bold), which help advance spliceosome assembly and the splicing reactions. Figure modified from (Sharp et al., 2017).

1.6 Recognition of splice site

Exons are categorised as either constitutive or alternative exons. This is related to whether they are always spliced (constitutive) or are sometimes alternative. The spliceosome machinery's recognition of the strength of the splice site signal controls alternative splicing. Such recognition is vital for correct splicing, as a strong splice site frequently flanks the constitutive exons compared to alternative exons. Cis-regulatory elements can regulate the recognition of splice sites. Cis-regulatory elements are sequences on the precursor mRNA that either enhance or inhibit splice site selection. These sequences are called exonic or intronic splicing enhancers and intronic or exonic splicing silencers. These splicing enhancer and silencer sequences regulate the identification of all constitutive and alternative exons (Matlin et al., 2005). Moreover, these cis-regulatory elements vastly exist near splice sites and within regulated exons (Sorek & Ast, 2003). However, enriched cis-regulatory elements have also been described far from splice sites (Ule et al., 2006).

1.7 Cryptic Splice sites

Spliceosomes are required to specifically identify splice sites which are short sequences at exon-intron junctions in order to join them together. There are additional short-length sequences known as “cryptic” because they share similarity with splice sites that are not usually selected by the spliceosome. These cryptic splice sites are known to be weakly recognised by the spliceosome machinery and usually be suppressed by nuclear RNA-binding proteins. In some cases they become activated by the loss of RNA-binding proteins (Aldalqaan et al., 2022). Cryptic splice sites can be found in non-annotated exons, introns, or larger exons. Figure 1.5 below shows examples of cryptic splice sites that may extend existing exons or introduce cryptic exons and exitrons. After analysing the mouse and human RNA sequencing (RNA-seq) data, almost half a million non-annotated splicing events were identified (Sibley et al., 2016). TDP43 is a nuclear RNA-binding protein that is known to suppress cryptic splice sites similar to RBMX family function (Aldalqaan et al., 2022).

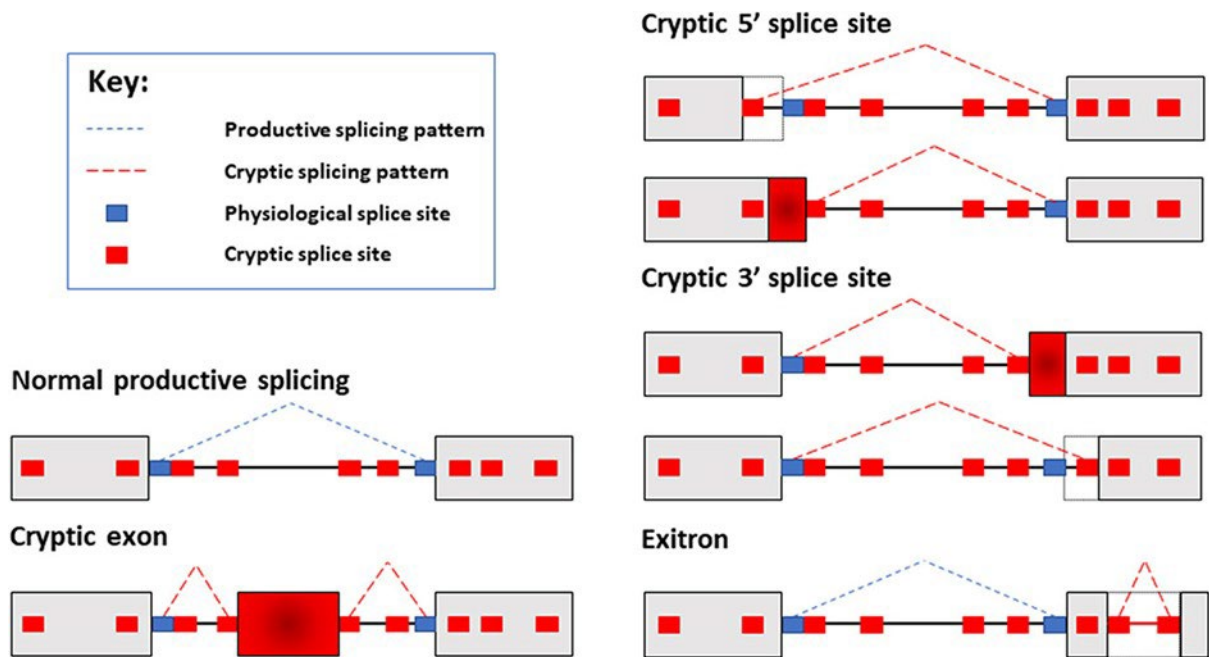


Figure 1. 5: A schematic showing some cryptic splice events such as the cryptic 5' and 3' splice site and the exitron. This diagram illustrates cryptic splicing patterns within genes, where most genes are made up of exons (grey boxes) separated by introns (connecting lines). The standard spliceosome process typically recognizes normal splice sites (blue boxes) and connects exons to form mature mRNA (dashed blue lines). However, cryptic splice sites (smaller red boxes), resembling normal ones, exist in both introns and exons. These cryptic sites are usually ignored but can mislead the spliceosome. If used, they lead to the production of different mRNA variants from genes. The conventional splicing pattern involves normal splice sites (broken blue lines), while cryptic splicing events, like cryptic exon inclusion within an intron, the use of cryptic 5' and 3' splice sites, or the mistaken recognition of cryptic splice sites within an exon (resulting in an exon being treated as an intron, known as an exitron), are depicted by dashed red lines. Figure taken from (Aldalaqan et al., 2022)

1.8 Male Germ Cell Development

Spermatogenesis is a highly prolific developmental pathway, with the human testis producing between 45 to 207 million sperm a day (Amann & Howards, 1980; Griswold, 2016). The testis contains both germ cells (which are in the developmental pathway leading to sperm) and supporting somatic cells (including Sertoli cells which support germ cell development, and Leydig cells which produce testosterone). These cells are organised within structures called seminiferous tubules. Located close to the basement membrane of seminiferous tubules are a population of germ cells called spermatogonia. Type A spermatogonia are stem cells. Approximately 0.1% of Type A spermatogonia undergo self-renewal in mice, whereas the remaining cells differentiate into type A paired (“Apr”) cells. Consistent with their name, type A paired cells are connected by intercellular bridges caused by incomplete cytokinesis after cell division. Further incomplete cell divisions result in the production of syncytia (joined cells) containing between 4 and 32 individual cells, referred to as A aligned (or “Aal”) cells. Both Apr and Aal cells are transit amplifying progenitor cells – by definition not yet differentiated, but also not part of the stem cell population. Spermatogonia then differentiate in response to retinoic acid stimulation (Griswold, 2016), which activates patterns of gene expression including *Stra8* (Kojima et al., 2019) and master regulators of gene expression such as the RNA binding protein DAZL (section 1.11.1.1) (Mikedis et al., 2020). Type B cells represent the last cell type in the mitotic stages of spermatogenesis, and their subsequent division results in the formation of preleptotene cells, in which chromosomes replicate and then enter meiosis (Legrand & Hobbs, 2018). Meiosis takes place within a group of cells called spermatocytes, and lasts for approximately 12 days in mice as seen in the figure below 1.6. In particular, meiotic prophase is much longer than mitotic prophase – corresponding to a difference between days in the mouse for meiotic prophase, compared with minutes for mitosis.

During meiotic prophase chromosomes initially condense (during a substage called leptotene). One of the key features of meiotic prophase is genetic recombination to produce new chromosome variants. During zygotene a protein complex called the synaptonemal complex forms between homologous chromosomes that start to form cross overs. Cells in the next meiotic substage, pachytene, contain shortened and thickened homologous chromosomes that are undergoing genetic recombination and are the largest of the spermatocytes. During the next substage (diplotene) homologous chromosomes are starting to separate, and this separation is complete by diakinesis. Immediately following the first meiotic division is a

second meiotic division that independently assorts individual sister chromatids into 4 haploid gametes called spermatids (Griswold, 2016). This is followed by the final stage of germ cell development (called spermiogenesis), which converts haploid cells into motile gametes capable of fertilization. Spermatogenesis takes 14-days in mice and comprises 16 steps, during which round spermatid cells differentiate into elongated spermatids, and eventually into sperm.

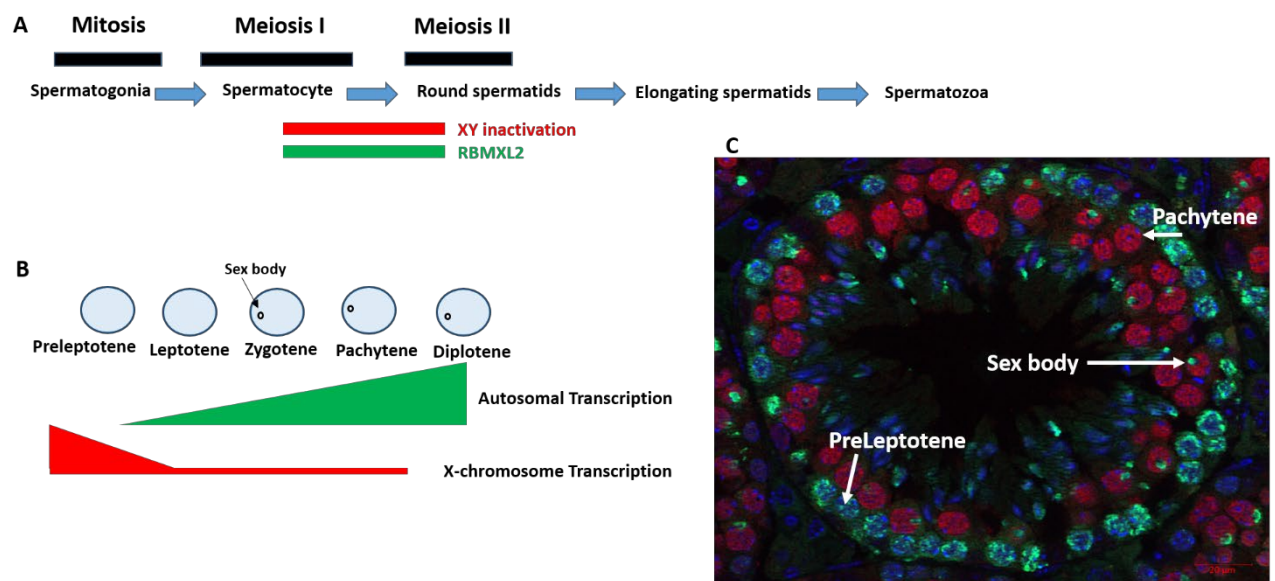


Figure 1. 6: Mouse spermatogenesis. A) Spermatogonia produces sperm through mitosis and two stages of meiosis. B) Meiosis phases displaying XY inactivation. C) Fluorescent staining of wild-type mouse testis (Sv/129), stage X, RBMXL2 protein is represented in red, and γ H2AX is represented in green. The testis staining was also used in my results in Figure (3-14). Figure taken from (Aldalaqan et al., 2022)

1.9 RNA binding proteins

RNA binding proteins (RBPs) are a group of more than 1914 proteins that regulate all aspects of RNA biology, including transcription, processing, localisation as well as nonsense-mediated decay (NMD). RNA-binding proteins (RBPs) encompass a combination of proteins featuring distinct domains (for instance, the RRM) and those that have been described by recent function-based high-throughput studies (Qin et al., 2020). Because of this well-controlled spatiotemporal interaction, gene expression can be precisely modulated. The lifecycle of each mRNA strand can be affected by the combination of RBPs it encounters. Some RBPs alter mRNA through alternative splicing or polyadenylation. RBPs can also act as positive or negative modulators of splicing depending on their binding position relative to an exon or poly-A site. RBPs play roles in the control of post-transcriptional events, such as the localisation and translation of mRNA and the maintenance of mRNA stability (Licatalosi, 2016).

The interactions between RBPs and RNA are accomplished by RNA-binding domains (RBDs), which are known specific regions within the proteins. RBPs are classified based on their RNA-binding domains into four main groups: RNA recognition motif (RRM), the K-homology domain, zinc finger (ZnF) domain, and the double-stranded RNA-binding domain ((ds-RBD)). However, approximately 50% of all RBPs have no known RNA-binding domains (Garcia-Moreno et al., 2018). Moreover, some RBPs have more than one RBDs with linker regions that control the affinity and specificity of the binding RNA, leading them to diversify their functions (Lunde et al., 2007). RBPs also contain intrinsically disordered regions that are repeats of amino acids with hydrophobic residues that sustain the dynamic disordered structures allowing the non-specific RNA binding, mediate interactions with other proteins, and regulate the localisation of RBPs (De Conti et al., 2017).

1.10 RNA binding proteins (RBPs) throughout spermatogenesis

Of all tissues in the body, the testes and the brain have the highest gene expression profile and experience the highest levels of alternative splicing events (Gamble et al., 2020). However, the exact mechanisms that regulate the alternative splicing of mRNA and exert post-transcriptional control during spermatogenesis have yet to be determined.

Schmid et al. (2013) identified significant changes in alternative splicing throughout the transition from mitotic to meiotic phases during germ cell development. In addition, their work revealed changes in mRNA splice isoforms between pre- and post-meiotic stages. The meiosis stage lasts for approximately 12 days in the mouse testis. During this time, they tracked the expression of different RNA splicing regulators, such as PTBP1 (Polypyrimidine tract-binding protein 1) and RBMXL2, among others (Schmid et al., 2013). PTBP1 was found to be downregulated during meiosis, likely replaced by the expression of PTBP2 (Hannigan et al., 2017). In the testis, some alternative splicing events are conserved between mice and humans; therefore, these events are thought to control important aspects of germ-cell development (Schmid et al., 2013). However, many splicing events between mice and humans are not conserved (Kan et al., 2005).

Altered mRNPs in male germ cells are believed to play roles in the modulation of mRNA stability, translation, and decay. RNA helicases are believed to restructure protein-RNA complexes by promoting or removing certain RBPs, and RNA helicases are localised to chromatoid bodies, which is a cytoplasmic structure within male germ cells that plays a vital role in processing and regulating RNA molecules (Licatalosi, 2016).

Several RBPs are involved in the regulation of post-transcriptional machinery in male germ cells, and can be categorised as either nuclear or cytoplasmic, based on their functions and localisation. Some important RBPs in spermatogenesis are shown in table 1.1 and described below.

1.10.1 Nuclear RNA binding proteins important in spermatogenesis.

Nuclear RBPs mediate nuclear gene expression and are involved in regulating the alternative splicing of pre-mRNA and polyadenylation as well as preventing cryptic splicing and cryptic polyadenylation (Licatalosi, 2016). A high number of alternative splicing and polyadenylation events occur in the mouse testis, which involve approximately 63% of the mouse genes (Song et al., 2020). Nuclear-RBPs can act as either enhancers or repressors of splicing events, depending on which region of the skipped exon they bind.

1.10.2 The RBMX family of RNA binding proteins

RBMX, RBMY, and hnRNP G-T (RBMXL2) are a family of RNA binding proteins as seen in figure 1.7, containing a conserved (RRM) recognition binding motif that mediates RNA binding as Figure 1.8 below and is involved in RNA processing and the alternative splicing of pre-mRNA transcripts. Moreover, in vitro, this family of proteins has been found to interact and control the splicing activity of SR (serine- arginine rich) proteins and Tra2 β (Ehrmann et al., 2019). *RBMX* and *RBMY* (located on the X and Y chromosomes respectively) are ancient genes which evolved 300 million years ago (Lahn et al., 1997). The RBMX family is essential to this thesis; therefore, the three family members are described in more detail below.

1.10.2.1 RBMX

RBMX is a homologue of *RBMY* that is found on the X chromosome and is ubiquitously expressed. First identified as heterogeneous nuclear ribonucleoprotein G (hnRNP G), *RBMX* was mapped to chromosome X in both humans and mice (Delbridge et al., 1999). RBMX protein is involved in splicing control and genome integrity (Elliott et al., 2019). RBMX protein is essential for proper development and lacking RBMX has been linked to cancer. RBMX family proteins are involved in splicing regulation, transcription, and genome integrity at the molecular level. RBMX is found in many tissues, including the brain and testis, and its involvement in brain development is particularly significant. Many studies show that RBMX is required for neural network creation, axon and dendrite development and gene expression regulation in the central nervous system. Furthermore, RBMX regulates splicing, altering the splicing of genes associated with brain development and neurological diseases. RBMX's importance goes beyond brain development. It has been linked to chromosomal biology, including kinetochore function and chromatid cohesion during cell division.

RBMX has been associated with cancer, where its expression can either suppress or promote tumour development. RBMX also plays a role in DNA repair pathways, impacting genome stability. RBMX interacts with the Borna Disease Virus in infectious diseases and contributes to viral replication. Chileleko Siachisumo from David Elliott's lab showed that RBMX functions as a splicing suppressor in breast cancer cell lines (Siachisumo et al., 2023). RBMX has been found to suppress cryptic splice sites within genes involved in genome stability, such as ETAA-1 and REV3L (Elliott et al., 2019).

1.10.2.2 RBMY

RBMY is a paralog of *RBMX* that is specifically expressed in male germ cells, indicating that *RBMY* may have evolved into a testes-specific gene involved in meiosis and spermatogenesis (Delbridge et al., 1999). In addition, several identical copies of the *RBMY* gene, including *RBMY1A1*, *RBMY1B*, and *RBMY1E*, are located on the human Y chromosome (Elliott et al., 2019). Moreover, one of the first identified genetic causes of male infertility was a microdeletion of the Y chromosome (Ma et al., 1993) and micro-deletions of the Y chromosome, including deletions that remove *RBMY*, have been linked to infertility in men, and associated with clear phenotypes, including azoospermia and oligospermia (Elliott, 2000). Lahn (1997) confirmed that 98% of ancestral genes are conserved on the X chromosome, whereas the Y chromosome retains only 3% of ancestral genes (Lahn et al., 1997), as many genes have been lost from chromosome Y; thus, the survival of *RBMY* on the Y chromosome indicates its likely importance (McCarrey & Thomas, 1987).

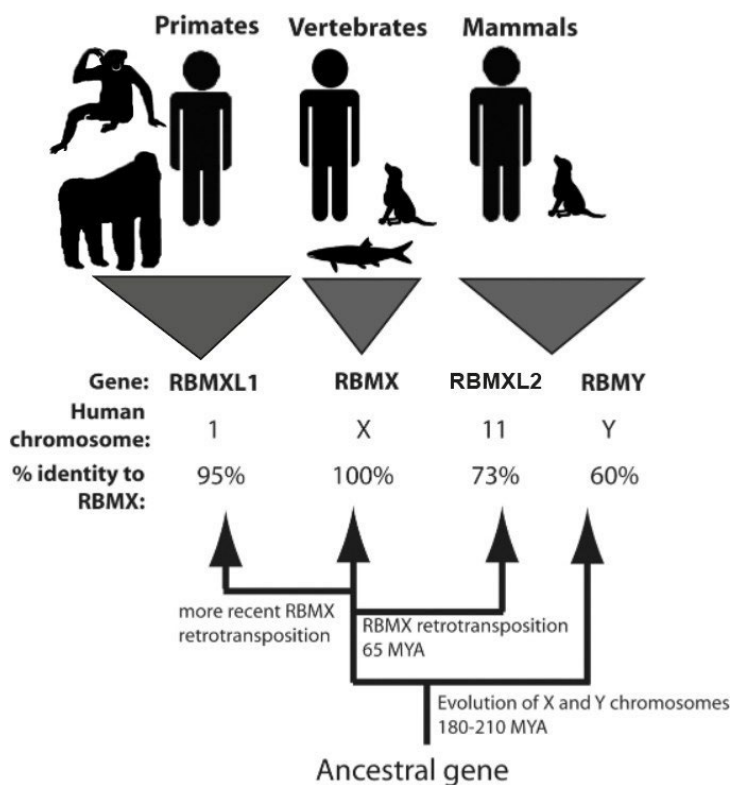


Figure 1. 7: the origin of the RBMX family of genes. *RBMY*, a paralog of *RBMX*, shares 60% of *RBMX*'s DNA. Similarity in sequence between *RBMX* and *RBMXL1* is 95%, and between *RBMX* and *RBMXL2* (hnRNPGT) is 73%. The cladogram was taken from (Elliott et al., 2019).

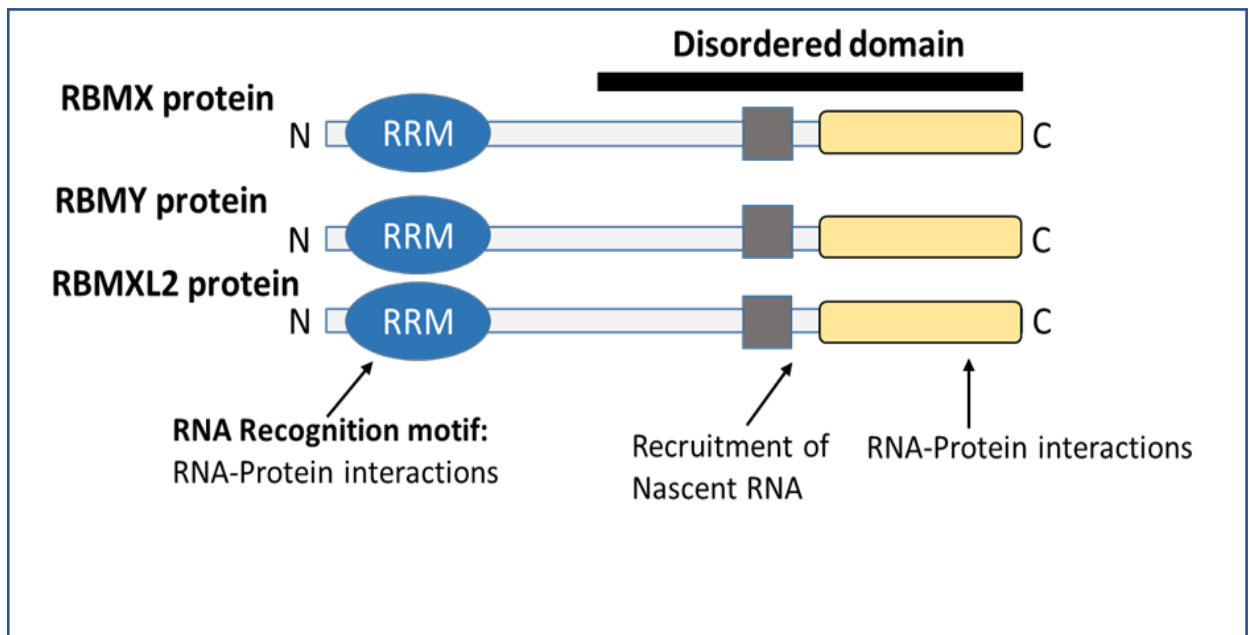


Figure 1. 8: The protein structures of the RBMX family. In blue the RRM (RNA recognition motif) that binds to RNA. A disordered domains contains conserved regions in grey, is used to recruit nascent RNA and C-terminus RNA-protein interaction domain is shown in yellow. This figure was adapted from (Elliott et al., 2019).

1.10.2.3 RBMXL2

RBMXL2 (also called Heterogeneous nuclear ribonucleoproteins G-testis, or HNRNPGT), is a retro- transposed copy of *RBMX* that evolved approximately 65 million years ago. In humans, *RBMXL2* is located on chromosome 11 (Venables et al., 2000; Elliott et al., 2000; Ehrmann et al., 2019). *RBMXL2* protein shares 73 % similarity with *RBMX* protein, suggesting that *RBMXL2* developed from *RBMX*. *RBMXL2* is a testis-specific nuclear RNA binding protein expressed within spermatocytes and round spermatids, suggesting that this protein may function as a substitute for *RBMX* when the X and Y chromosomes are inactivated during the pachytene stage of meiosis (Elliott et al., 2019). *RBMXL2* may either act to replace *RBMX* while the chromosome X is inactive or have a specific and independent function during meiosis (Elliott, 2004). Deleting one copy of *Rbmxl2* gene resulted in abnormal sperm production in mice (Ehrmann et al., 2008). Also, the study of Westerveld *et al.*, (2004) discovered a point mutation in the human *RBMXL2* gene resulting in infertility in men (Westerveld et al., 2004).

However, the role played by RBMXL2 during spermatogenesis remains unclear. Another study identified a frameshift mutation in *RBMXL2* that led to meiotic arrest and sterility (Ghieh et al., 2022). Recently, a study reported by Ehrmann *et al.*, (2019) showed that the deletion of *Rbmxl2* causes infertility in mice, and indicated that RBMXL2 plays an important role in splicing regulation, including the suppression of cryptic splice sites that would otherwise disrupt gene expression. RNA-seq analysis of 18-day old wild type and *Rbmxl2* knockout mice showed that RBMXL2 protein prevents the use of weak splice sites during meiosis. Use of these weak splice sites in the *Rbmxl2* knockout mouse causes accumulation of aberrantly spliced mRNA isoforms, including the inclusion of cryptic exons (which are not normally spliced). One high amplitude cryptic exon that was activated without RBMXL2 protein was within intron 1 of the *Kdm4d* gene. The evidence suggested that this *Kdm4d* cryptic exon is likely to be directly repressed by RBMXL2 protein, as cryptic exon splicing could be repressed when expressed from a minigene after co-transfection of a RBMXL2-GFP fusion protein. Loss of RBMXL2 protein also caused production of novel protein isoforms. *Rbmxl2* knockout mice have aberrant splicing of exon 5 of the *Meioc* gene, which was converted into an exitron (splice sites within a long exon can introduce a new intron called an exitron), resulting in the production of a shorter *Meioc* protein isoform. The splicing defects in *Rbmxl2* knockout mice cause meiotic arrest and infertility by preventing the expression of key proteins needed in meiotic prophase. The UV-crosslinking immunoprecipitation technique (CLIP) can aid the identification of RBPs associated with cryptic splice sites (König et al., 2012), and HITS-CLIP identified some RBMXL2 binding sites in *Meioc* among others (Ehrmann et al., 2019).

RNA-binding protein	Motif	Function	Expression	Exp. In the Germ cells	Knockout phenotype
Tra2β	RRM, arginine-serine (RS) domains	Splicing factor ¹	Nucleus	Gonocytes to elongated spermatids.	Spermatogenic arrest sterile.
hnRNPA-1	2 RRM	Splicing and transcription factor ²	Nucleus	Spermatogonia	Unknown
ADAD-1	RRM and an AD-domain	Splicing factor ³	Nucleus	Round and elongated spermatid.	Faulty sperm morphology, sterile.
r-Cstf64	RRM	Polyadenylation Factor ⁴	Nucleus	Spermatocytes and early spermatids.	Disrupted meiotic and post-meiotic germ cells. Sterile
RBMXL2 (HNRNPG-T)	RRMs	Splicing Factor ⁵	Nucleus	pachytene spermatocytes and round spermatids	Spermatogenic arrest sterile.
DAZ-1	2 RRM and DAZ repeats.	Translational factor ⁶	Nucleus and Cytoplasm	Spermatogonia and early spermatocytes.	Oligozoospermia, Infertile
YBX2 (MSY2)	Y-box DNA-binding motif	Translational and transcription factor ⁷	Cytoplasm	Round spermatids.	Elongated spermatid failure. Sterile.
NANOS2	2 zinc finger domain-like cysteine-histidine (C/H domain)	Translational Factor ⁸	Cytoplasm	Gonocytes and Spermatogonia	Loss of spermatogonia, Sterile.
PUMILIO (PUM1)	a sequence specific RRM	Translational Factor ^{9,10}	Cytoplasm	Spermatocytes and spermatids.	spermatogenic defects (oligozoospermia), Although fertile.
PUMILIO (PUM2)	a sequence specific RRM	Translational Factor ^{9,10}	Cytoplasm	Gonocytes and spermatogonia	Fertile
CPEB	2RRMs, a zinc finger domain-like cysteine-histidine (C/H domain)	Polyadenylation and Translational Factor ^{11,12}	Cytoplasm	Early spermatocytes	Spermatogenic arrest sterile.
Sam68(khdrbs1)	KH-Domain	Splicing, transcription, and Translational factor ^{13,14}	Nucleus and Cytoplasm	Spermatocytes and round spermatids.	Defects in Spermatogenesis, sterile.
CELF-1 (CUG-BP)	3 RRM	Splicing and Translational factor ¹⁵	Nucleus and Cytoplasm	spermatocytes	Defective elongation, sterile.
HuR (ELAVL1)	Several RRM	Translational and RNA editing Factor ¹⁶	Nucleus and Cytoplasm	Spermatocytes and early spermatids.	Spermatogenic failure, Azoospermia.
Ptbp1 and 2	4 RRM	Splicing, mRNA stability and translational factor ^{17,18}	Nucleus and Cytoplasm	Spermatogonia and spermatocytes.	Spermatogenic arrest sterile.
RBM-5	2RRM, and 2 zinc-finger (ZF) domains	Splicing and mRNA stability factor ¹⁹	Nucleus and Cytoplasm	Spermatogonia and round spermatid.	Spermatid arrest, Sterile
DDX-5	DEAD-box helicase domain	Splicing and transcription factor ²⁰	Nucleus and Cytoplasm	spermatogonia	Lost spermatogonia, sterile.

Table 1.1: Examples of some important RBPs and their functions in spermatogenesis.
References are included in superscript, and details given in appendix 2 of this thesis.

1.10.3 Other nuclear RNA binding proteins important in spermatogenesis:

Other nuclear RNA-binding proteins have been identified as essential or potentially essential for spermatogenesis, where they may play a role in nuclear RNA processing steps such as splicing and polyadenylation.

1.10.3.1 Transformer-2 (Tra-2 α and Tra-2 β)

Tra-2 β protein has one RNA recognition motif (RRM) and two RS domains (N-terminal and C-terminal). In humans, there are two homologues of Tra-2, called Tra-2 α and Tra-2 β , both of which specifically bind to GAA-repeats in RNA (Venables & Eperon, 1999; Sciabica & Hertel, 2006). Moreover, both Tra-2 α and Tra-2 β proteins function as splicing activators (Tacke et al., 1998). The two proteins have different expression profiles, and Tra-2 α is less studied than Tra-2 β . Tra-2 β contains 10 exons that produce five isoforms by alternative splicing (Tra-2 β 1 to Tra-2 β 5) (Stoilov et al., 2004; Kuwano et al., 2015).

Tra-2 β is highly phosphorylated, and this phosphorylation influences its localisation and activity. Furthermore, phosphorylation increases the binding between Tra-2 β and RBMY (Venables et al., 2000). Best and colleagues (2014) reported that Tra-2 β auto-regulates itself via a poison exon (this is a negative feedback loop that controls exon splicing) (Best et al., 2014). Tra-2 β interacts with the testes-specific proteins RBMXL2 and RBMY, suggesting the importance of Tra-2 β in male gametogenesis (Venables et al., 2000; Stoilov et al., 2004; Grellscheid et al., 2011). Knockout of *Tra-2 β* in mice resulted in embryonic lethality (Mende et al., 2010). Therefore, Tra-2 β 's involvement in the regulation of alternative splicing events in mouse spermatogenesis is not yet reported in the literature but possibly important. Analysis of a conditional allele in mouse spermatogenesis indicates an important role for Tra-2 β (Caroline Dalgliesh, Newcastle University, personal communication).

1.10.3.2 Sam68 (KHDRBS1)

Src-associated in mitosis (sam68) is a part of the signal transduction and activation of RNA (STAR) family. Sam68 protein is 68 kDa and binds RNA through a K homology (KH) domain. Sam68 is expressed in all cell types (Ehrmann et al., 2013). Sam68 plays roles in transcription, splicing, translation, signal transduction, and RNA export (Huot et al., 2012; Licatalosi, 2016). Male *Sam68* knockout mice were infertile and showed many defects, including an increased number of apoptotic cells and abnormal/immotile sperms (Licatalosi, 2016). Moreover, the up- and downregulation of up to 300 mRNA transcripts was observed in *Sam68* knockout testis, and detection of abnormal alternative splicing (Licatalosi, 2016).

1.10.3.3 PTB protein family

The polypyrimidine tract binding (PTB) family includes PTBP1, PTBP2 and PTBP3 (Licatalosi, 2016). PTB proteins are important for the regulation of gene expression (Zagore et al., 2015). Moreover, PTB proteins likely play crucial roles in germ cell development (Licatalosi, 2016). PTBP1 is expressed by almost all cell types, unlike PTBP2 which is a cell type-specific protein and expressed primarily in the testis and brain. PTBP2 plays a crucial role in the regulation of exon splicing and neurogenesis and is thought to be critical for postnatal survival (Licatalosi, 2016; Senoo et al., 2019). In the mouse testis, PTBP1 was found to be expressed in spermatogonia, whereas PTBP2 was found in spermatocytes and round spermatids. PTBP2 is a strong mediator of alternative splicing and acts in a stage-specific manner to regulate alternative splicing (Song et al., 2020). During spermatogenesis, PTBP2 was found to bind to the 3' splice site of cassette exons in some important genes. Additionally, the depletion of PTBP2 leads to the generation of several mis-spliced isoforms of these genes (Hannigan et al., 2017).

1.10.3.4 RNA binding motif-5 (RBM-5)

RBM-5 is an RBP that is ubiquitously expressed in all tissues. RBM-5 contains two RRM and is found in the nucleus and cytoplasm of cells. RBM-5 is known to be a splicing factor and to be evolutionarily conserved across different species (Legrand & Hobbs, 2018). RBM-5 is highly expressed in mouse testes, particularly in spermatocytes and round spermatids. A point mutation identified in the second RNA recognition motif of *RBM-5* causes loss of function. Male mice with this mutation presented with azoospermia, with spermatogenic arrest observed at the round spermatid stage (Jamsai et al., 2017). Finally, a point mutation (R263P) in *RBM-5* that compromised pre-mRNA splicing, resulting in the generation of abnormally spliced transcripts, spermatogenic arrest, and male sterility (Song et al., 2020). A recent study identified 6 men with infertility caused by a de novo missense mutation in *RBM-5* (Oud et al., 2022).

1.10.3.5 DEAD-box helicase-5 (DDX-5)

DDX-5 is part of the DEAD-box protein family. This family of proteins is known to have a conserved DEAD (Asp-Glu-Ala-Asp) motif inside the helicase core associated with ATP-hydrolysis. Other motifs across the DEAD-box family allow these proteins to bind to DNA as transcriptional factors or to play a role in RNA processing (Legrand et al., 2019; Sone et al., 2020). Knock-out of the *DDX-5* gene in male mice led to the loss of spermatogonia. However, the exact role of the DDX-5 protein is still unknown (Sone et al., 2020). Recently, Legrand and colleagues (2019) reported that *DDX-5* is essential for spermatogenesis maintenance in mice (Legrand et al., 2019).

1.10.3.6 Polyadenylation Factors (Testis cleavage stimulation factor-2 τ)

During spermatogenesis, alternative polyadenylation (APA) can also regulate the expression of genes in the testis. APA can significantly impact gene expression, which was found to be frequently occurs in a tissue- or developmental stage-specific manner. APA permits testis-specific genes to produce multiple protein isoforms from a single gene (Liu et al., 2007) as well as permitting a change in expression levels of a protein (Mittleman et al., 2020). In the absence of the polyadenylation protein *Cstf64*, spermatogenesis is impaired, resulting in male sterility (Dass et al., 2007).

Testis-Cstf64 is an RBP that generated from the cleavage stimulation factor-2 (*CSTF2*) gene, thought to be retro-transposed from *CstF64* mRNA that is essential for the cleavage and polyadenylation of the 3' end of mRNA. Testis-Cstf64 is a paralogue of *Cstf64*, which is conserved in all mammalian species, including humans and mice, and is known to be expressed primarily in the testes and brain, although it can also be found at varying levels in many other tissues, such as the heart and liver (Licatalosi, 2016; MacDonald, 2019).

Cstf64 is localised on the X chromosome, and during meiosis, the inactivation of the X-chromosome results in the loss of *Cstf64* expression in early spermatocytes. The expression of testis-Cstf64 protein has been detected in pachytene spermatocytes and remains in early spermatids (Licatalosi, 2016). The knockout of *testis-Cstf64* results in infertile male mice due to a lack of post-meiotic germ cells, including reduced numbers of spermatocytes and elongated spermatids. These abnormalities were associated with the polyadenylation of mRNA and the alternative splicing of some critical genes in spermatogenic cells such as testis-specific histone genes (*H1fnt*, *Hils1*, *H2afb1*) among others (Dass et al., 2007; Harris et al., 2016; Grozdanov et al., 2018).

1.11 Examples of important Cytoplasmic-RBPs in spermatogenesis

Cytoplasmic RBPs can bind sequences in mature mRNA found in the cytoplasm. Cytoplasmic RBPs have the potential to affect mRNA localisation, stability and translation (Zagore et al., 2018), mRNA transport (Sutherland et al., 2015). However, some RBPs are thought to be involved in various nuclear and cytoplasmic processes (Sutherland et al., 2015). Below is a description of some important cytoplasmic RBPs involved in spermatogenesis.

1.11.1 Translational Factors important during spermatogenesis

1.11.1.1 DAZ protein family

DAZ-like (DAZL) protein is expressed specifically by vertebrate germs cells and is known as a germ cell- specific RBP. *DAZL* is an autosomal homolog of *DAZ*, which is found in primates and encoded by *DAZ* genes on the Y chromosome (Licatalosi, 2016). *DAZL* belongs to the *DAZ* family, which is comprised of 3 genes, *BOULE*, *DAZL*, and *DAZ*. The *DAZ* family is located in the distal euchromatic segment of the long arm of the human Y chromosome (Yq11.23), in the azoospermia factor c (AZFc) region (Fernandes et al., 2002; Collier et al., 2005). The *DAZ* family functions as a translational regulators in germ cell development, interacting with poly(A) binding proteins (PABPs), which are also required for translational initiation (Collier et al., 2005).

DAZL and BOULE, are commonly expressed in all vertebrates (Paronetto & Sette, 2010; Smith et al., 2011). DAZ family proteins contain one RRM and at least one DAZ repeat, comprised of 24 amino acids, including Asn, Tyr, and Gln. The RRM motif is responsible for binding mRNA and is highly conserved. There is up to 85% sequence similarity of *Dazl* between mouse and human. (Fu et al., 2015).

DAZ and DAZL are essential during spermatogenesis and are expressed by spermatogonia, (Paronetto & Sette, 2010). In humans, DAZ is expressed in spermatogonia, spermatocytes, late spermatids, and sperm tails (Fernandes et al., 2002). In spermatogonia, these proteins are expressed in the cell nucleus. The production DAZL continues into the spermatocyte stage, at which point the expressed proteins are translocated from the cell nucleus to the cytoplasm (Paronetto & Sette, 2010). The partial deletion of DAZ (approximately 90%–95%) in males can lead to spermatogenic failure (a condition characterised by the absence or reduced production of spermatozoa), and the complete elimination of *DAZ* can result in oligozoospermia (low sperm count), both of which are common causes of male infertility (Licatalosi, 2016). A study examining the knockout of *Dazl* in mice revealed complete infertility in both male and female mice (Paronetto & Sette, 2010). Thus, both genetic and epigenetic mutations in *DAZ genes* and *DAZL* can lead to infertility (Li et al., 2019). Additionally, the regulation of mRNA translation by *DAZL* is essential for germ cell differentiation and, thus, fertility (Paronetto & Sette, 2010).

1.11.1.2 Y-Box proteins

The Y-box proteins are conserved testis-specific proteins. In mouse testes, three Y-box genes have been identified. These are expressed at similar levels: YBX1, YBX2, and YBX3 (Licatalosi, 2016). YBX2 and YBX3 are primarily expressed in meiotic and round spermatids, indicating their importance for spermatogenesis. These proteins are also complexed with more than 75% polyadenylated RNA in the testis (Licatalosi, 2016). YBX2 has also been found to be expressed during the pachytene stage and in round spermatids. Another study found that YBX2 binds to the 3'UTR of mRNAs in elongated spermatids, and the absence of YBX2 results in loss of translational repression leading to sterility (Snyder et al., 2015; Jan et al., 2017). Those studies identified this protein functions as a repressor of translation on specific mRNAs. YBX2 is localised primarily in the cytoplasm, where it is thought to play roles in cytoplasmic mRNA storage, stabilisation, and translation suppression (Gatta et al., 2010; Sutherland et al., 2015). *Ybx2* knockout leads to spermatogenic arrest, with no spermatozoa formation (Legrand & Hobbs, 2018). Snyder et al. (2015) reported that the deletion of *Ybx3* in mice resulted in low to no sperm production, suggesting the interruption of spermatogenesis (Snyder et al., 2015).

1.11.1.3 Pumilio and Nanos

The Pumilio proteins (PUM) belong to the Pumilio and fem-3 binding-factor (FBF) family of RNA- binding proteins (RBPs). These proteins regulate translational repression and mRNA nonsense- mediated decay (Chen et al., 2012; Licatalosi, 2016) . In mammals (mice and humans), two PUM proteins have been identified, named PUM1 and PUM2. These proteins localise to the cytoplasm and have expression pattern in multiple tissues. Moreover, these two proteins are highly conserved, with 76% similarity. *Pum2*, unlike *Pum1*, is not essential for spermatogenesis in male mice, and fertility was not compromised in animals harbouring a homozygous mutation in *Pum2*. In contrast, the loss of *Pum1* expression in mice revealed spermatogenic defects, including a high number of apoptotic spermatocytes and low sperm counts (Licatalosi, 2016; Goldstrohm et al., 2018). Nanos proteins are conserved, translational regulatory factors that bind to PUM and recruit specific mRNAs. Moreover, by recruiting the deadenylation complex, Nanos proteins can downregulate their mRNA targets. Transgenic mouse strains revealed essential roles for *Nanos2* and *Nanos3*, the two mouse *Nanos* genes involved in mouse spermatogenesis. The loss of *Nanos3* results in a decrease in the number of primordial germ cells (PGCs) before they are completely lost. Defects in spermatogenic cells were also observed following the depletion of *Nanos2* in mice. *Nanos2* plays a vital role in maintaining the balance between self-renewal and differentiation in spermatogonia. The post-natal loss of *Nanos2* causes reduced numbers of undifferentiated spermatogonia, whereas the accumulation of undifferentiated spermatogonia was observed following increased *Nanos2* expression (Licatalosi, 2016; Legrand & Hobbs, 2018).

1.11.1.4 Cytoplasmic polyadenylation element-binding protein (CPEB)

Cytoplasmic polyadenylation element-binding protein (CPEB) is a conserved RBP that binds to the 3'U-rich sequence of cytoplasmic polyadenylation elements (CPEs) and can be found in all multicellular eukaryotes. After transcription, CPEB proteins play multiple roles in gene regulation. For example, they promote cytoplasmic polyadenylation, which lengthens specific mRNA molecules' poly (A) tail. Furthermore, they influence mRNA stability, help with mRNA transport and localization, and control mRNA translation. These processes enable precise gene expression regulation and support various cellular functions and developmental processes (Huang et al., 2023). CPEB has two RNA recognition motifs (RRMs) and a zinc finger domain-like cysteine-histidine repeat sequence (C/H domain), which are both critical for RNA binding. Vertebrates regularly express four CPEB paralogues (Licatalosi, 2016; Singh et al., 2017). The loss of one *Cpeb* gene (*Cpeb-1*) in mice leads to a spermatogenic arrest, indicating the importance of *Cpeb* during mouse spermatogenesis. Moreover, other *Cpeb* paralogues show overlapping and varying levels of expression during mouse spermatogenesis, indicating the added complexity of *Cpeb*-associated regulation due to specific isoforms (Tay & Richter, 2001; Licatalosi, 2016).

1.12 Hypotheses and project aims.

1.12.1 Background to this thesis

This thesis is based on the work of Dr. Ingrid Ehrmann in her eLife paper (2019), where she knocked out *Rbmxl2* in mouse testes and found that the mice were completely sterile. Following this she did RNA-seq and HITS-clip for RBMXL2 to better understand the role of RBMXL2 protein in the testes. In addition, she noted that the testes from knock out mice were smaller in comparison to their wild-type littermates. Histological examination and analysis of chromosome spreads of mouse testes showed that the knockout of *Rbmxl2* resulted in an arrest in diplotene and no sperm were detected. This indicates the importance of RBMXL2 in meiosis. Moreover, her work showed the importance of RBMXL2 protein in preventing the mis-splicing of long exons in essential genes during meiosis. She also showed that RBMXL2 can suppress the cryptic splicing of exons that might impact the meiotic gene expression like *Kdm4d*, *Lrrcc1*, and *Meioc*.

In this project, I aimed to further investigate the mechanism of RBMXL2 function. I did experiments to determine whether the phenotype Ingrid observed in C57BL/6 could be replicated in other mouse strains. I performed iCLIP experiments in order to identify new RBMXL2 RNA targets. Finally, I wanted to determine whether RBMXL2 replaces RBMX during meiosis, or whether these proteins have a different role in a male germ cells and somatic cells.

1.12.2 Project aims:

1.12.2.1 Investigating whether Knockout of *Rbmxl2* will show different phenotypes in different mice strains?

To extend the previous findings by Dr. Ehrmann (2019) in our lab that knockout of *Rbmxl2* causes spermatogenesis arrest at diplotene in meiosis by examining different mouse strains. *Rbmxl2* knockout in C57BL/6 background mouse caused male infertility with a block at diplotene. But preliminary data suggested that this phenotype might be different in the mouse Sv/129 strain.

Therefore, our question was, does the same phenotype manifest in different mouse strains? The reason for testing this is another very well-known spermatogenesis gene *Dazl* has strain specific effects (Saunders et al., 2003).

1.12.2.2 Identification of RNA targets of RBMXL2 in mouse testis using RNA-seq and iCLIP:

To identify RBMXL2 protein-RNA interactions, using individual cross-linking and immunoprecipitation (iCLIP) will be used. The principle behind iCLIP is to crosslink RNA-protein interactions in vivo, then immune-precipitate the resulted complexes using specific antibody. With the higher-resolution positional data, the mechanism of splicing regulation of RBMXL2 may be revealed. Further analysis included using our previous HITS-CLIP and RNA-Seq data, to correlate RNA binding with changes in the splicing pattern.

1.12.2.3 Is RBMXL2/RBMY provides a direct or specialist replacement for RBMX during Meiosis?

RBMXL2 is expressed exclusively in the testes, especially during meiosis and directly after the inactivation of XY chromosomes. Moreover, *RBMXL2* is a paralog of *RBMX* that evolved 65 million years ago. Therefore, we hypothesised that RBMXL2 might replace RBMX during meiosis (in which case it would have shared targets), or might have another important role in spermatogenesis (in which case the proteins might have their own specific targets). To test this I used a rescue experiment to see if RBMXL2 could replace the function of RBMX in somatic cells.

Chapter 2: Material and methods:

2.1 Mouse lines

The testes of all mice used in my Ph.D. project were from the C57BL/6 and Sv/129 inbred lines. For Sv/129, mice were backcrossed for 6 to 9 generations before being used in the experimental cohort. Mice were looked after by Mrs Caroline Dalglish from David Elliott's lab. She assisted in collecting adult mouse testes and day 21 mouse testis, both for the iCLIP experiments and the histological studies for the two mouse strains.

2.2 Animal care

All mice were kept in the animal unit (Centre for life). Mice were looked after and humanely culled via cervical dislocation; the mouse was then dissected, and the testes were removed and placed on ice before being weighed (average of both testis were calculated) and placed in fixative solutions by Dr. Ingrid Ehrmann and Mrs Caroline Dalglish from David Elliott's lab. All mice were treated according to the Comparative Biology Centre (CBC) guidelines, and bred under Home Office License PPL number: PP4033272.

2.3 Genotyping

All of the mouse genotyping was done by Dr. Ingrid Ehrmann and Mrs. Caroline Dalglish (Ehrmann et al., 2019). Mouse ears were clipped before she did the PCR to confirm the knockout of *Rbmxl2*.

2.4 Histological sections:

All mouse testes were stained for a qualitative assessment; for each staining experiment, 3 mouse testes were examined (n=3) with appropriate controls (negative controls) or positive controls, for instance, wild type vs. knockout in both mouse strains. I looked at 50 fields to choose tubules with cells to be described.

2.4.1 Paraffin wax embedding of mouse testes

All collected testes were fixed in 4% paraformaldehyde for 24hrs. The wax-embedding time was modified based on the mice's age (and related testes size) Table 2.1. All of the fixation and embedding solutions were formulated with DEPC-treated water.

	Day 21	Adult
4% PFA at RT	1 night	2 night
50% EtOH Shaking at RT	30 mins	2 hours
70% EtOH (x2) Shaking at RT	30 mins	2 hours
Testes could be stored at room temperature		
95% EtOH Shaking at RT	30 mins	60 mins
100% EtOH (x2) Shaking at RT	30 mins 60 mins	60 mins O/N
Histoclear (x2) Shaking at RT	15 mins 15 mins	20 mins 20 mins
Histoclear/Wax Heat 60-65°C	20 mins	30 mins
Wax (x3) Heat 60-65°C	30 mins	60 mins
	30 mins	60 mins
	30 mins	60 mins

Table 2.1: Steps for testes paraffin wax embedding.

2.4.2 Testis sectioning

A Leica RM2235 manual rotary microtome (Biosystems Swaziland) was used to section paraffin wax tissues to a thickness of 5 μ m. First, the microtome surface was cleaned and wiped down before use. Next, tissue sections were aligned on microscope glass slides (Marienfeld Superior, 1000612) and then adhered to the slides by heating them with 37°C water. The slides were then placed in an incubator at 37 °C for the night. Finally, the slides were stored the following day at 4°C in the refrigerator.

2.5 Staining:

All testes were fixed in Bouin's or 4% paraformaldehyde (PFA) solutions for 24 h and then processed through serial ethanol dilutions (50%, 70%, 95%, and 100%). Then, the testes were treated with Histo-Clear (National Diagnostics, NAT1334) twice, then wax was added at 65°C. Following the protocol from the above table 2.1. Formalin-fixed paraffin- embedded (FFPE) sections were cut into 5- μ m thick sections for all histological stain.

2.5.1 H&E staining

Slides were dewaxed with Histo-Clear (National Diagnostics, NAT1334) twice. Then they were re-hydrated by serial dilutions of ethanol (100%, 95%, and 75%) and washed in PBS-T. Slides were then placed in Harris hematoxylin for 45 seconds. Next, slides were washed in water and placed in Blue Scot's tap water substitute. The slides were quickly rinsed with water before differentiating for about 15 seconds in 1% acid alcohol. Slides were then rinsed with water and put in Blue Scots, a tap water substitute. Slides were placed in Eosin for 10 seconds. The slides were rinsed in running water. Slides were then dehydrated through serial ethanol dilutions (70%, 95%, and 100%), and placed in Histo-Clear twice (5 minutes each). Finally, slides were mounted in DPX (a mixture of distyrene, a plasticizer, and xylene), and a coverslip was added. PBS-T made using Phosphate-buffered saline (PBS) tablets (Thermofisher, cat# 18912014), five tablets added to 1L of dH₂O with 1ml of Tween 20 (Thermofisher, cat# 85113).

2.5.2 Immunolocalisation using Fluorescent staining

Slides were de-waxed using Histo-Clear, then re-hydrated by ethanol (100%, 95%, and 75%) and washed in PBS-T. Antigen retrieval was performed by microwaving the slides in 0.01 M sodium citrate, pH 6.0, for 20 min. Slides were subsequently blocked with 10% horse serum in PBS-T, and incubated with the primary antibody, overnight at 4°C, in a humidified box.

The next day, the slides were washed with PBS-T buffer before adding the secondary antibody, at room temperature for 1 hour. One drop of Vectashield Anti-fade, containing 4',6-diamidino-2-phenylindole (DAPI, Vector Laboratories, H-1200), was added, followed by addition of a coverslip, which was sealed with nail polish. Slides were wrapped in foil and stored at 4°C.

Primary antibodies used were anti-RBML2 (in house, 1:100), anti-RBMY (in house, 1:100) and anti-γHa2X (Sigma-Aldrich, cat# 05-636 1:200). Secondary antibodies used were Anti-Rabbit Alexa Fluor® 488 secondary antibody (Thermofisher, Cat#A-21206), and Anti-mouse Alexa Fluor® 594 secondary antibody (Thermofisher, Cat# A10040).

2.6 Minigene assays

2.6.1 Cloning of mouse *Esco1* exon 3 into an exon trap vector and analysis in HEK293 cells

In order to investigate the splicing regulation of *Esco1* gene (exon 3), a region of mouse genomic DNA containing exon and flanking intron sequences for exon 3 of the *Esco1* gene was amplified by Platinum™ PCR SuperMix High Fidelity (Invitrogen, cat; 12532016) for 30 cycles in Thermocycler. The product length was 2 kb using the primers shown in table 2.2. The *EcoR1* restriction enzyme was used to digest the PCR product, which was then cloned into the de-phosphorylated-pXJ41 vector cut at the *Mfe1* site (Bourgeois et al., 1999). The pXJ41 vector and the insert (2.4kb) were digested by two restriction enzymes (*EcoR1* and *Mfe1* respectively) then ligated together. Following ligation, a re-cleavage was carried out using the *Mfe1* enzyme in order to remove the religated vector without the insert (that can often contaminate ligation mixes). The ligation mix was transformed into *E.coli* DH5α cells, and plated on agar plates that contain ampicillin. 40 colonies were then screened by PCR with insert forward primer (*Esco1*-F) and reverse primer PXJR (complementary to a sequence within the vector).

A colony was selected and grown on LB medium with 50µg/ml ampicillin, incubated overnight on 37°C shaker. Plasmid purification was carried out, and plasmid sequencing was done to ensure the presence of the insert. Sanger sequencing (Source Bioscience) was used to verify that the cloning was correct. To perform transfection of plasmids to carry out the splicing analysis the HEK293 cell-line and lipofectamine 3000 (Invitrogen) were used. Cells were transfected with expression constructs encoding different RNA- binding proteins (RBMXL2, RBMX). An illustration of the minigene experiment seen below in figure 2.1. Cell pellets were separately used for RNA extraction and protein analysis. Cell pellets for RNA extraction were re-suspended in 150µl Trizol (Lifetechnologies). TRIzol was used to extract RNA using the standard protocol provided below, then was analysed with a Qiagen kit (PCR with reverse transcription) as instructed by the manufacturer. (Qiagen) one-step RT-PCR was carried out using the primers pXJRTF and PXJB1 whose sequences are shown in table 2.2. Capillary gel electrophoresis was used to analyse and quantify RT-PCR reactions using the PSI (%) formula (found in chapter 2nd section **2.6.13**). For protein analysis the cell pellets were immediately lysed in a 25-µl (2X) SDS sample loading buffer.

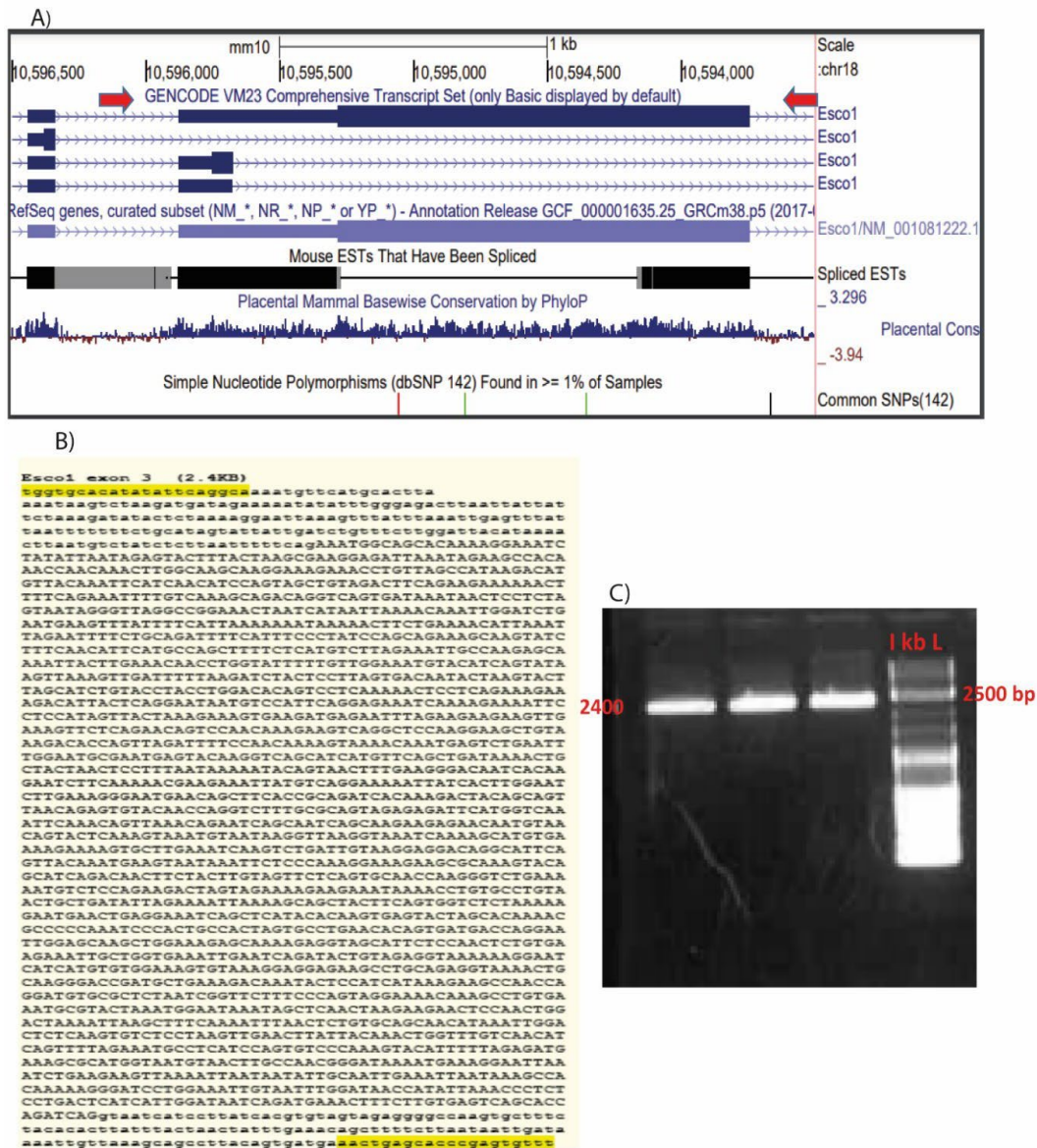


Figure 2. 1 : cloning an Esco1 minigene: A) The Esco1 exon 3 track from UCSC browser (<http://genome.ucsc.edu>). Red arrows represent the *Esco1*-3F and *Esco1*-3R primers used to amplify the exon with flanking introns. B) Shows exon 3 sequence with flanking introns, lower case letters indicates the introns sequences while the upper case is representing the exon sequences. C) 1% agarose gel shows electrophoresis of exon 3 and flanking intron amplification product from mouse genomic DNA. The PCR products (2.4 kb) are shown in triplicate along with 1kb plus DNA ladder.

2.6.2 Cloning of human *Esco1* exon 4 into an exon trap vector and analysis in HEK293 cells.

The same strategy was conducted to analyse splicing of human *ESCO1* exon 4. My colleague Chile detected *ESCO1* exon 4 skipping using RNA-seq analysis after *RBMX* knockdown.

Therefore, a minigene construct was produced using PCR-amplified human genomic DNA. The amplified product was 2.4 kb using the primers shown below in table 2.2. Exon 4 of human *ESCO1* is another ultra-long exon that covers 2.6 KB. Exon 4 and flanking introns were PCR amplified, cut with *EcoR1*, and cloned into the pXJ41 exon trap vector, digested by the restriction enzymes *MfeI*. Like the previous minigene, re-cleavage was done using the *MfeI* enzyme to remove the re-ligated vector without the insert. Transformation into *E.coli DH5* cells was performed, followed by plating on ampicillin-containing agar plates. Insert forward primer (ESCO1-F) and reverse primer (PXJR) were used for colony screening. A colony was grown on LB medium and then incubated overnight at 37°C. Purification and sequencing were performed to confirm the insert's presence. In order to analyse splicing, HEK-293 cells were transfected with the insert as seen in the figure 2.2.

Primers	Forward	Reverse
<i>Esco1</i> exon 3 (mouse)	AAAAAAAAAAGAATTcgggtgcacatatattcaggca	AAAAAAAAAAGAATTCaaacactcgggtgctcagtt
<i>ESCO1</i> exon 4 (human)	AAAAAAAAACAATTGGGGATACTTCGTTCTGTTA G CC	AAAAAAAAACAATTGACTCTCGCTAACAACTTA A CAGG
pXJRTF (forward) and pXJB1 (reverse)	GCTCCGGATCGATCCTGAGAACT	AGCAGAACTTGTTTATTGCAGC

Table 2.2: primer sequences used for cloning mouse and human *ESCO1*, Upper case indicate the A residues and restriction site and lower case shows the exon.

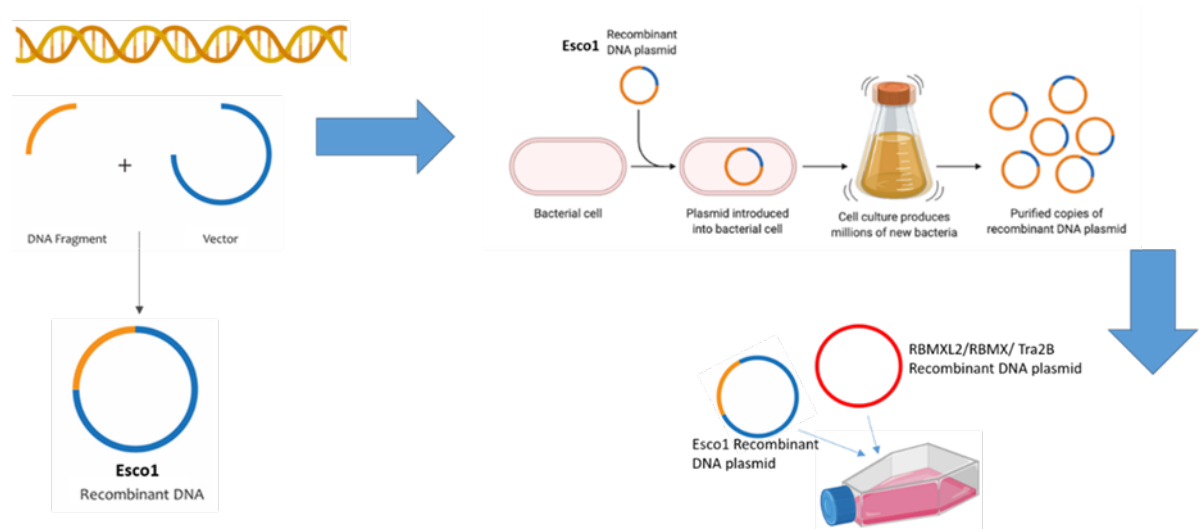


Figure 2. 3: illustrating minigene experiment. From amplification of genomic DNA to inserting the studied exon into a vector, producing more copies using bacterial cells, transfecting different RNA-binding proteins (RBMXL2 and RBMX) alongside this recombinant DNA, and isolated RNA and proteins from these cells to investigate splicing changes. Figure made with BioRender.

2.6.3 Primer design

Primers for mouse *Esco1* exon 3 and human *ESCO1* exon 4 were designed using UCSC genomic browser website and then Primer3web website (<https://primer3.ut.ee/>). The whole exons were included with small intronic regions from both sides around 20 bp. Primers were ordered from IDT DNA (<https://eu.idtdna.com/>)

The exonic region of genomic DNA (from both mouse and human samples) was amplified using PCR. PCR products were examined by agarose gel electrophoresis on a 1% TAE gel with a ladder to confirm the correct product size using Platinum™ PCR SuperMix High Fidelity (Invitrogen, cat; 12532016). The PCR recipe and thermocycler programme are seen below in tables 2.3 and 2.4.

PCR reaction	1X RXN
5X SuperFi™ II Buffer	2µl
Forward primer(10µM)	1µl
Reverse primer(10µM)	1µl
10 mM dNTPs	0.2µl
Platinum™ SuperFi™ II DNA Polymerase	0.1µl
Template (insert, vector, 20ng/µl)	0.5µl
dH2o	2µl
Total	10 µl

Table 2.3: PCR reactions recipe from thermofisher website. Thermocycler programme

Cycle step	Temperature	Time	Cycle
Initial denaturation	98°C	30 seconds	1
Denaturation	98°C	5–10 seconds	30
Annealing	60°C	10 seconds	
Extension	72°C	60 seconds	
Final extension	72°C	5 minutes	1
Hold	4°C	∞	-

Table 2.4: Thermocycler programme used.

2.6.4 Purification of DNA fragments from agarose-TAE gels

PCR products were digested with restriction enzymes following the manufacturer's protocol (QIAquick® PCR & Gel Cleanup Kit from Qiagen). Then, the products were sent for sequencing to confirm that the whole exon is amplified.

2.6.5 Preparation of plasmid DNA and bacterial cloning

Select chemically competent cells (Bioline) were used to transform ligated plasmids using the manufacturer's heat shock technique. Transformed cells resistant to ampicillin were grown for 16 hours at 37 °C on LB agar plates with ampicillin. PCR was used to screen colonies using primers that were unique to the vector and insert combinations. All cloning-related PCR reactions were conducted using the Phusion High-Fidelity DNA Polymerase Kit (Thermo Scientific). After the screening process, plasmids were sent to Source BioScience for sequencing. Samples of purified plasmid DNA were diluted to 100ng/l and analysed in automated gel electrophoresis using 10 µM of pXJRTF and pXJB1 primers. The obtained sequences were aligned onto the genome (using the BLAT on UCSC genome browser) and visually examined.

2.6.6 Sanger sequencing

Samples were sent to Source BioScience LifeSciences for DNA sequencing. First, since many exons are ultra-long, I investigated the PCR products to ensure that the entire exons and flanking intron were amplified. Then, after cloning, primer sets for the pXJRTF forward primer and the pXJB1 reverse primer (10 mM each) was sent along with 100 ng/l samples of plasmid miniprep to make sure that the correct sequences had been cloned into each minigene.

2.6.7 Cell culture

HEK293 and flp-in HEK293 cells were seeded in Medium (DMEM), a high-glucose-pyruvate-medium provided by (Life Technologies), and the medium was enriched with 10% foetal bovine serum (FBS) and treated with 1% penicillin/streptomycin (Life Technologies). To prepare cells for minigene assays, 6-well plates of cells were cultured at 37 °C and 5% CO₂. Due to its high transfection rate, the human embryonic kidney cell line HEK293 (Graham et al., 1977) was used for the minigene assays. Before cell harvesting, expression of GFP-tagged proteins was observed under a fluorescent microscope. Then media was removed, and cells were washed with PBS and 300µl Trypsin-EDTA were added, and the cells were returned to the incubator for 5 minutes at 37°C to allow cells to detach, after which 1ml of the media (DMEM) was added to the wells. From each well, two aliquots of the re-suspended cells were taken and pelleted by microcentrifuge.

2.6.8 Cell passages and maintenance

All cell culture work was done in a Class II laminar flow microbiological safety cabinet. The HEK293 cell line was seeded in 25 cm² and 75 cm² tissue culture flasks in a humid incubator at 37 °C containing 5% CO₂. HEK293 cells were maintained and passaged every 3 to 5 days in DMEM media supplemented with 10% FBS (Sigma-Aldrich) with 1% penicillin/streptomycin (Sigma-Aldrich). Cell passages with around 70 to 80% confluence were removed by removing the media, rinsing the flask with sterile PBS, and treated with 2mM trypsin-EDTA (Sigma-Aldrich) for 5 min in the incubator. Media were introduced to counteract the impact of trypsin-EDTA. Centrifugation was used to collect all detached cells, the pelleted cells were then resuspended in DMEM media and passed at a ratio of 1:4.

2.6.9 Cell counting

A Neubauer chamber haemocytometer was used to count cells before experiment. Media were added to the cell pellets, then an aliquot of single cell suspension (10µl) was placed into haemocytometer. Using low-power (10x) magnification, the number of cells that surrounded the ruled grid area was counted. The appropriate number of cells was then seeded for each experiment by diluting the cell suspension by the cell density per millilitre.

2.6.10 Lipofectamine 3000 transfection

Following the instructions from the manufacturer (Invitrogen), Lipofectamine® 3000 (Invitrogen) was used to transfect the cells in 6-well plates with 100,000 cells in each well. The volumes used and incubation times are given in the table below 2.5.

Transfection protocol	Components	volume	Incubation time
Reaction mix			
Mix A	Opti-MEM® Medium	125µL	5minutes
	Lipofectamine 3000 reagent	5 µL	
Mix B	Opti-MEM® Medium	125 µL	5minutes
	DNA (200ng)+ plasmid (500ng)	5 µL	
	P3000 reagent	5 µL	
A+B	Mix A	125 µL	10-15 minutes
	Mix B	125 µL	
Short spin			
Add 250 µL to each well in 6 well plate and incubate at 37c overnight			

Table 2.5: Transfection protocol from Thermofisher scientific website.

2.6.11 RNA extraction

150 μ l of Trizol (Life Technologies) was added per tube for RNA extraction and incubated at RT for 5 min. The mixture was then vigorously mixed with 20 μ l of chloroform. After 10 min of incubation at RT, samples were centrifuged at 13,000 rpm for 10 min. The top layer (about 50 ml) was taken out and put into a new Eppendorf. An equal amount of isopropanol was then added to the new Eppendorf.

After a 15-minute incubation at 37 °C, samples were centrifuged at 13,000 rpm for 10 min at 4 °C to form a white RNA pellet. The supernatant was removed cautiously, and the pellet was washed with 70% ethanol and air-dried. Lastly, samples were re-suspended in 30 μ l dH₂O treated with diethylpyrocarbonate (DEPC). The RNA concentration was measured using a Nanodrop spectrophotometer. Prior to RT-PCR, RNA samples were diluted to 50 ng/ μ l using RNase-free dH₂O.

2.6.12 RT-PCR

RT-PCR was done following the one step-PCR protocol from Qiagen. RT-PCR samples of (5µl) were diluted with equal amount of the QIAxcel DNA dilution buffer (Qiagen). Samples were then electrophoresed on the QIAxcel multi-capillary electrophoresis system (Qiagen). Analysis of the samples was done using the QIAxcel Biocalculator software (Qiagen) to determine the size of each PCR product (bp) with the concentration of each band (ng/µl).

RT-PCR was used more than once during my PhD project. The table below 2.6 shows the primers I used for chapter 3 where I detected similar patterns for RBMXL2 target expression during neonatal development in both mouse strains.

Primers	Forward	Reverse	Internal
<i>Kdm4d</i>	AAGGCGCAAATAA GTACGGG	TCTTGTAGGCTACTGGGTGC	CCAGAAACTACTTT GGCTGTG
<i>Meioc</i>	AGACCGAAAGAAA TGACTATGGC	ATCTGCTTGTGTTTCATTGGCC	GCTCGCCATTGTTCA TAGCA

Table 2.6: primers used for time course experiment in the two mouse strains.

2.6.13 Calculation of percentage splicing inclusion (PSI %):

Bands from products detected in RT-PCR experiments were quantified and compared using the formula below:

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

2.6.14 Western blotting

HEK293 cells were resuspended in loading dye (2x sodium dodecyl sulfate (SDS)). The mixture was then sonicated and heated for 5min at 95 °C. The proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The proteins were transferred to a nitrocellulose membrane through wet transfer for 1 hour. Next, the membrane was incubated with a blocking buffer [5% Milk, dissolved in phosphate-buffered saline (PBS) and 2.5% horse serum]. The membranes were stained with primary antibodies (at the concentrations indicated below) diluted in blocking buffer overnight at 4°C. After incubation, the membranes were washed three times with Tris-buffered saline containing Tween-20 (TBS-T). Secondary antibodies conjugated with horseradish peroxidase were used to detect primary antibody binding. Finally, an X-ray film developer machine used Clarity Western ECL Substrate (GE Healthcare Systems) to detect the bands developed in the dark for 2 seconds. The following primary antibodies were used: anti-RBMXL2 (generated in-house, 1:1,000) anti-Tubulin (Abcam, 1:2,000), anti-RBMX (Cell Signalling, 1:1000), anti-GAPDH (Abcam, 1:2000) and anti-FLAG antibody (Sigma, 1:1000).

Secondary antibodies were as follows: anti-mouse HRP (Jackson ImmunoResearch Labs, 1:2000), anti-rabbit HRP (Jackson ImmunoResearch Labs, 1:2000).

2.7 iCLIP

2.7.1 Definition

The iCLIP protocol was obtained from the Ule Lab (unpublished) and used to identify RBMXL2 protein targets in mouse testes. The new protocol utilises some changes from the irCLIP method (Zarnegar et al., 2016) and the eCLIP (Van Nostrand et al., 2016) but is mainly based on the old iCLIP protocol (Huppertz et al., 2014). Figure 2.4 illustrates the iCLIP protocol. Wild-type mouse testes (4 testes) were collected and used fresh. First, each testis was cut into small pieces with a sterile scalpel over ice, then 1 ml ice-cold PBS (Gibco) with protease inhibitors (Protease Inhibitor Cocktail Set III, EDTA-Free, Merck) were added. Consecutively, testis was passed through serial pipetting from a P1000 tip, then a P200 and finally a P10 tip. Homogenised testes were then saved in an Eppendorf tube in a -80 °C freezer.

2.7.2 Steps:

2.7.2.1 UV cross-linking of mice testes

Mouse testes were placed in sterilised plates on ice and irradiated four times (with shaking in between) with 100 mJ/cm² in a Stratalinker 2400 at 254 nm. Cells were collected and centrifuged for 1 min at 514 g at 4°C. Then, the supernatant was removed, and the pellets were stored in a -80 °C freezer.

2.7.2.2 Preparation of magnetic beads

Lysis buffer was prepared as advised in the protocol table 2.7, then was used twice (900 µl) to wash the 100 µl of protein G Dynabeads (Invitrogen). Then, beads were re-suspended in 100 µl lysis buffer with 5 µg of anti-flag antibody (Sigma F1804) per tube. Next, the room temperature incubation was carried out for an hour. Beads were then washed with 900 µl of the lysis buffer three times and then placed in 400 µl lysis buffer until the lysate was ready.

2.7.2.3 Cell lysis, partial RNA digestion and immunoprecipitation

The pellet containing the UV cross-linked cells was resuspended in 1 ml of lysis buffer treated with protease inhibitors (Protease Inhibitor Cocktail Set III, EDTA-Free, Merck), transferred to a new (1.5 mL) Eppendorf tube, and sonicated using a Bioruptor for ten cycles with 30 seconds on/off intervals at low intensity. The lysate concentration was determined using the Bradford assay. All samples were diluted to 1 mg/ml or the concentration of the least concentrated sample (whichever was less), and 2 µl Turbo DNase (Ambion, AM2238) were added to the cell lysate. Two conditions of RNase I (Thermo Scientific, EN0602) were used: these were 0.8 (Low RNase) or 2.5 (High RNase) units per ml. At 37°C, the cell-lysate was incubated shaking at 1100 rpm for 3 min for the low RNase and 5 min for the high RNase, then transferred into ice for 3 min. The supernatant from centrifuging the lysates at 15000 rpm for 10 minutes at 4°C was moved to a new 1.5 ml tube. Finally, cell lysates were moved to a Proteus Clarification Mini Spin Column (Generon; GEN- MSF500) and centrifuged at 13000 rpm for 1 min at 4°C (500 µl at a time). Cell extracts were added to the beads from step 2.7.2.2. The beads/lysate mix was rotated overnight at 4°C, in the cold room, washed 2x with high-salt wash table 2.7 (the second wash was rotated for at least 1min in the cold room). Beads were washed 1x with PNK buffer table 2.7, resuspend in 1X PNK wash buffer (Table 1) and moved to a new tube.

2.7.2.4 De-phosphorylation of RNA 3' end and linker ligation

8 µl of 5x PNK pH 6.5 buffer, 1 µl of PNK (NEB M0201L), 0.5 µl of FastAP alkaline phosphatase (ThermoFisher, EF0654), 0.5 µl of RNasin, and 30 µl of water were mixed to make up a 40 µl dephosphorylation mix. This was used to resuspend the beads, which were then incubated at 37 °C for 40 minutes while shaking at 1100 rpm in the thermomixer.

1x Ligation Buffer (DTT free) (containing 6.3 µl of 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), 2.5 µl pre- adenylated adaptor L3-???-App (see table 2.8 for sequence) and 9µl 50% PEG8000 was used to wash the beads once. Then 25 µl of the ligation reaction was added to the beads and incubated at RT for 75 min (while flicking the tube every 10 min). Finally, the beads were washed twice in the cold room with rotation for at least a minute before being re-suspended in 1 mL of PNK buffer. The beads were transferred to a new tube, put on a magnet to remove the buffer, and 20 µl of a mix was added 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).

Then beads were incubated for an hour at 30°C, followed by 30 mins at 37°C while shaking in a thermomixer at 1100 rpm. The beads were washed twice with high salt wash while rotating for at least 1 minute in the cold room and once with 1 ml of PNK buffer.

2.7.2.5 SDS-PAGE and membrane transfer

The PNK buffer was discarded, and the beads were resuspended in 20µl of preheated at 70°C 1x NuPAGE loading buffer (made by combining 4x stock; Invitrogen) with water and reducing agent DTT; the final concentration was 100 mM (from 1M stock = 10%). Then the samples were incubated in 1x NuPAGE loading buffer at 70°C for 1 min before being put on a magnet to collect the supernatant for loading onto the gel. According to the manufacturer's recommendations, the supernatant was added onto a 4-12% NuPAGE Bis- Tris gel (Invitrogen) alongside 6 µl of a pre-stained protein size marker. The gel was run for 65 minutes at 180 V with 0.5L of 1 x MOPS running buffer (Invitrogen) for size separation. The gel's dye front was removed and discarded (the dye might interfere with imaging). The protein-RNA-L3 complexes were transferred 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 washed in PBS buffer, placed in a clear document sleeve, and scanned in both the 700nm and 800nm channels using the Odyssey LI-COR CLx imager.

2.7.2.6 RNA isolation

The low-RNase samples were eluted from the membrane using a printed template of the infrared gel. The membrane slices were placed in a 1.5ml (Eppendorf Test tube DNA LoBind 1.5mL) microfuge tube containing 10 µl proteinase K (Roche, 03115828001) per 190 µl PK+SDS buffer table 2.7, and incubated at 50 °C for 60 minutes while shaking at 1100 rpm. The sample and 200 ul Phenol:Chloroform:Isoamyl Alcohol (Sigma P3803) were added to a separate tube, mixed, and added to a pre-spun 2 ml Phase Lock Gel Heavy tube (713-2536, VWR), incubated for 5 minutes at 30 °C while shaking at 1100 rpm, and then spun for 5 minutes at 13000 rpm at room temperature. 800 ul of chloroform was added to the top phase of the Phase Lock Gel Heavy tubes, which were then inverted and spun at 13000 rpm for 5 minutes at 4°C. I then

transferred the aqueous layer to a new tube and precipitated it overnight at -20 °C with 0.75 ul Glycoblue (Ambion, 9510), 20 ul 5 M sodium chloride, and 500 l 100% ethanol.

2.7.2.7 Reverse transcription

The RNA was resuspended in 5.5 µl of dNTP mix (10 mM) and reverse transcribed in 1 µl of primer irCLIP ddRT_## (1 pmol/l) and 0.5 µl of dNTP mix (10 mM). Table 2.8 gives the sequence of RT primers with individual barcodes for each reference.) The RNA samples were placed at 65°C for 5 min and then cooled to 25°C until the RT-PCR mix was prepared with 2 µl 5x SSIV buffer (Invitrogen), 0.5 µl 0.1 M DTT, 0.25 l RNasin, 0.25 µl Superscript IV. Reverse transcription was done using the following 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, incubated for 15 minutes at 85°C for alkaline hydrolysis of the RNA, and then neutralised with 1.25 ul of 1M HCl.

2.7.2.8 cDNA purification

37.5µl (3x) of Agencourt AMPure XP beads and 21µl (1.66x) of isopropanol was added to the newly transcribed cDNA from the previous step (2.7.2.7). First, this was vigorously mixed and incubated at room temperature for 5min. Then the beads were kept on a magnetic rack for 2–3 mins. Next, the supernatant was discarded and washed twice with 200 ul of 85% ethanol for 30 seconds. Next, the beads were dried for one minute at room temperature. After the ethanol was removed, cDNA was eluted into 9µl of water, incubated for 3 minutes, and then transferred to a new tube.

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

The purified cDNA was circularised in the following combination: (1.5 µl 10x CircLigase Buffer II (Epicentre), 0.75 µl CircLigase II (Epicentre), 0.75 µl 50 mM MnCl₂ (included in the CircLigase II kit), and 3 µl 5M betaine (included in CircLigase II kit)) and incubated at 60 °C overnight. The circularised cDNA was purified using AMPure XP beads as described before (step 2.7.2.8), with the reaction volume increased to 45 µl of AMPure XP beads, 25 µl of isopropanol, and 10 µl of water.

2.7.2.10 Circularised cDNA amplification

The circularised cDNA was amplified by PCR using the following PCR mixture: 4 µl cDNA, 15 µl water, 1 µl primer mix P5Solexa/P3Solexa 10 M (Table 2-8), and 20 µl Platinum HS master mix. The PCR protocol consisted of 98°C for 40 seconds, 19 cycles of [98°C for 20 seconds, 65°C for 30 seconds, and 72°C for 45 seconds], 72°C for 3 minutes, and a 25°C hold.

2.7.2.11 Purification with AMPure

AMPure XP was utilized to purify amplified cDNA at a volume twice that of the cDNA. First, the mixture was thoroughly mixed using a vortex. Then, the mixture was sat for 5 minutes at room temperature. The beads were then placed on a magnetic rack and washed with 500 µl of 70% ethanol. The beads were then placed on a magnetic stand for 30 seconds, after which the ethanol was removed, and they were allowed to dry for one minute. The purified PCR product was then eluted using 40 µl of TE buffer and 30 seconds of vortexing.

2.7.2.12 Size selection of cDNA library by gel purification

The PCR product was size separated on a 6% TBE gel at 180V for 30 minutes, stained with SYBR Green, and visualised using a blue light transilluminator. DNA between 145 and 400 nt was collected from the gel by crushing the gel in 500 µl of Crush-Soak Gel buffer table 2.7, DNA was obtained from the gel over two hours of incubation at 65 °C with thermomixer settings of 15 seconds at 1000 rpm followed by 45 seconds of rest. The liquid component of the supernatant was added to a Costar SpinX column (Corning, Inc., 8161), which included two 1 cm glass pre-filters (Whatman, 1823010). The sample was spun at 13000 rpm for 1 min, and the liquid was combined with 1 l glycobule, 50 µl 3M sodium acetate, pH 5.5, and 1.5 ml of absolute ethanol overnight at -20 °C.

2.7.2.13 High-throughput sequencing

The iCLIP protocol contains two critical steps: (1) the visualisation of the infrared protein- RNA-L3 complexes after transfer to the Protan BA85 Nitrocellulose Membrane, and (2) the visualisation of the cDNA library using the 6% TBE gel. Before sequencing, samples were sent to be analysed using the Tapestation and then submitted for high-throughput sequencing on the Illumina NextSeq 500 in collaboration with Mr. Rafiqul Hussain at the Newcastle University Genomics Core Facility.

(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
(C)	PNK Wash Buffer
	20 mM Tris-HCl, pH 7.4
	10 mM MgCl ₂
	0.2% Tween-20
(D)	4x Ligation Buffer
	200 mM Tris-HCl, pH 7.4
	40 mM MgCl ₂
	4 mM dithiothreitol
(E)	10x Ligation Buffer (DTT free)
	500mM Tris-HCL 7.5
	100mM MgCl ₂
(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

Table 2.7: List of buffer components used in the iCLIP experiment.

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_ddR T_30 RT primer	/5Phos/ WWW CGATC NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC
irCLIP_ddR T_32 RT primer	/5Phos/ WWW CTCGA NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC
irCLIP_ddR T_34 RT primer	/5Phos/ WWW GAAAC NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC
irCLIP_ddR T_35 RT primer	/5Phos/ WWW GAGTT NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC
P3 Solexa	CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCCTGCT GAACCGCTCTTCCGATCT
P5 Solexa	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACG ACGCTCTTCCGATCT

Table 2.8: 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.

2.7.2.14 Bioinformatics analysis of iCLIP sequencing data

My colleague Chile Siachisumo assisted me with the iCLIP analysis. Firstly, the raw files of the four replicates were given to Dr. Graham Smith (Newcastle University Bioinformatics Support Unit) to create the Fastq files to be later used in the iMAPs pipeline. The iMAPs pipeline eliminate the PCR duplicate copies, and used STAR to line iCLIP tags up with the mouse genome sequence (mouse GRCm39). Then merged files were created for the four samples that contain the crosslink sites. First, Bedtools merged the four files by adding 5 nucleotides to each file (Quinlan & Hall, 2010). Next, a TRIMGALORE was used to remove residual Illumina-sequencing adaptors or low-quality reads at the 5' end. Then BOWTIE_ALIGN was used to align the demultiplexed reads against the rRNA and mature tRNA indexes to prevent contamination. Finally, Galaxy hub (<https://usegalaxy.org>) was utilised to perform PCA and correlation analysis for the four samples and the no-antibody (negative control).

2.7.2.15 Gene Ontology (GO) enrichment analysis

From the iMAPs analysis, a list of genes was generated. Then, I performed Gene Ontology Enrichment Analyses on the first 500 genes using ShinyGO 0.77 (<http://bioinformatics.sdstate.edu/go/>). I developed the graph myself. The p-value cutoff for the analysis was set at 0.05.

2.7.2.16 Analysis of long mouse exons

Using Ensembl Genes (<http://www.ensembl.org/biomart/>), Dr. Sara Luzzi annotated all mouse exons and produced a graph comparing mouse RNA-seq to RBMXL2 iclip.

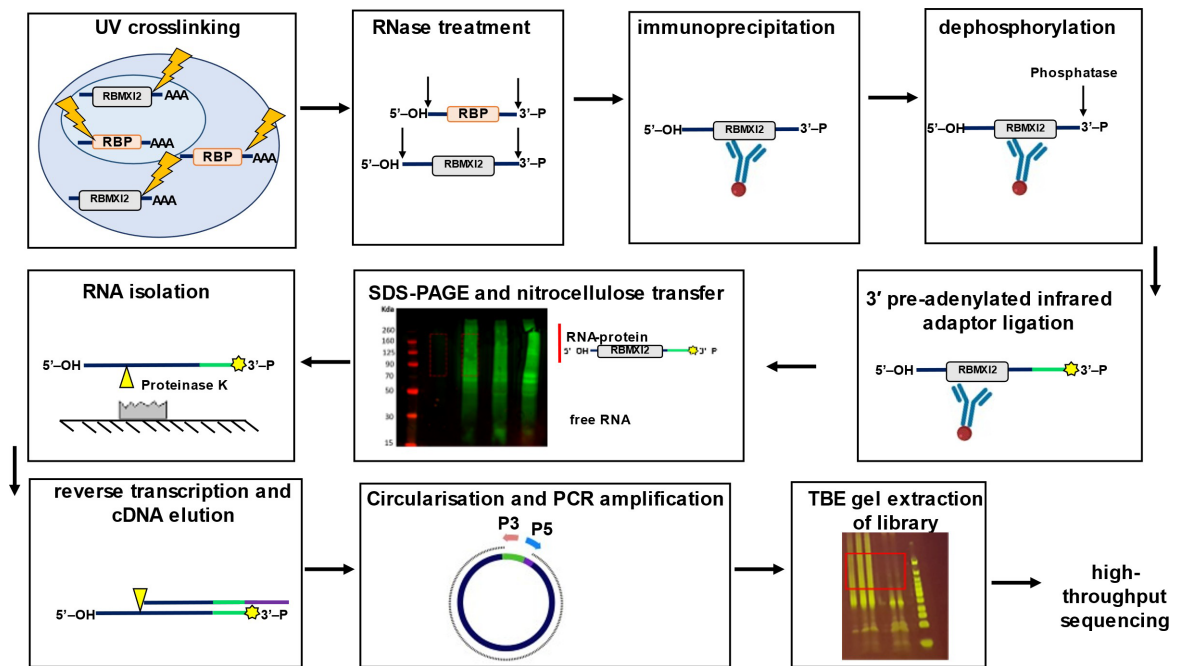


Figure 2. 4: Visual illustration of the iCLIP protocol. (1) UV radiation was used to induce irreversible covalent RNA-protein crosslinks in mouse testes. After sonication, RNase was used to digest the cells. (2) Cells were digested with RNase after sonication. (3) RNA was immunoprecipitated with the anti-RbmX12 antibody to form an RNA-protein complex. (4-5) RNA was dephosphorylated, and a pre-adenylated L3-IR adaptor 3' RNA linker was bound to the RNA complex. (6) The protein-RNA complexes were separated by size using the SDS-PAGE and nitrocellulose membrane. The area of the membrane with the protein-RNA complex was cut out with the help of a printed template of the area. (7) The protein-RNA region was fragmented with a scalpel and digested with proteinase K. (8) After purifying the RNA, a primer containing a barcode sequence was used to reverse-transcribe the RNA into cDNA. (9) AMPure XP beads were used to purify the cDNA, and then (10) Agarose gel was used to separate DNA in the appropriate size range. (150-400 nt) that was purified and sent for sequencing using the Illumina 500 machine.

2.8 SiRNA knockdown

Following the instructions from the manufacturer, Lipofectamine® RNAiMAX (Invitrogen) was used to transfect the cells in 6-well plates with 100,000 cells in each well. In addition, pre-designed siRNA was ordered from Integrated DNA Technologies (IDT), as seen in the tables below 2.8 and 2.9.

Transfection protocol Reaction mix	Components	volume	Incubation time
Mix A	Opti-MEM® Medium	125µL	5minutes
	Lipofectamine® RNAiMAX Reagent	5 µL	
Mix B	Opti-MEM® Medium	125 µL	5minutes
	siRNA (10 µM)	5 µL	
A+B	Mix A	125 µL	5 minutes
	Mix B	125 µL	
Short spin			
Add 250 µL to each well in 6 well plate and incubate at 37c overnight for 3 days			

Table 2.8: Transfection protocol from Thermofisher scientific website.

RBMX siRNA sequence	
hs.Ri.RBMX. 13.1-SEQ1	5'- rArUrC rArArG rArGrG rArUrA rUrArG rCrGrA rUrArG rArGA T -3'

Table 2.9: RBMX siRNA sequence used to create the knockdown.

2.9 RNA-seq

RNA sequencing made the study of alternative splicing across the entire transcriptome possible. RNA sequencing involves mRNA enrichment, fragmentation, cDNA synthesis, adapter ligation, library amplification, and sequencing. Then an expression profile for each gene is then generated by processing the sequencing reads and mapping them to a reference genome (Wang et al., 2009).

The Newcastle University Genomics Core Facility carried out the RNA sequencing. TruSeq-stranded mRNA-Seq was used to sequence the RNA, producing more than 40 million paired-end 100-bp reads. For each condition, a single sample was sequenced.

2.10 Statistical analysis:

To compare the testis weight between two backgrounds, I used two-way ANOVA to investigate whether there were statistically significant differences in the testis measurements when considering both the mouse background (categorised into two groups) and the presence or absence of *Rbmxl2* knockout (categorised into two groups).

In this study, I also used a time course experiment (over the first wave of spermatogenesis in the neonatal mouse) to compare splicing patterns in WT and *Rbmxl2* knockout mice.

Traditional parametric statistical tests, such as ANOVA, were not possible due to the lack of replication because there was only one subject or sample in each condition (KO and WT) for each time point days 12, 14,18 and adult mice (n = 1). As a result, I relied on descriptive statistics, such as means, medians, and standard deviations, to summarise the data, allowing me to describe trends or differences between KO and WT across different days. Although the small sample size prevented formal statistical hypothesis testing, I was aware of this limitation. I chose a descriptive, data-driven approach, acknowledging the inherent

limitations and aiming to extract exploratory insights from the available data. This preliminary finding aims to confirm the distinctions between WT and KO.

Finally, a statistical analysis was conducted for the rescue experiment. A non-parametric t-test was performed for each of the three replicates (n=3), comparing WT and RBMX KD with and without the overexpression of RBMXL2, RBMY, and RBMXL2 without RRM.

2.11 Details of rescue experiment:

To first establish this protocol I used a stable cell line from our lab that can overexpress RBMXL2-FLAG, RBMY-FLAG or RBMXL2 Δ RRM-FLAG after tetracycline induction. A stable cell line (Flp-In HEK293) was used. In the Flp-In HEK293 cell line, an FRT site, a Tet repressor, and a blasticidin resistance gene are expressed and integrated within these cells' genome. In my stable cell lines RBMXL2 is integrated into the FRT site. Expression of RBMXL2 cloned into the FRT site is regulated by the tetracycline promoter. Therefore, adding tetracycline (1mg/ml) will allow the expression of *RBMXL2* in those cells. In order to maintain the expression of RBMXL2, blasticidin was added on every third passage.

The rescue protocol starts with plating two 6-well plates—one as a control (without expression) and another for overexpression. The main difference between the two plates is the addition of tetracycline. First, cells were seeded at 100,000 cells per well as mentioned in section 2.6.7. Then, I performed *RBMX* depletion in the first 3 wells using an siRNA directed against *RBMX* using Lipofectamine™ RNAiMAX Transfection Reagent (Invitrogen). The other wells were transfected with a control siRNA to compare with the knockdown of *RBMX* (section 2.8) contains the siRNA sequences purchased from IDT. The 6 well plates were kept in incubators at 37 °C with 5% CO₂ for 48 hrs. Then, the experimental plate was induced with tetracycline to overexpress the RBMXL2-FLAG. After adding the tetracycline (1mg/ml) (2 μ l per well), I continued the incubation for 24 hrs before cells were harvested for RNA extraction (to be used for RT-PCR mentioned in section 2.6.12) and protein extraction (for a western blot mentioned in 2.6.14). Details of the experiment are shown in the figure below 2.5. I used Western blot analysis to confirm the overexpression of RBMXL2.

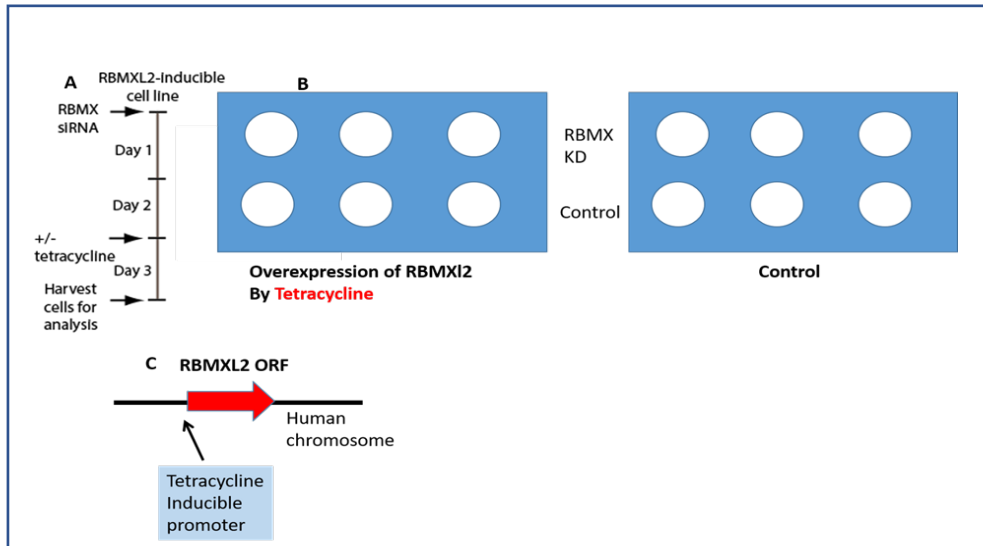


Figure 2. 5: The experiment design of rescue experiment. A) The experimental time course to knockdown *RBMX* and overexpress *RBMXL2* in stable HEK293 cells. B) The two (6-well plates) one for overexpression and the other as a negative control (no tetracycline induction). C) A schematic showing how this stable cell line was designed with an *RBMXL2* open reading frame that will express *RBMXL2* if Tetracycline is added.

Chapter 3: Do different mouse strains have a different *Rbmxl2* knockout phenotype?

3.1 Introduction:

The laboratory mouse has been used as a model to examine the effect of genetic mutations on germ cell development. The reason is since mice are a relatively easy system to generate gene mutations or knock out some genes and investigate the impact of these genes. This Chapter is based on a recent elife paper (2019), in which Dr. Ingrid Ehrmann from our lab successfully knocked out the *Rbmxl2* gene rendering the male mice infertile (Ehrmann et al., 2019). She also reported that no sperm were seen, and spermatogenesis was arrested at the diplotene stage in the C57BL/6 background. The testis-specific RBMXL2 protein is expressed in the germ cells from the pachytene to the round spermatid stage. In the work described in this Chapter I was trying to establish whether RBMXL2 might have roles outside of diplotene. The strategy was to cross the *Rbmxl2* knockout allele onto a different genetic background (Sv/129). For example, *Dazl* knockout mice showed an earlier embryonic phenotype in C57BL/6 compared to the mixed background, which shows a postnatal phenotype (Ruggiu et al., 1997; Lin & Page, 2005).

This Sv/129 background has previously been reported to sometimes have different phenotypes from the C57BL/6 background. Tissue damage, inflammation, and angiogenesis were found to be more aggressive in Sv/129 mice than in C57BL/6 mice (Kübler et al., 2021). In addition, C57BL/6 mice showed more sensitivity to renal damage triggered by ischemia/reperfusion injury (IRI) than what was observed in Sv/129 mice (Lu et al., 2012). Imai et al., 2020 discovered a significant increase in the incidence of testicular teratomas in the Sv/129 mice strain when *Dnd1* was knocked out, whereas no teratomas were found in the C57BL/6 strain. Also, this same study observed that the heterozygous *Dnd1*-Δ mutant allele has a decreased number of sperm and is generally less fertile in Sv/129 mice than in the C57BL/6 strain. A study by Ogonuki et al. (2010) investigated the intracytoplasmic sperm injection (ICSI) procedure and effects of freeze and thawing for sperm and oocytes in different mouse backgrounds (Ogonuki et al., 2010). They detected less ICSI efficiency for round spermatids frozen and thawed in Sv/129, by 0.3% compared to C57BL/6 background. The

study of (Adham et al., 2001) indicated that the knockout of transition protein 2 (*Tnp2*) shows normal fertility on a hybrid background (C57BL/6- Sv/129), although in Sv/129 mice was completely infertile (where mice had teratospermia). Another example is deletion of the *Smcp* gene (Sperm Mitochondrion-Associated Cysteine-Rich Protein) in hybrid background (C57BL/6- Sv/129) was fertile, while Sv/129 was infertile despite producing sperm. In the Sv/129 mouse strain, knocking out the (*Smcp*) gene reduced motility and sperm's ability to penetrate oocytes. Recently, many studies identified different outcomes and phenotypes with different mouse backgrounds, and it is most likely that those backgrounds have diverse genetic modifiers (Lu et al., 2012; Miyata et al., 2020). In addition, a mutation of *Usp26* (ubiquitin-specific peptidase 26) results in sterility or subfertility when crossed with the DBA/2 mouse strain. However, fertile mice were bred when the same mutation was introduced into a C57BL/6 background. This example confirms that genetic background modified the effects of the *Usp26* mutation on male reproductive capacity (Sakai et al., 2019). These data shows the important of the genetic background in mouse phenotype.

Identifying the exact stage of any block in spermatogenesis can help describe and interpret what a gene's effects in spermatogenesis are. During spermatogenesis, the mouse testis undergoes 12 stages of seminiferous tubule development. Staging can be done following the Meistrich and Hess's (2013) mouse staging roadmap. Therefore, in order to study the effect of *Rbmxl2* knockout on mouse testis development, I had to learn how to stage testes. With the assistance of Professor Ian Adams from the University of Edinburgh, I began learning how to stage mouse testes. We staged wild-type mouse testes using the mouse testes charts shown in figure 3.1. In addition, I used some markers to check for apoptosis and DNA breakage, including an antibody that recognises gamma-H2AX to identify leptotene, zygotene, and the sex body in pachytene (Meistrich and Hess, 2013). However, the staging of mice is a complex process. For example, if mice are not producing sperm, it is hard to identify the first wave of spermatogenesis without round spermatids. Furthermore, it can occasionally be challenging to stage dying cells like pachytene during the first wave.

A complete absence of germ cells is called Sertoli cell only syndrome, and this can also apply to individual seminiferous tubules (called SCO tubules). Presence of SCO tubules suggests a dramatic defect in germ cell development. There are two types of SCO: primary and secondary. Primary refers to when prenatal damage occurs during migration to the male gonads. In contrast, secondary SCO is postnatal damage to a healthy testis (Gashti et al., 2021).

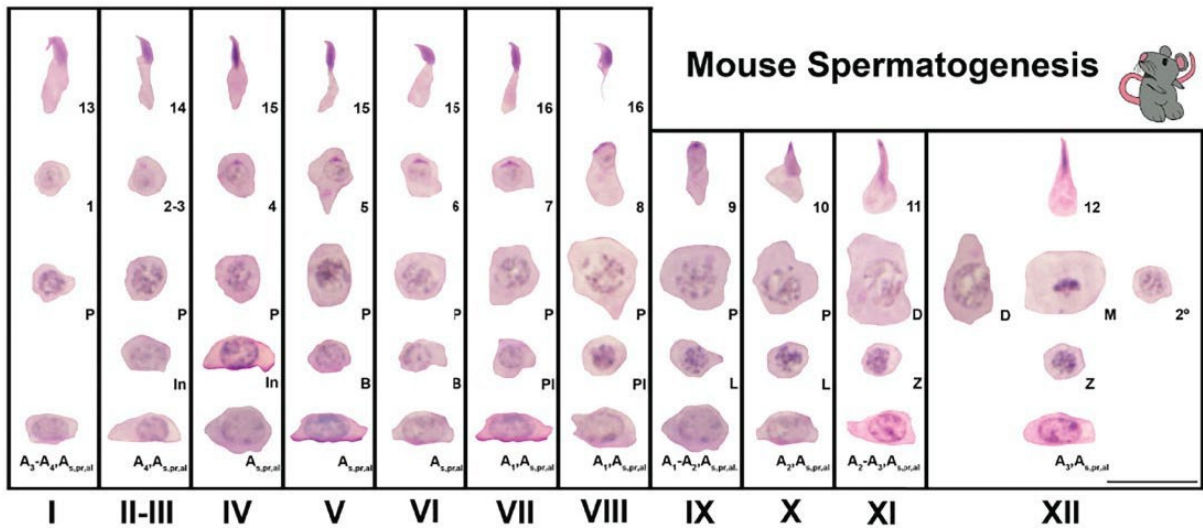


Figure 3. 1: Mouse testes staging chart from stage 1 to 12. Taken with modification from (Ruthig & Lamb, 2022).

3.2 Hypothesis:

I hypothesised that *Rbmxl2* knockout alleles may have cause different phenotypes in different mouse strains. To test this the C57BL/6 with the *Rbmxl2* knockout originally studied in the eLife paper (Ehrmann et al., 2019) was backcrossed to a different background: Sv/129. We were interested in learning whether *Rbmxl2* is essential during the diplotene stage or earlier/later, and whether RBMXL2 expression would be the same in both backgrounds.

3.3 Aims:

- 1-To investigate whether *Rbmxl2* knockout mice have different phenotypes between different strains.
- 2-To investigate if different mouse might strains have different RBMXL2 protein expression patterns.
- 3-To investigate whether already known RBMXL2 target genes in meiosis respond to *Rbmxl2* depletion in a strain-specific background.

3.4 Results

3.4.1 *Rbmxl2* knockout mice do not have different testis weight between the C57BL/6 and Sv/129 strains.

Dr. Ingrid Ehrmann did a preliminary study to see if the *Rbmxl2* knockout phenotype in a C57BL/6 background differs from the same knockout allele in Sv/129 using mouse testes weights. Using Ingrid's data, I did a testes weight comparison using a two-way ANOVA between the two mouse strains, shown in figure 3.2. The first thing to notice is that there is a strong change in testis size for homozygote versus wild type mice in both strains – hence the *Rbmxl2* knockout likely does cause male infertility and germ cell depletion on the Sv/129 background. However, the difference in testis weight between the Sv/129 and C57BL/6 mouse strains was insignificant, suggesting they might arrest at the same time.

Combined testes weight

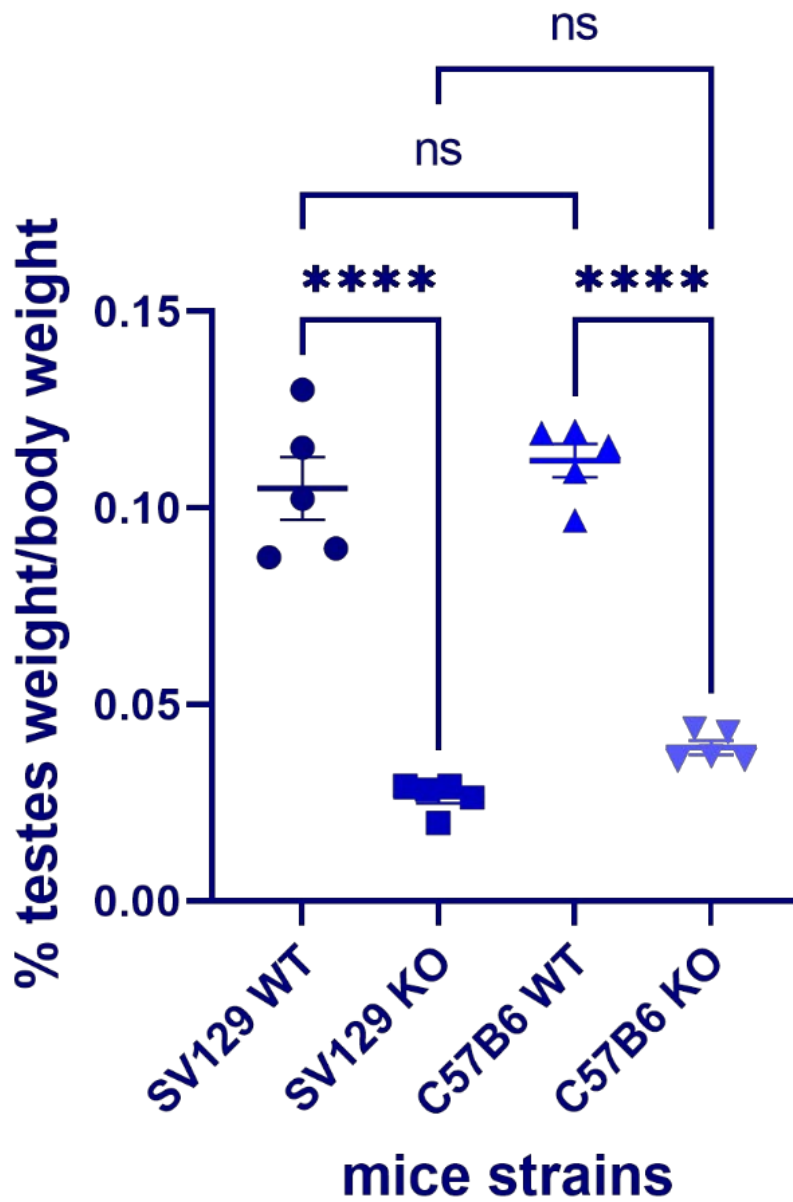


Figure 3. 2: Testis weight comparison between the *Rbm12* knockouts in the C57BL/6 and Sv/129 strains (n=5). A significant difference between the wild-type and knockout mice within both strains were seen. However, there were no significant differences in the testis weight of the KO between the mouse strains. Statistical analysis (two-way ANOVA) was done using Prism GraphPad. ns=not significant ****= $P \leq 0.0001$.

3.4.2 *Rbmxl2* knockout mice have an increased content of “Sertoli Cell Only” tubules in the Sv/129 strain compared to C57BL/6

Next I started staining testis sections with H&E (Haematoxylin and eosin) to check for possible more subtle differences between the two mouse strains. My first observations at X20 magnification suggested that testis morphology in the Sv/129 strain was more disrupted, and has more Sertoli cell only tubules than C57BL/6 sections as shown in figure 3.3. No round spermatids or elongating spermatids were detected in either mouse strain.

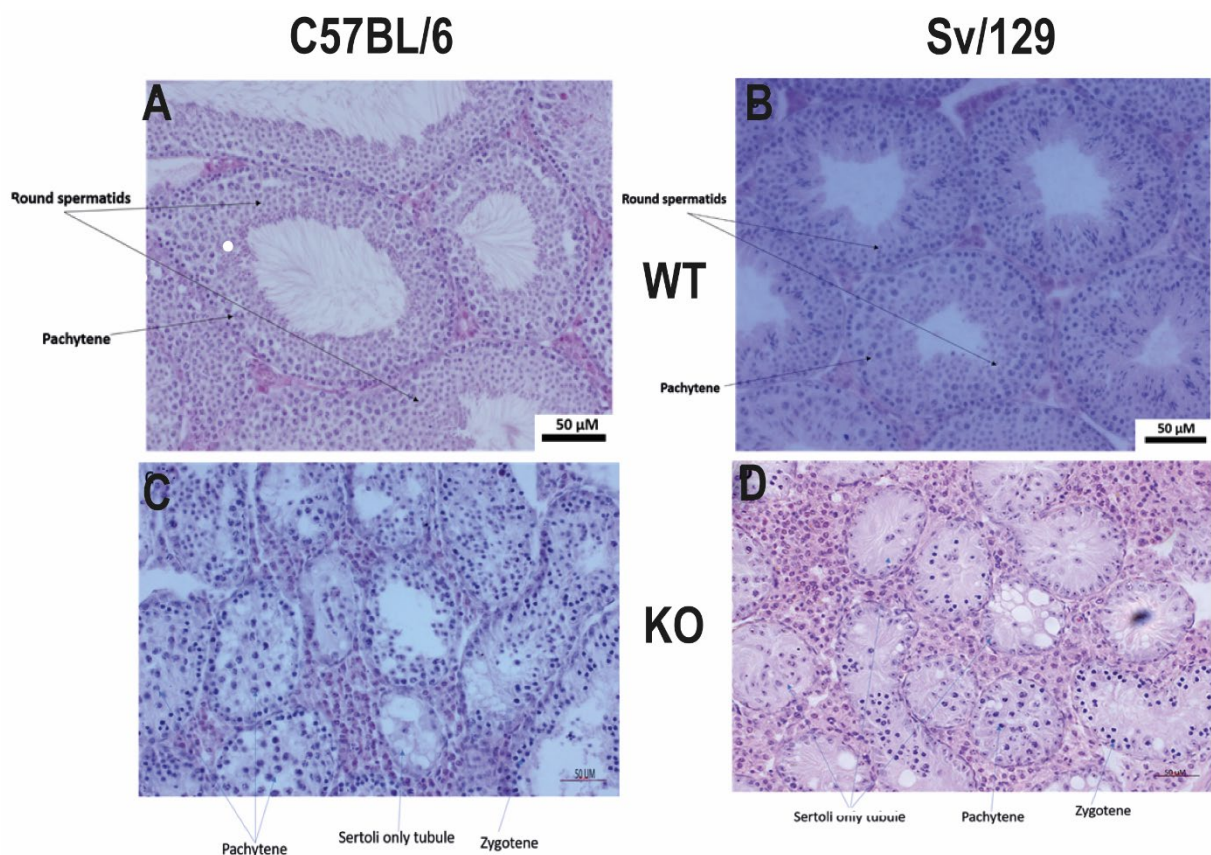


Figure 3. 3: Histological comparison of testis development for *Rbmxl2* knockout mice between the Sv/129 and C57BL/6 mouse strains at X20 magnification (n=1). (A) A micrograph of H&E stained testis from a wild type C57BL/6 mouse shows round spermatids and pachytene cells. (B) A micrograph of H&E stained testis from a wild type Sv/129 mouse shows normal round spermatids and pachytene cells. (C) A micrograph of H&E stained testis from a C57BL/6 mouse shows that one tubule appears to have only Sertoli cells. (D) In a micrograph of H&E stained testis from a Sv/129 mouse, the tubules seemed more empty compared to C57BL/6, with more frequent occurrence of tubules that have only Sertoli cells.

SCO tubules were seen in three adult *Rbmxl2* knockout mice in the Sv/129 strain background, numbered 109, 54 and 98. All of these mice were adult males. However, individual 109 was around 8 months, which might explain the higher number of empty tubules seen in comparison to the other replicates (Mice numbers 54 and 98 were 3 months old). Higher magnification images of histological staining with H&E seen below in the figure 3.4.

Due to the COVID-19 pandemic, we had to cull all SV/129 mouse strains during the lockdown. Consequently, upon our return to normal laboratory operations, there was a considerable delay in procuring new knockout mice. Therefore, I resorted to utilizing an archived set of testes samples, specifically from individual number 109, which had been fixed in Bouins. These testes were found to consist solely of Sertoli cell-only tubules upon microscopic examination. With Dr. Ingrid Ehrmann's help, we confirmed that the mouse was eight months old. Therefore, no power tests were executed due to the unavailability of appropriate replicate samples for this analysis.

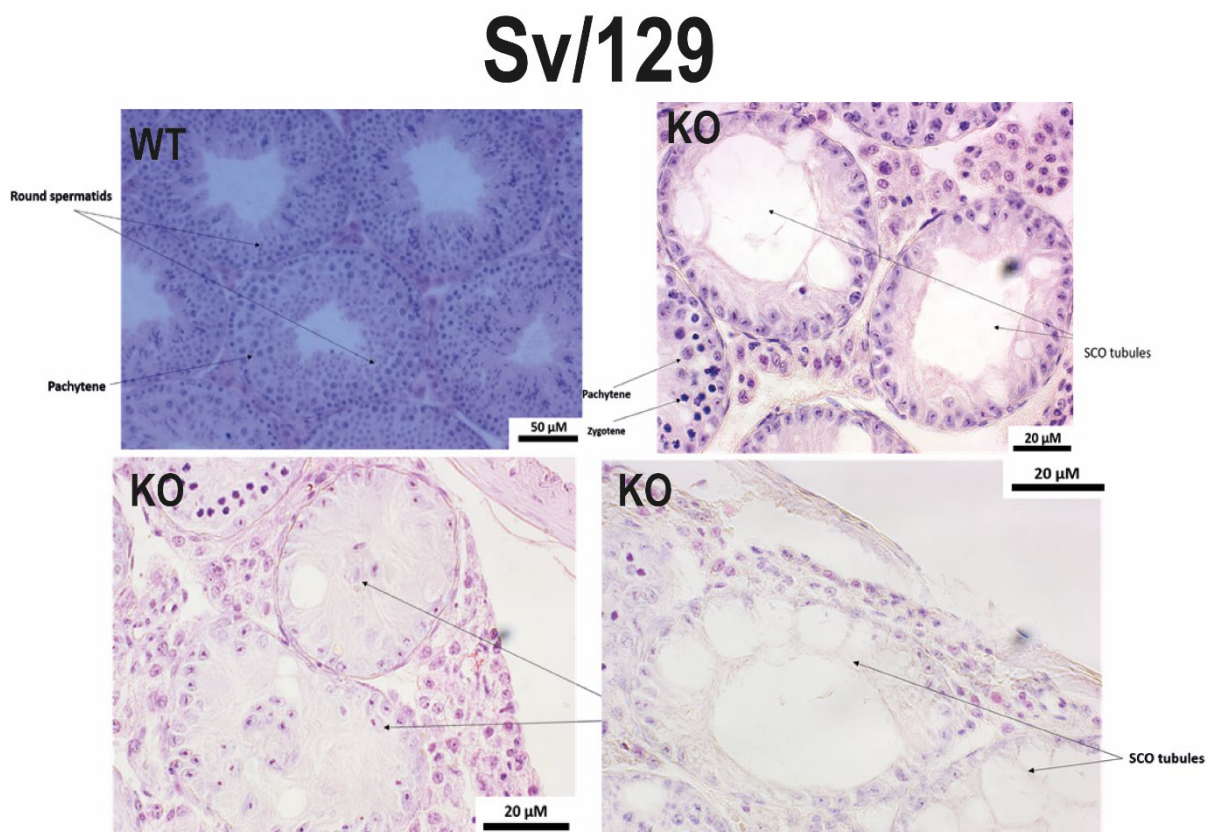


Figure 3. 4: Histological section of WT and *Rbmxl2* knockout mice testis (n=3). WT exhibit all cell types. Unlike RBMXL2 knockout individual number 109, 54 and 98 shows Sertoli cells only tubules. Exact tubule stages cannot be identified due to complete cell loss. No quantification analysis done beyond describing the data (qualitative assessment**).**

3.4.3 *Rbmxl2* deletion causes a Stage IV arrest in the Sv/129 background compared to C57BL/6

Before investigating the stage of germ cell arrest within the Sv/129 background, I confirmed Ingrid's findings reported in her eLife paper (Ehrmann et al., 2019). She reported that knocking out *Rbmxl2* in C57BL/6J resulted in the arrest at diplotene (stage XI). Consequently, I repeated the analysis as demonstrated below 3.5. At a superficial level, the C57BL/6 tubules lack round and elongating spermatids, and the pachytenes and diplotenes are not quite right. However, diplotene was clearly detected, and another tubule showed pachytene cells. These data confirm the stage XI defect in C57BL/6 mice caused by *Rbmxl2* deletion (cells up to diplotene).

C57BL/6

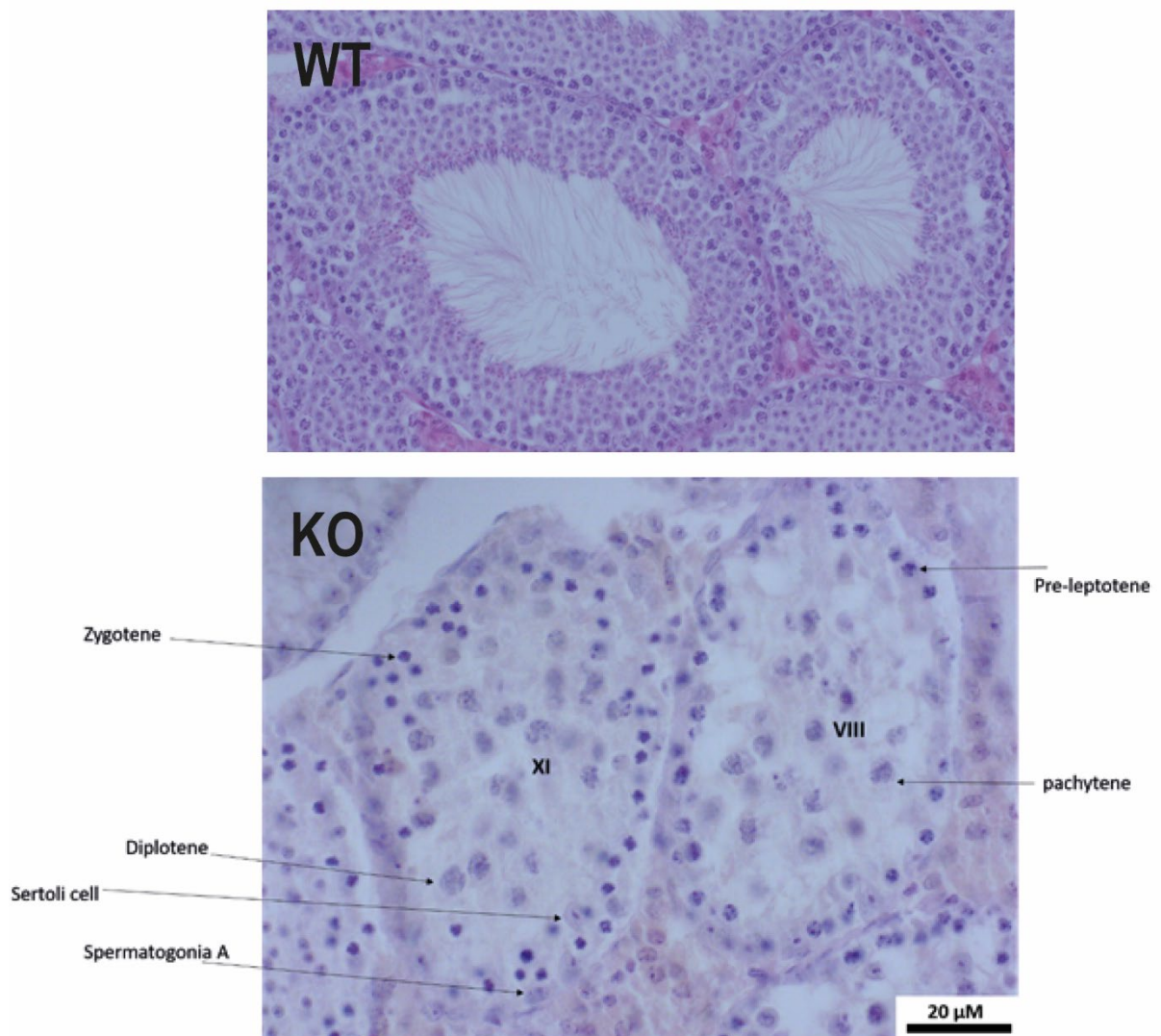


Figure 3. 5: Histological section of WT and Rbmxl2 knockout (C57BL/6). Wild type shows all cells type while the KO shows arrest at diplotene. The left side tubule shows a stage XI tubule, where diplotene and zygotene were present. Preleptotene and pachytene are seen on the right side tubule, indicating a stage VIII tubule.

In order to find the precise germ cell defect in the Sv/129 strain testis, more comprehensive staging of mouse testis sections was done, as explained below. However, defining the exact stage is more complex when the testis has a spermatogenesis defect. In the figure below, a seminiferous tubule showed early pachytene cells, suggesting it corresponds to stages I-III (exact staging is difficult without the later spermatid stages). However, there are some dying/odd-looking cells. This might be due to the fact that this mouse is around 8 months old, and so there could be secondary effects.

Staging mouse testis was done through Hematoxylin and Eosin (H&E) staining. Hematoxylin stained cell nuclei blue, while eosin stained the cytoplasm and extracellular matrix pink. Initially, I aimed to identify one or two layers of spermatids. The first layer of spermatids represented stages IX-XII, whereas the second layer represented stages I-VIII. Following this, I proceeded to examine the cells within the tubules, starting with spermatogonia A, which had prominent round nuclei. Primary spermatocytes displayed larger and more irregularly shaped nuclei. Sertoli cells were characterised by their elongated nuclei located along the basement membrane. As I investigated further, I identified the preleptotene stage, which had round nuclei. The leptotene and zygotene stages were identifiable through the appearance of thread-like chromatin in the nuclei, with zygotene cells appearing smaller and darker. Finally, the pachytene stage showcased densely stained chromosomes, similar to the Diplotene stage but smaller in size.

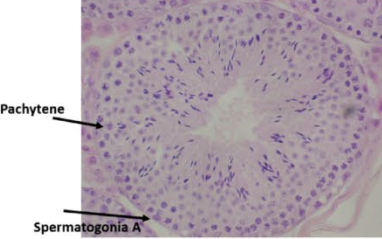
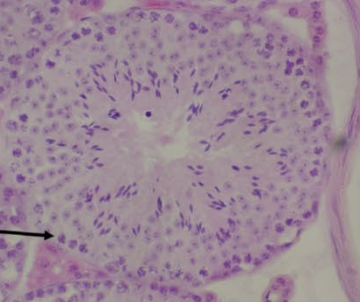
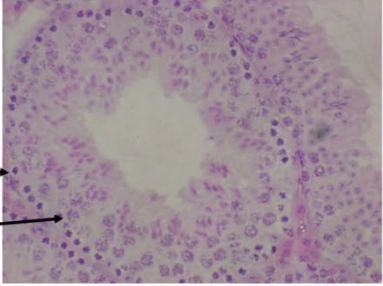
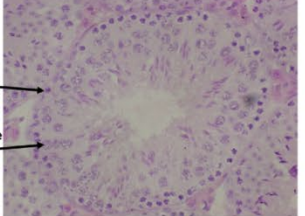
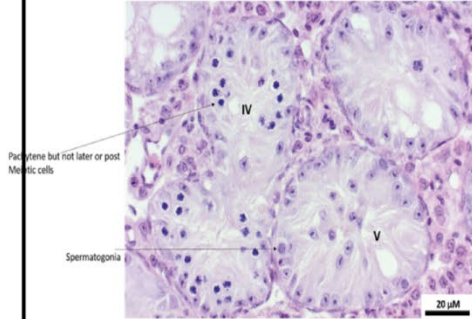
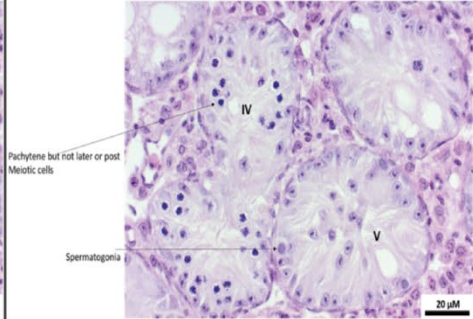
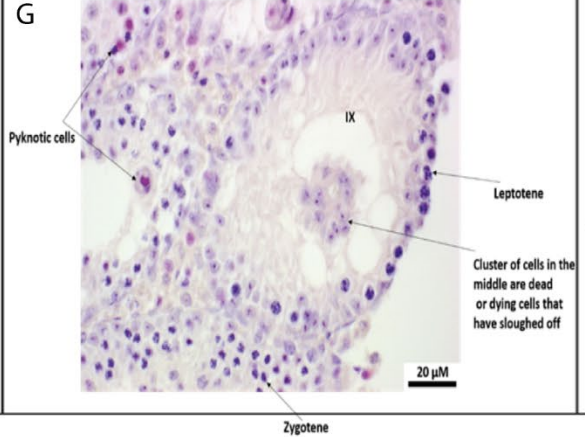
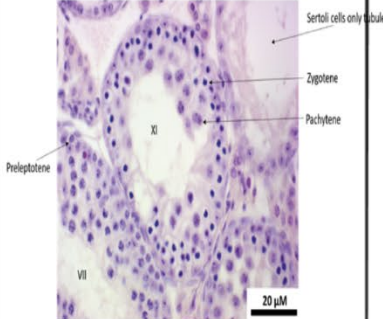
	Stage I-III	Stage V-VI	Stage IX	Stage XI
WT Sv/129	<p>A</p>  <p>Pachytene</p> <p>Spermatogonia A</p>	<p>B</p>  <p>Pachytene</p>	<p>C</p>  <p>Leptotene</p> <p>Pachytene</p>	<p>D</p>  <p>Zygotene</p> <p>Diplotene</p>
RBMXL2 ko Sv/129	<p>E</p>  <p>Pachytene but not later or post Meiotic cells</p> <p>Spermatogonia</p> <p>IV</p> <p>V</p> <p>20 μM</p>	<p>F</p>  <p>Pachytene but not later or post Meiotic cells</p> <p>Spermatogonia</p> <p>IV</p> <p>V</p> <p>20 μM</p>	<p>G</p>  <p>Pyknotic cells</p> <p>IX</p> <p>Leptotene</p> <p>Cluster of cells in the middle are dead or dying cells that have sloughed off</p> <p>20 μM</p> <p>Zygotene</p>	<p>H</p>  <p>Sertoli cells only tubule</p> <p>Zygotene</p> <p>Pachytene</p> <p>Preleptotene</p> <p>XI</p> <p>VII</p> <p>20 μM</p>

Figure 3. 6: Histological section of Rbmxl2 knockout (Sv/129) vs. wild type showing different stages from (I-IX). **A-D)** Sections of testis from WT control mice. **E)** shows early pachytene cells, suggesting it corresponds to stages I-III (exact staging is difficult without the later spermatid stages). There are also some dying cells. This advanced phenotype might be due to the fact that this mouse is around 8 months old. The tubule on the right side were Sertoli cells only tubules. **F)** Individual number (109) showing Stage IV and Stage V tubules. This section highlights two tubules. One showed pachytene cells but no diplotene cells, suggesting it corresponds to stage IV. These pachytene cells looked abnormal, they are smaller and darker than usual compared to above Wild type. A second tubule showed spermatogonia suggesting it corresponds to a stage V-VI tubule in which cells died after pachytene. **G)** Individual number (98) showing a stage IX tubule with Zygotene and Leptotene cells, and with some dying sloughing cells. There were no diplotenes, suggesting earlier cell death compared to WT. The presence of leptotene suggest this is a stage IX tubule. The tubule on the extreme left contains some pyknotic (dying) cells; however, no cells beyond pachytene were detected however, diplotene and round spermatid were detected in the WT. **H)** Individual number 54 showed tubules with preleptotene and pachytene, suggesting stage XI; and tubules with zygotene and pachytene cells, suggesting stage XI which indicates the cells apoptosis during pachytene. A Sertoli cells-only tubule was identified as well unlike the WT (**A-D**). Appendix 3 has the RBMXL2 KO stages.

Summary of staging analysis. The *Rbmxl2* deletion in a C57BL/6 background shows a Stage XI arrest. This meant the last stage to be detected was diplotene of meiotic prophase. Hence this is what was expected for Sv/129 if it was the same as C57BL/6. My examination of Sv/129 testis slides concluded that no diplotene or later cells were detected in this background. In many slides, pachytene cells were abnormal as seen in the figures above. Some cells were clustered in the middle or about to slough, suggesting apoptosis. Sertoli cells only tubules were difficult to stage due to the complete loss of cells. On all *Rbmxl2* knockout slides in the Sv/129 background I only saw up to pachytene cells; no cells were seen beyond that, suggesting that the later cells were dead or did not progress beyond pachytene. This is hence an earlier meiotic defect in the Sv/129 background.

3.4.4 C57BL/6 and Sv/129 strains have similar patterns of expression of RBMXL2 protein over germ cell development

The above analyses showed that deletion of the *Rbmxl2* gene caused apparent histological differences between germ cell developments in the two mouse strains. The Sv/129 arrest at pachytene (stage IV arrest) was different to that at diplotene (stage XI) reported in C57BL/6 (Ehrmann et al., 2019).

Therefore, one possibility to explain this difference might be that RBMXL2 protein is expressed differently within testes from the Sv/129 and the C57BL/6 strains. In order to test this hypothesis I did immunofluorescence experiments as discussed below.

Ehrmann et al. (2008) demonstrated that RBMXL2 protein is expressed during meiosis and right after meiosis (Ehrmann et al., 2008). But no detailed staging was done. I stained wild-type testes from the two mouse strains to do more detailed histological analysis of RBMXL2 expression in these two backgrounds 3.7 and 3.8 and negative control (with no secondary antibody to show the specificity of the RBMXL2 antibody as seen in the figure 3.9.

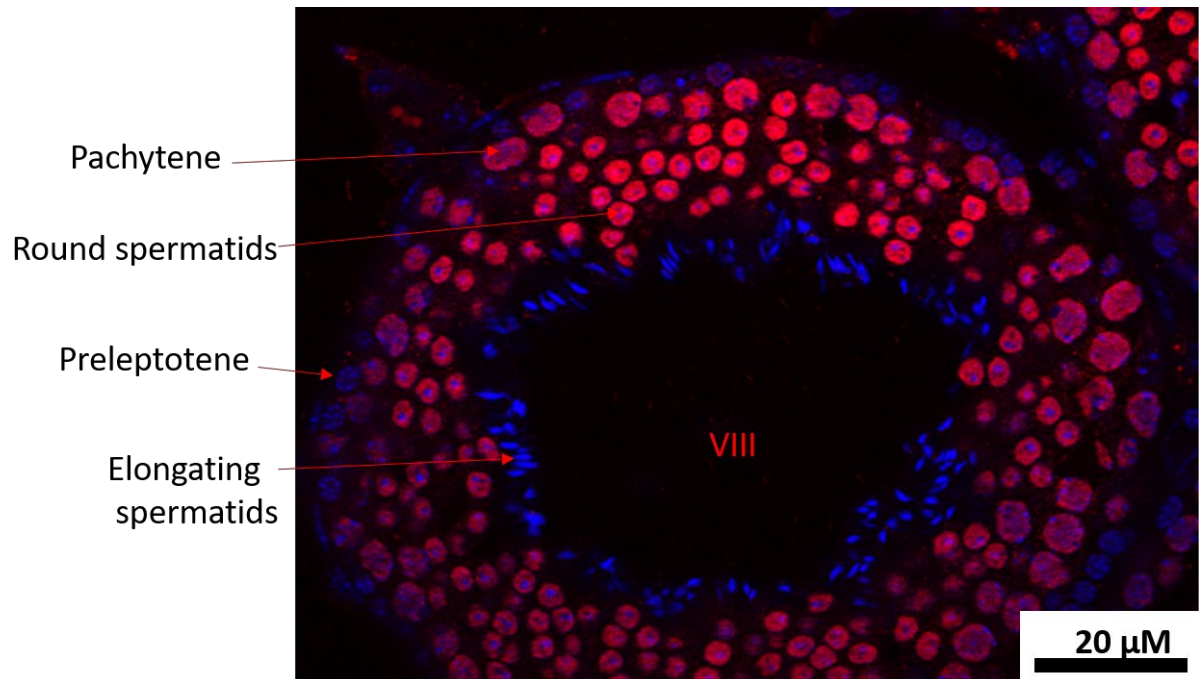


Figure 3. 7: Fluorescent staining of RBMXL2 protein in wild type testis (Sv/129). RBMXL2 antibodies stained pachytene cells and round spermatids. Preleptotene and elongating spermatids were not stained and so, are in blue counterstain DAPI.

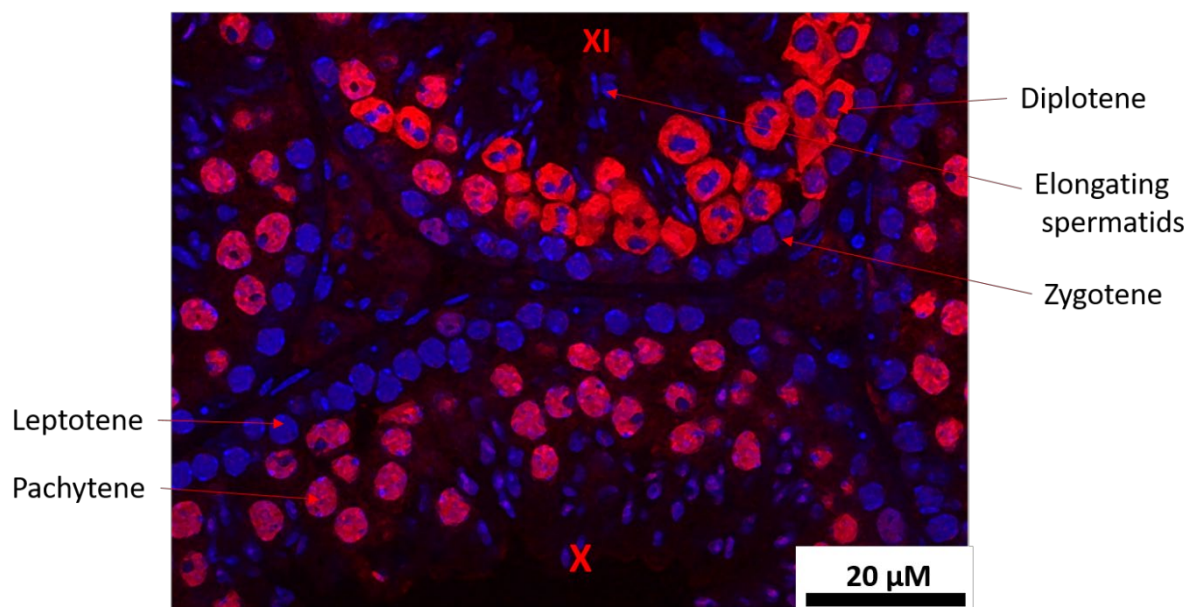


Figure 3. 8: Fluorescent staining of wild type testis (Sv/129). Antibodies specific to RBMXL2 stained diplotene cells undergoing the first meiotic division—the presence of diplotene in the above tubule suggests stage XI. While, the other tubule shows pachytene with leptotene cells suggesting stage X.

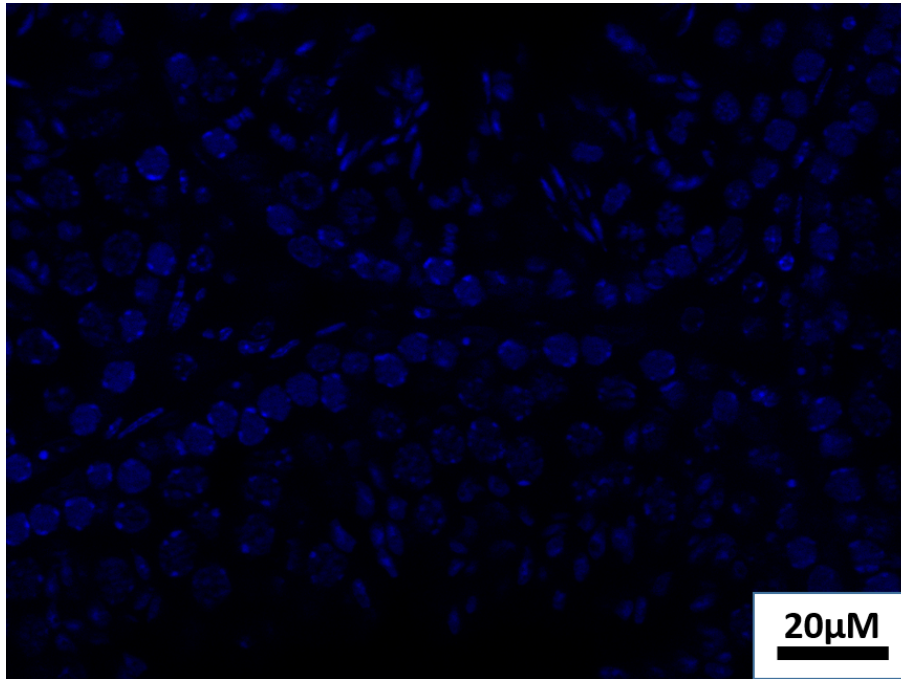


Figure 3. 9: Fluorescent staining without secondary antibody of RBMXL2 on wild type testis (Sv/129). This shows the specificity of RBMXL2 antibody compared to background auto-fluorescence that is apparent after fixation in PFA and Bouins.

First I examined the distribution of RBMXL2 protein using indirect immunofluorescence and using an antibody that recognises RBMXL2 protein. RBMXL2 protein expression was detected in pachytene and diplotene and round spermatids, i.e. during and right after meiosis only. The presence of two layers of spermatids (round and elongating) suggests that the tubule in Figure 3.7 is stage I to VIII. Therefore, finding another marker would be useful to identify the exact mouse stages. Another example shows the diplotene cells in the wild-type mouse testis. Diplotene presence is implicated with stages XI to XII as seen in figure 3.8. However, the first layer of the tubule needs to be differentiated using another marker to give an exact stage.

3.4.5 Staining of wild type testes with γ H2AX as a marker for the sex body

Next to solve the staging more precisely, I used an antibody that recognises γ H2AX, which is a marker used to check for double-strand breaks (DSBs). γ H2AX stains the sex body in pachytene and has a different distribution within Zygotene, Preleptotene, and Leptotene. Therefore, to confirm the presence of pachytene cells, analysis of sex body staining would be helpful. However, in the figure 3.10 below, I only detected the preleptotene stage. Zygotene can be identified with γ H2AX staining, as it exhibits smaller and darker cells with thread-like nuclei. On the other hand, leptotene is larger than zygotene and can be differentiated by the presence of a single layer of spermatids, unlike preleptotene, which is characterised by two layers of spermatids as seen below.

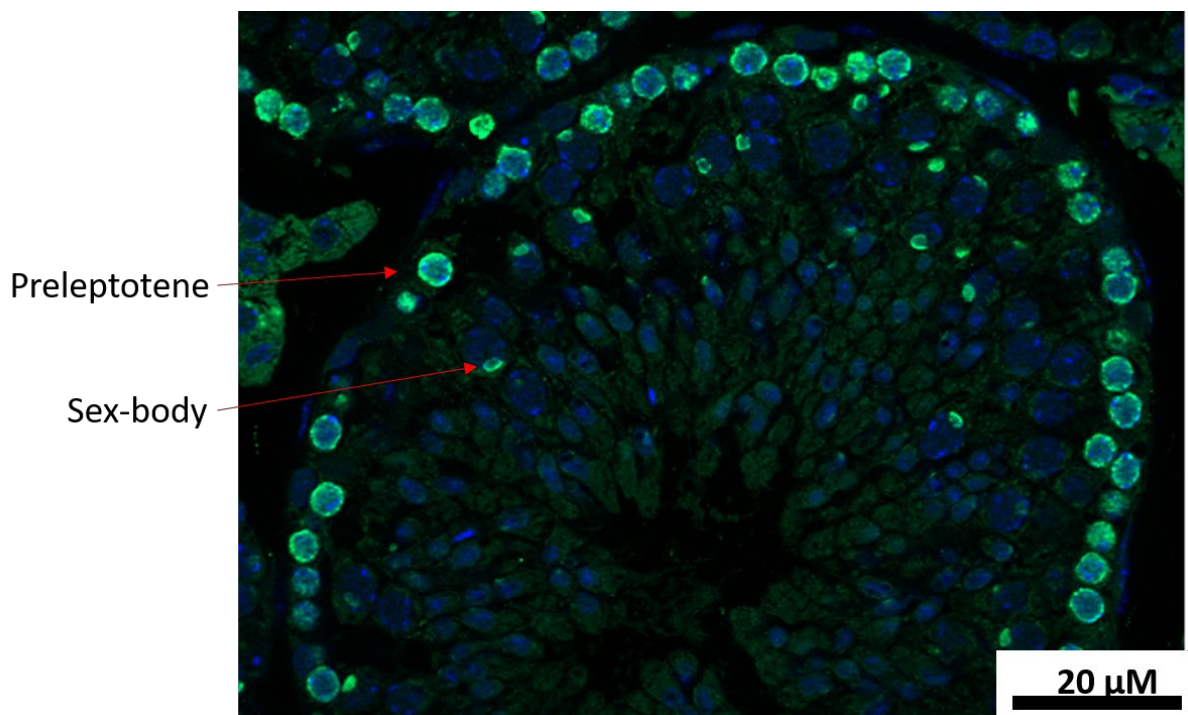


Figure 3. 10: Fluorescent staining of wild type testis (Sv/129). γ H2AX stained preleptotene and sex body of pachytene in green.

One possibility for the difference in phenotype between strains might be that RBMXL2 protein is expressed differently within testes from the Sv/129 and the C57BL/6 strains. In order to test this hypothesis I did double immunofluorescence to detect RBMXL2 protein and γ H2AX.

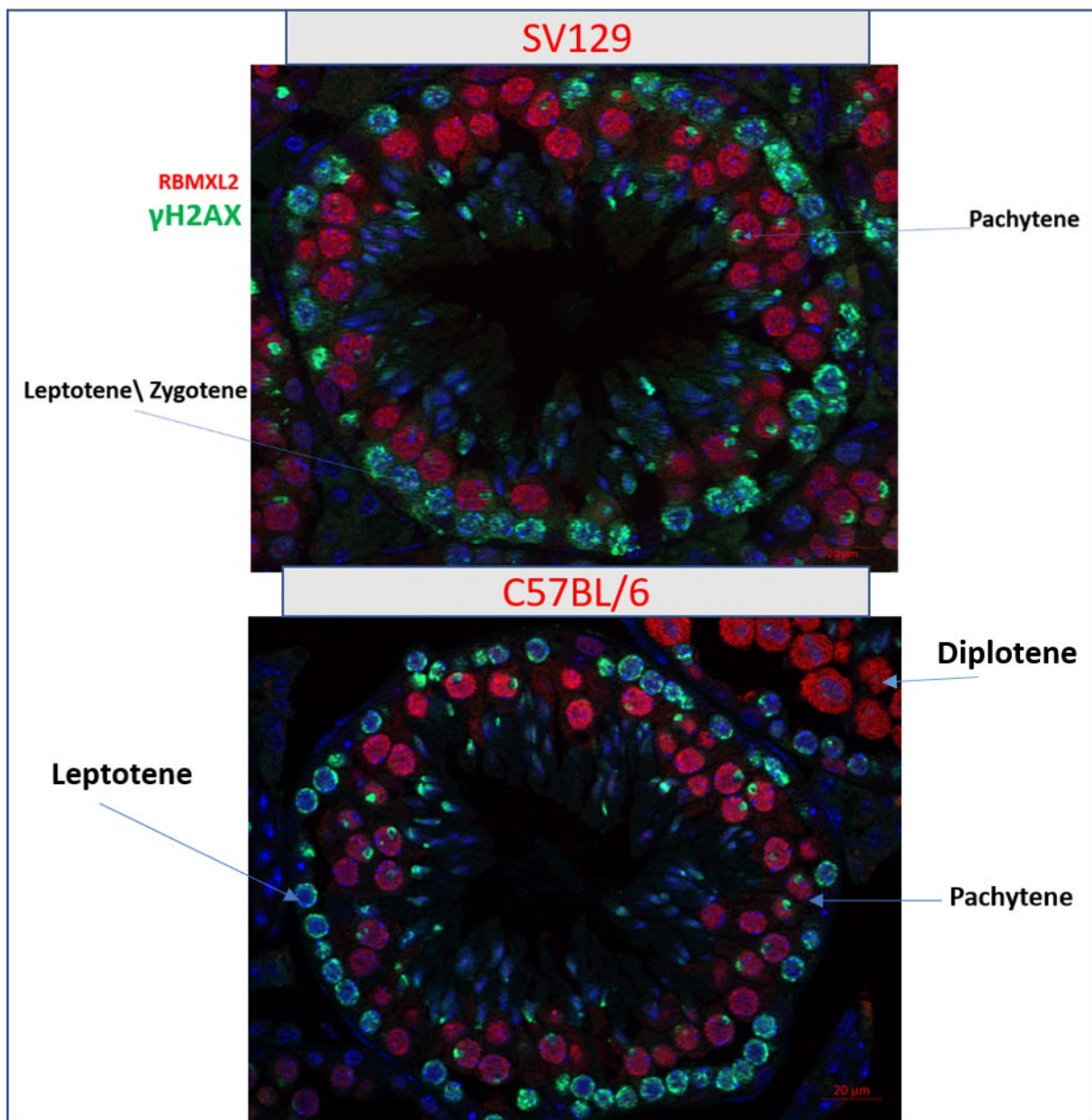


Figure 3. 11: Fluorescent staining of two wild-type mouse testes (Sv/129 and C57BL/6), stage X. RBMXL2 protein is pseudocoloured red, and γH2AX is pseudocoloured green. The RBMXL2 protein cell type expression was identical in both tubules.

Figure 3.11 shows that the RBMXL2 staining pattern is similar for the two wild-type mouse strains (Sv/129 and C57BL/6). Both tubules shown are at stage X, containing leptotene and pachytene cells (sex body stained with green). RBMXL2 protein is pseudocoloured red, while elongating spermatids are in blue (the blue represents the counterstain with DAPI). The expression of RBMXL2 protein between cell types is similar between the two mouse strains. Staining wild-type mouse testis Sv/129 with an antibody recognising RBMXL2 reveals its expression in pachytene cells that were also stained with γ H2AX. A similar expression of RBMXL2 was observed in C57BL/6. RBMXL2 expression was also detected in diplotene cells in the right tubule. γ H2AX displayed a pattern comparable to that of Sv/129. RBMXL2 protein had a similar pattern of expression in wild type C57BL/6 and Sv/129 mouse testes. Therefore, the differences in phenotype is not caused by a change in RBMXL2 protein expression pattern.

3.4.6 Investigating important genes regulated by RBMXL2 over the first wave of spermatogenesis

3.4.6.1 *Kdm4d* has an earlier cryptic splicing between in Sv/129 over the first wave of spermatogenesis.

After finding that the *Rbmxl2* knockout allele had a different phenotype within the Sv/129 strain compared to C57BL/6, yet RBMXL2 protein expression patterns are similar in both mice strains, we anticipated that RBMXL2 protein might have different gene targets in the two different strains, or perhaps Sv/129 has earlier splicing defects than C57BL/6.

From the literature, *Rbmxl2* is expressed only in the testis and found to suppress cryptic splice sites of some important genes in meiosis (Ehrmann et al., 2019). In order to test if there was an earlier defect in cryptic splicing in the Sv/129 background I took advantage of the first wave of spermatogenesis in the neonatal mouse which is synchronous. Spermatogenesis is a continuous, complex process. During the first wave of spermatogenesis, spermatogonial cells go through mitosis and then two stages of meiosis to produce sperm (Bellve et al., 1977).

I used the time course of neonatal days 12, 14, 18, and the adult mouse testes dissected from both strains. I then used RT-PCR to compare RNA processing patterns within the wild type and the *Rbmxl2* knockout backgrounds.

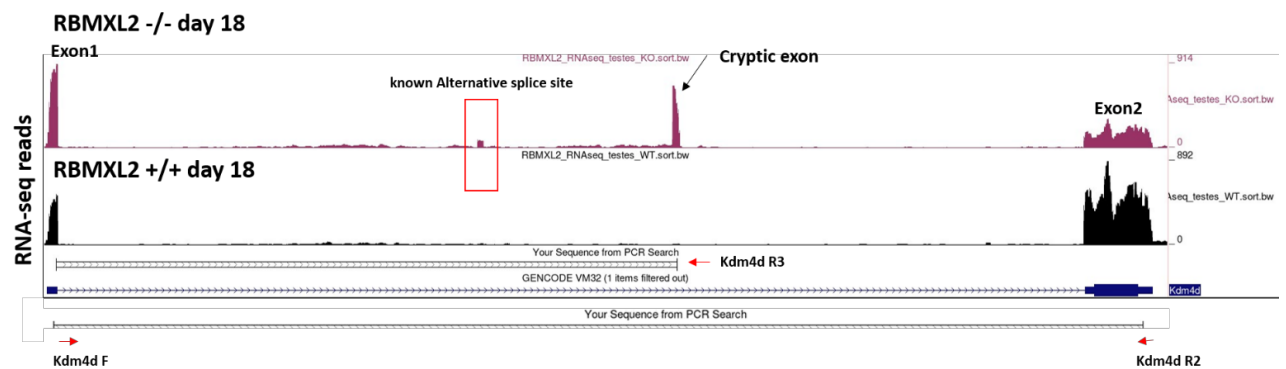


Figure 3. 12: RNAseq track of the *Kdm4d* gene on the UCSC website. RNAseq data was generated from *Rbmxl2* knockout testes and wild type (Ehrmann et al., 2019). The inclusion of the cryptic splice exon is shown in the *Rbmxl2* knockout mouse testis. Multiplex primers were designed to analyse splicing, including a forward primer in exon 1 of the *Kdm4d* gene and a reverse primer in exon 2; another reverse primer was designed to include the region between the forward and the cryptic splice site. The primers and RNAseq data were taken from the Ehrmann et al. (2019) paper. This screenshot was taken from the UCSC genome browser (<http://genome.ucsc.edu>).

The figure above 3.12 was generated using RNA-Seq reads from the Ehrmann et al. 2019 paper. The *Kdm4d* gene encodes a protein involved in chromatin regulation (Zoabi et al., 2014). The *Rbmxl2* knockout testis includes a cryptic exon not normally included in the wild type. This cryptic exon is sometimes joined to the downstream exon 2, and sometimes used as a terminal exon

I used RT-PCR to test splicing of *Kdm4d* in the two different mouse strains. The figure below 3.13 shows a diagram of what we would expect if RBMXL2 was expressed versus if RBMXL2 was knocked out. cDNA was generated by Dr. Ingrid Ehrmann in the Elliott lab. I did the RT-PCR using the primers from (Ehrmann et al., 2019) and loaded my samples in the qiaxcel machine (Qiagen) following the protocol described in the materials and methods chapter (2.6.12).

Kdm4d gene:

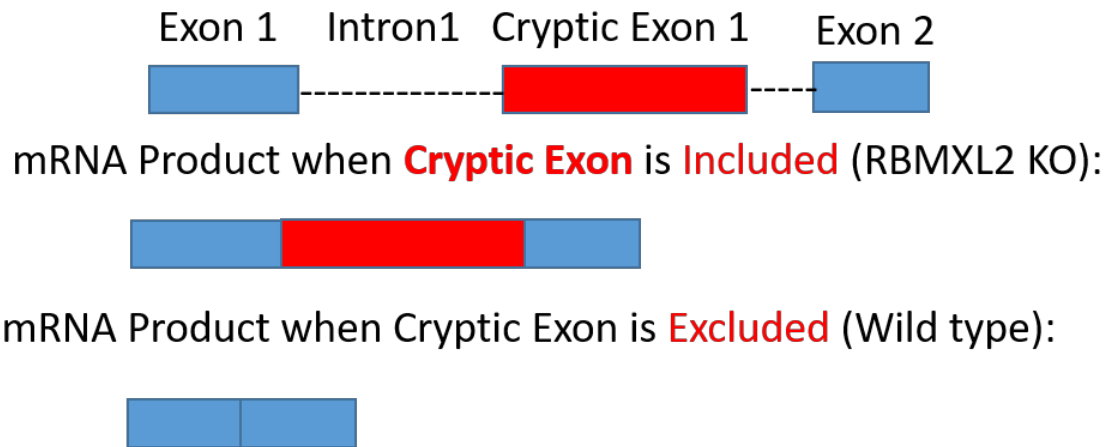


Figure 3. 13: Diagram depicts the mRNA product with and without RBMXL2. Boxes represent exons, while a dashed line represents introns. Because of the cryptic exon, the RBMXL2 knockout would produce two products. In the presence of RBMXL2, however, only one product would be visible.

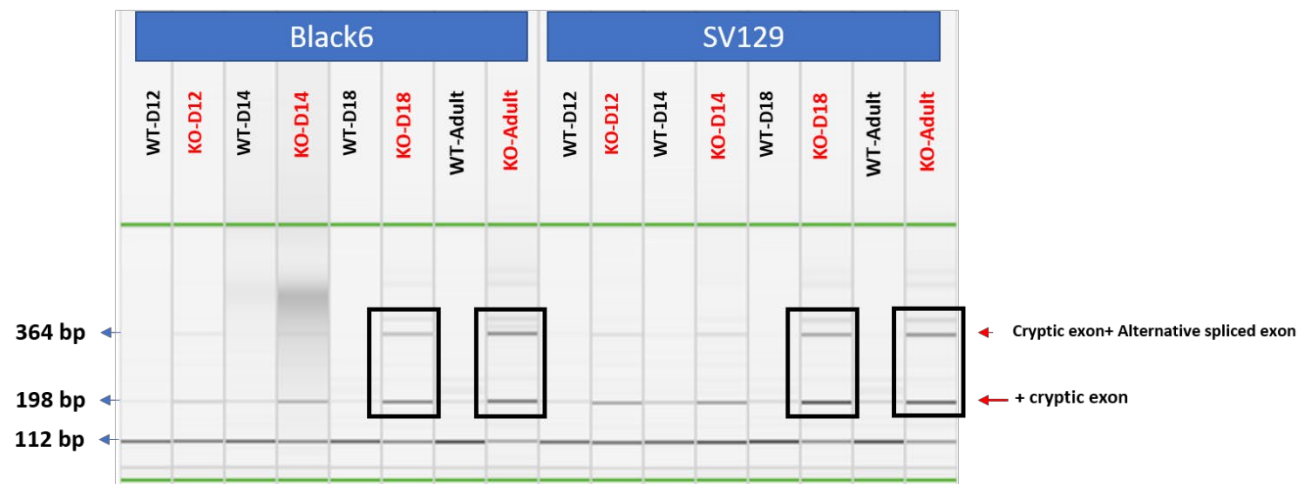


Figure 3. 14: Capillary gel electrophoresis showed the splicing patterns of *Kdm4d* between the wild type and *Rbmxl2* knockout testes in both mouse strains (n=1). The results detect three bands: the first band (112 bp) is where exon1 and exon2 get spliced. The band in the middle is (198 bp), indicating the inclusion of the cryptic splice site that was not seen in the wild-type conditions in both strains. Finally, the third band (364bp) represents splicing of the cryptic exon and the annotated alternative exon.

The RT-PCR results (done once per sample) showed that mainly one band was seen on the wild type, indicating the splicing of exon 1 and exon 2. However, in the *Rbmxl2* knockout, increased use of the cryptic splice site was detected, and splicing of the annotated alternative exon that was less evident in the wild type. Only in the *Rbmxl2* knockout were the two bands strongly observed (198 bp and 364 bp). On days 12 and 14, faint bands corresponding to cryptic splicing were observed in the *Rbmxl2* knockout that were not present on the same days in the wild type. Furthermore, a smear was observed in the day 14 knockout sample on a C57BL/6 background, as depicted above. Consequently, a quantitative analysis was conducted in Figure 3.15 to illustrate the differences in expression between the two mouse strains. However, it is essential to perform this RT-PCR in triplicate before conducting any statistical analysis.

Much less cryptic splicing was observed in wild type mouse testis at all ages. The expression of *Kdm4d* in wild type testis was consistent in both mouse strains. However, there was evidence of increase cryptic splicing of *Kdm4d* in the Sv/129 background at days 12 and 14 that were less obvious in the *Rbmxl2* knockout of C57BL/6 background.

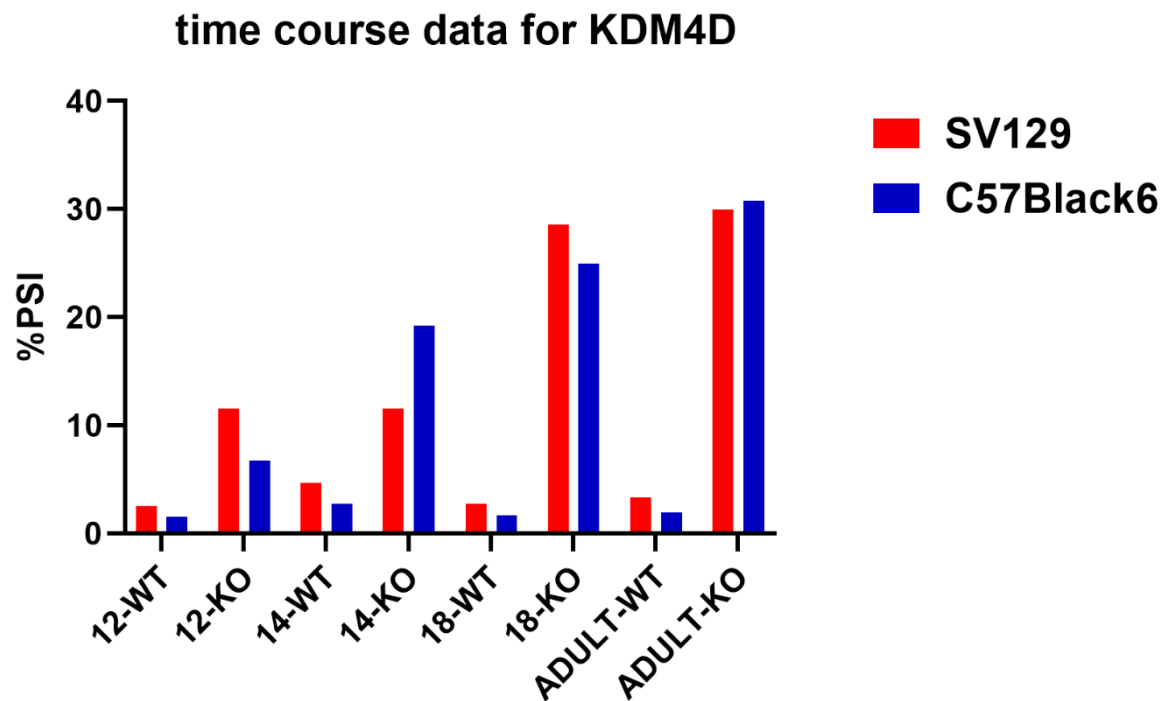


Figure 3. 15: KDM4D expression was investigated in this time course study comparing C57BL/6 and SV/129 backgrounds (n=1). It is important to note that each sample was analysed only once, and there was an observed smear on day 14 in the RBMXL2 knockout sample, indicating the need for repetition.

3.4.6.2 *Meioc* has an earlier cryptic splicing in Sv/129 over the first wave of spermatogenesis.

The second example gene I investigated was *Meioc*. MEIOC is a conserved meiosis-specific protein in all metazoans with an uncharacterised coiled-coil domain known as PF15189 (Soh et al., 2017; Wojtas et al., 2017). *Meioc* is expressed in male and female mouse germ cells during meiotic prophase I. *Meioc* is expressed exclusively in the testes of male mice (postnatal and adult). A knockout of *Meioc* revealed germ cells initiating meiosis normally and reaching the preleptotene stage before germ cell arrest, while some cells developed into the zygotene stage. However, *Ccna2* misexpression was seen in spermatocytes that were missing *Meioc*. This showed that the cell-cycle program did not correctly move to the meiotic prophase (Abby et al., 2016; Soh et al., 2017).

Knockout of *Rbmxl2* leads to the use of an exon within exon 5 of *Meioc* that is not normally used in the wild type. Mice lacking *Rbmxl2* have weak splice sites included in exon 5 of *Meioc*, which leads to the production of a shorter *Meioc* protein isoform (Ehrmann et al., 2019). The figure below 3.16, was generated using RNA-Seq reads from the Ehrmann et al. 2019 paper. The *Rbmxl2* knockout testis includes an exon within exon 5 of the *Meioc* gene that is not seen in the wild type. An inclusion of weak splice sites on 5' and 3' on exon 5 of *Meioc* represents the exon. *Rbmxl2* is known to suppress the use of these cryptic splice sites within ultra-long exons such as *Meioc* exon 5.

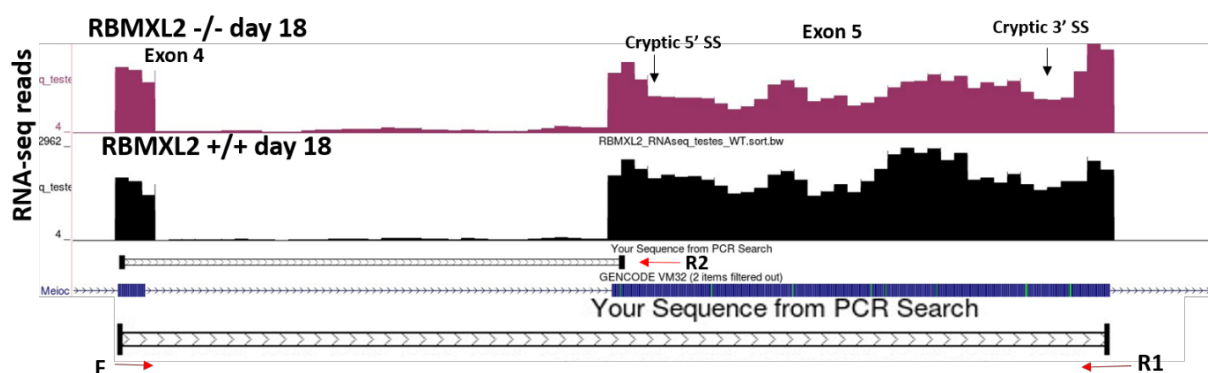


Figure 3. 16: RNAseq track of the *Meioc* gene on the UCSC website including RNAseq data generated from *Rbmxl2* knockout testes and wild type. The inclusion of the exon is shown in the *Rbmxl2* Knockout. Multiplex primers were designed to include forward in exon 4 of the *Meioc* gene and reverse to include exon 5; another reverse primer was designed to amplify the region between the forward and the cryptic 5' splice site.

I used RT-PCR to test splicing of *Meioc* in the two different mouse strains. cDNA was generated by Dr. Ingrid Ehrmann in the Elliott lab. I did the RT-PCR using the primers from (Ehrmann et al., 2019) and loaded my samples in the Qiaxcel machine (Qiagen) following the protocol in the materials and methods chapter section (2.6.12).

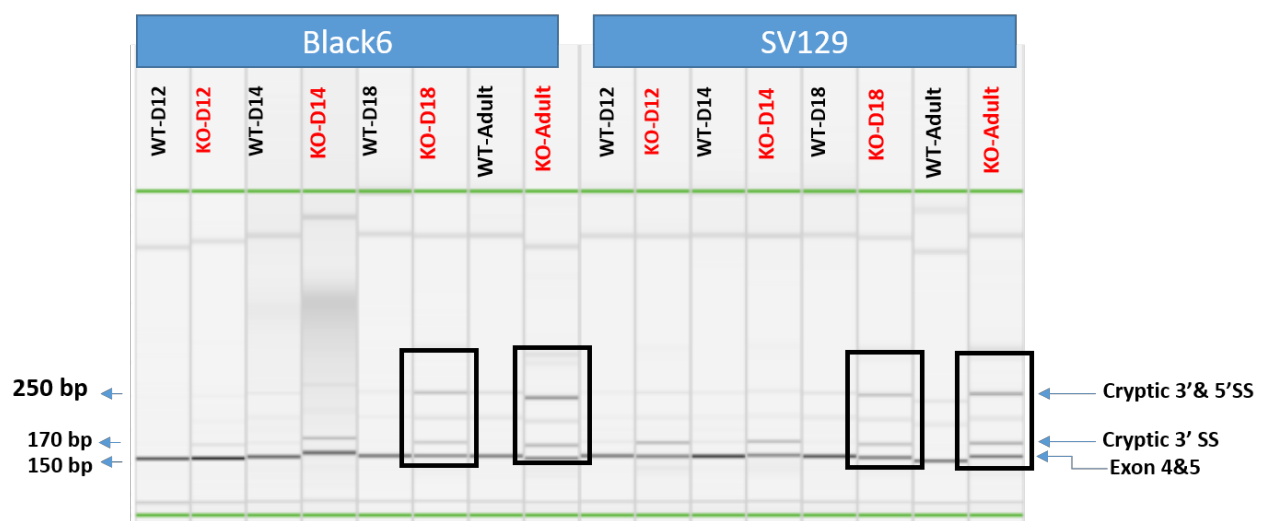


Figure 3. 17: Gel electrophoresis showed the splicing of *Meioc* between testes from wild type and *Rbmxl2* knockout testes in both mouse strains (n=1). The results indicate three bands. The smallest band (150 bp) is where Exon4 and Exon5 get spliced. The band in the middle (170 bp) represents the inclusion of the cryptic 3' splice site that was not seen in the wild-type conditions in both strains. Finally, the third band (250 bp) included the cryptic 3' and 5' splice sites.

The RT-PCR results showed that only one band was strongly seen in the wild type testis, indicating the inclusion of exon 4 and exon 5. However, in the *Rbmxl2* knockout, the use of the 3' and 5' cryptic splice sites were observed. Only in the *Rbmxl2* knockout were the two cryptic bands strongly observed (170 bp and 250 bp). On days 12 and 14, faint bands representing cryptic splicing were observed that were not present on the same days in the wild type. While days 12 and 14 in the wild type showed one band only at 150 bp, indicating exon 4 and 5 inclusion as usual, three bands were observed in both the 18-day-old knockout mice and the adults (150 bp, 170 bp and 250 bp). These cryptic splice bands were expressed at a high level in the day 18 knockout testis. While again, only one band was seen in the wild type of days, 18 and adults. The splicing pattern of *Meioc* showed an increase in cryptic splicing on days 12 and 14 in the Sv/129 background compared to in the *Rbmxl2* knockout of the C57BL/6 background. This splicing pattern demonstrates cryptic splicing following *Rbmxl2* deletion. This suggests stronger cryptic splicing changes occurred earlier in the Sv/129 mouse strain than in the C57BL/6 background. The earlier splicing defects observed in the Sv/129 may explain the developmental arrest observed in the Sv/129 testis. A Moreover, as shown earlier, there was a detectable smear in the day 14 knockout sample from the C57BL/6 background. Subsequently, Figure 3.18 includes a statistical analysis demonstrating the disparities in expression between the two mouse strains. Nevertheless, it is crucial to emphasize the necessity of conducting the RT-PCR in triplicate before proceeding with any statistical assessments.

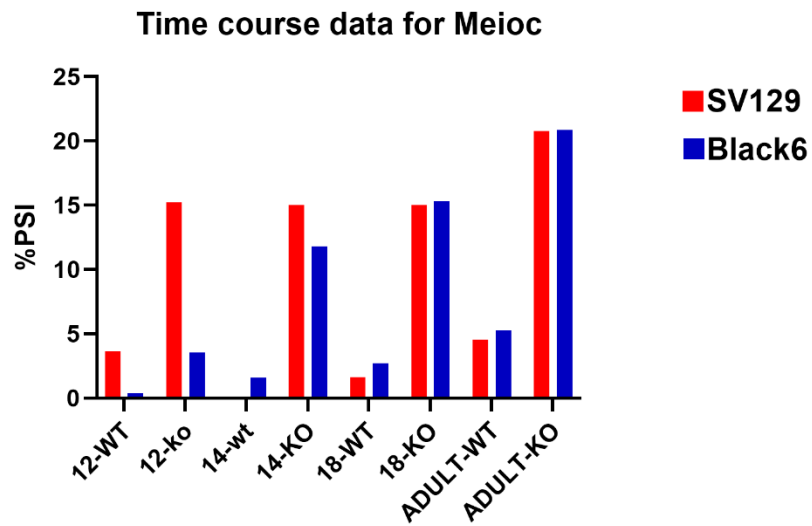


Figure 3. 18: *Meioc* expression was investigated in this time course study comparing C57BL/6 and SV/129 backgrounds (n=1). It's worth highlighting that each sample underwent a single analysis, and a smear was observed in the RBMXL2 knockout sample on day 14, suggesting the necessity for further replication.

3.4.7 *Rbmxl2* knockout has additional DNA breakage in the first wave of spermatogenesis of Sv/129 mice strain.

I saw earlier that there had been a pachytene arrest in the Sv/129 strain after the *Rbmxl2* deletion. We suspected that this might be caused by a synaptic defect that leads to cell apoptosis. Cells create several DNA double-strand breaks to start homologous recombination, making it possible for homologous chromosomes to share genetic information and ensuring that chromosomes are split correctly during meiosis (Murakami & Keeney, 2008). The presence of Sertoli cells only tubules in testes from old mice indicate apoptosis had cleared germ cells, which means something wrong was happening in the early stages. Therefore, in order to look into the earlier arrest at the pachytene stage in the Sv/129 mouse strain, I stained three younger *Rbmxl2* knockout mouse testes with γ H2AX antibodies and compared them to the wild type. Three testes were taken from mice at 21 days old to capture the first wave of spermatogenesis. Analysis of these Sv/129 21 day samples indicated that spermatogenesis did proceed to diplotene in the first wave of spermatogenesis. Evidence of unresolved DNA double-strand breaks were seen during meiosis in the metaphase stage, as seen in the figure below 3.20. However, these DNA double-strand breaks were not detected in the wild-type testes. These images demonstrate that during the first wave, cells enter metaphase, with unresolved DNA damage.

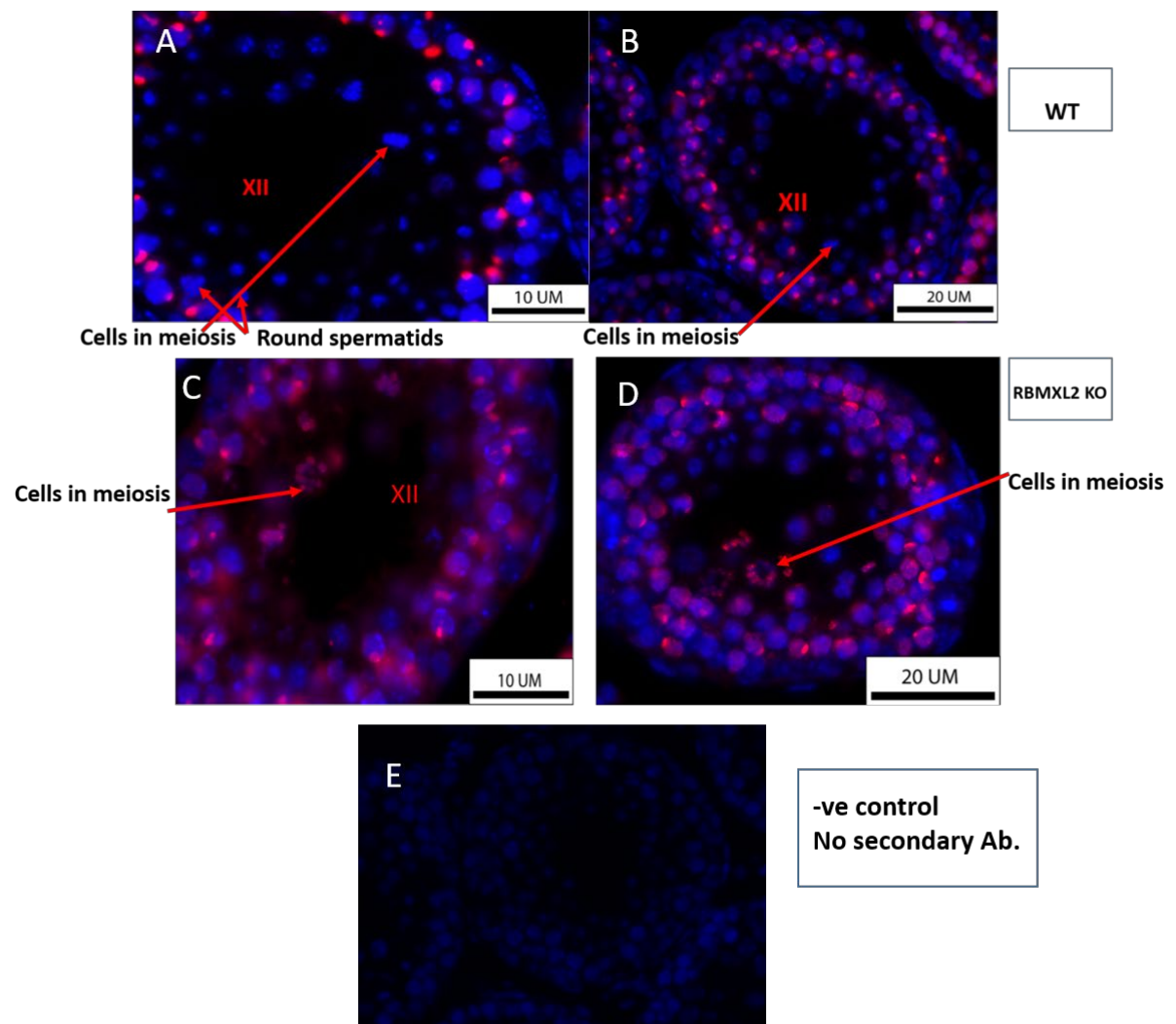


Figure 3. 19: Analysis of DNA damage in SV/129 Wild type vs. *Rbmxl2* knockout testis (n=3). Fluorescent staining of wild type mouse testes tubule stage XI, at (40X) and (63X) magnification (**A and B**), stained with γ H2AX revealed normal round sex-body in pachytene cells (no extra dots outside the sex body). Moreover, fluorescent staining of *Rbmxl2* knockout testis examined at (40X) and (63X) magnification (**C and D**), and stained with γ H2AX, displayed some abnormal sex-body staining in pachytene cells (dots outside the sex body and worm-shaped sex body) and additional γ H2AX staining in metaphase. Last picture showing negative control were no secondary antibody added (**E**).

Moreover, staining with γ H2AX antibodies showed some dots that were observed outside the sex body of pachytene cells at stage XI in the *Rbmxl2* Knockout. Unresolved DNA damage were seen as snake- or worm- shaped sex body in *Rbmxl2* knockout. In contrast, in the Wild type mouse testes, the sex bodies were round. In metaphase, additional γ H2AX staining was detected. In order to confirm the DNA damage abnormality that was see in the *Rbmxl2* knockout, three wild testes mouse were used as a control. In stage XI, no abnormal sex bodies were detected in pachytene cells. Wild-type testes were stained with γ H2AX antibodies, similar to the *Rbmxl2* knockout testes. Two mouse testes showed normal pachytene staining in stage XII, and cells entered metaphase without the additional double-stranded DNA breakage that was detected in the *Rbmxl2* knockout. Moreover, some round spermatids were detected, indicating the normal spermatogenesis process in the wild type, as seen in the figure 3.19 above.

3.5 Discussion:

3.5.1 *Rbmxl2* knockout mice have different phenotypes between the Sv/129 and C57BL/6

Deletions or mutations in mouse models affect fertility depending on the genetic background—for example, mutations in the nuclear export factor called *NXF-2*. The C57BL/6 *Nxf2* knockouts are subfertile but do not show meiotic arrest. In contrast, *Nxf2* knockout in mixed backgrounds resulted in sterility with meiotic defects in a third of the males (Pan et al., 2009). *Rbmxl2* knockout on a C57BL/6 background was previously reported to have a stage XI seminiferous tubule block, equivalent to when diplotene cells appeared. However, all earlier stages are histologically normal (Ehrmann et al., 2019). An earlier histological phenotype was seen when the *Rbmxl2* knockout C57BL/6 mice were backcrossed into the Sv/129 background, a stage IV arrest was identified. However, earlier stages (I-III) were histologically normal. Beyond stage IV, a significant loss of pachytene cells were seen. Sv/129 *Rbmxl2* knockout mice testes have more Sertoli Cell Only tubules than the C57BL/6 mice. Below is Figure 3.20, which illustrates the diplotene arrest in C57BL/6 mice compared to the stage IV arrest on the Sv/129 background.

Histological analysis showed that there were also fewer Sertoli cell-only (SCO) tubules in the C57BL/6 background. However, many SCO tubules were seen in the Sv/129 background. Ehrmann et al., (2019) reported, sometimes round spermatids are detected in the C57BL/6 background. But on Sv/129 no round spermatids were observed. The testis-to-body weight ratios are comparable in the *Rbmxl2* Sv/129 knockouts and the C57BL/6 background, despite the early cell loss that was seen in the Sv/129 *Rbmxl2* knockout.

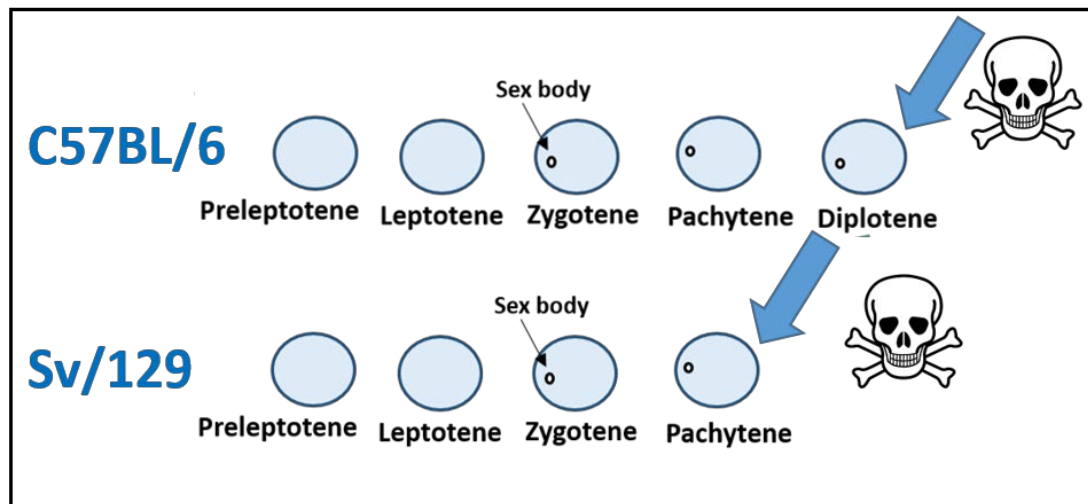


Figure 3. 20: A diagram illustrates the meiotic arrest between the two mouse strains, Sv/129 and C57BL/6, after knockout of the *Rbmxl2* gene.

3.5.2 RBMXL2 protein expression pattern is similar in both mouse strains

Rbmxl2, a testis-specific gene, is expressed during and immediately after meiosis (Ehrmann et al., 2019). Previous staining on two mouse strains (Sv/129 and C57BL/6) revealed that RBMXL2 protein is expressed in the leptotene, pachytene, diplotene, and round spermatids. Both mouse strains exhibited the same expression pattern. Consequently, there is no difference in protein expression that might explain the differences in phenotype.

3.5.3 Defective splice patterns appear earlier in Sv/129 in the first wave of spermatogenesis

RBMXL2 protein is expressed only in the testis. *Rbmxl2* knockout causes mouse sterility. RBMXL2 protein was reported to preserve the correct splice isoforms of a number of genes such as *Kdm4d*, *Meioc* and *Esco1* among others, by preventing the cryptic splice sites. To analyse if there is a change in RNA processing caused by the *Rbmxl2* knockout I analysed *Kdm4d* and *Meioc*, both genes involved in meiosis (Ehrmann et al., 2019). The expression of both genes increased with the older mouse (day 18 and adults), and the cryptic splice sites for *Kdm4d* and exon of *Meioc* were evident with the *Rbmxl2* knockout. Moreover, stronger cryptic splicing defects appeared to be observed earlier in Sv/129 compared to C57BL/6 background.

3.5.4 Investigating the first wave of spermatogenesis of Sv/129 *Rbmxl2* knockout

After analysing adult testes from the two mouse strains, I first confirmed Dr. Ingrid Ehrmann's findings from Professor Elliott's lab. She reported an arrest at diplotene in C57BL/6 background caused by *Rbmxl2* knockout. However, some random round spermatids were detected. In contrast, knockout of *Rbmxl2* in the Sv/129 background caused pachytene arrest with no sperm detected. SCO tubules were seen in Sv/129 but were less frequently observed in older mice on the C57BL/6 background. To see if there is more germ cell death I did investigate the first wave of spermatogenesis in the Sv/129 *Rbmxl2* knockout, with the prediction that there might be less cell death earlier in neonatal development. I did detect later stages than Stage IV in the Sv/129 background at day 21 compared to adult testis. I also detected increased DNA breakage in stage XI and additional double-stranded DNA breaks in metaphase I that were not seen in wild type mice. Normal spermatocytes show gamma-H2AX immunofluorescent staining during leptotene. This is when homologous sister pairs come together to form the synaptonemal complex. After that, γ H2AX staining begins to vanish from autosomal chromosomes but retains on the X-Y sex body in pachytene (Celeste et al., 2002).

Arrest at stage IX, during meiosis when pachytene cells are present, was reported before; for instance, mice lacking *H2AX* showed a pachytene arrest (Celeste et al., 2002). Another example is the knockout of *Mlh1* in mice shows arrest at pachytene (Edelmann et al., 1996). In addition, knockouts of genes like *Brca1* and the RNA-binding protein *Hnrnpk* were found to cause infertility and pachytene arrest (Xu et al., 2003, 2022). Furthermore, extra double-stranded DNA breakage was observed in the *Rbmxl2* knockout that was not seen in the wild type. All DNA breakage must be resolved before initiating metaphase I. Therefore, these additional DNA breakages imply a problem with DNA repair mechanisms that could result in abnormal sperm and infertility. In these circumstances, more investigation may be needed to find out what caused the DNA to break and what effects it could have.

3.6 Chapter summary:

The testis-specific *Rbmxl2* gene is essential for mouse fertility; mice lacking *Rbmxl2* on a C57BL/6 background exhibited diplotene arrest. In this Chapter I investigated the effect of *Rbmxl2* on the Sv/129 mouse strain and identified an earlier arrest at pachytene. In order to investigate this further, staining with γ H2AX during the first wave of spermatogenesis in *Rbmxl2* knockout mouse testes revealed extra dots in pachytene cells other than the sex body, indicating extra DNA damage in the Sv/129 background. In addition, additional DNA breakage was observed during metaphase I that was not present in the wild type. These observed histological differences require further investigation. Moreover, RBMXL2 protein expression was comparable between the two mouse strains. . However, the cryptic splicing defects were seen earlier in Sv/129 compared to C57BL/6 background. Analysis of further mice are needed to confirm this result.

Chapter 4: Identification of novel targets of RBMXL2-RNA protein contact using iCLIP

4.1 Introduction:

Previous work and my own work in chapter 3 showed (Ehrmann et al., 2019) that RBMXL2 is only expressed in the testis during and right after meiosis. This findings was proved by histological staining with antibodies against RBMXL2 in my previous chapter. The results of an RBMXL2 HITS-CLIP experiment was published in the same 2019 paper but using one mouse testis. The HITS-CLIP technique generated a library of only a few thousand reads mapped to the mouse genome. Accordingly, the primary purpose of this chapter was to identify and characterise direct RBMXL2 targets in mouse testes through individual crosslinking and immunoprecipitation (iCLIP) in experiments performed in quadruplicate, so increasing the number of replicate testes and the number of sampled mice.

Konig et al. (2010) developed iCLIP to identify global protein-RNA interactions at single-nucleotide resolution as shown in the figure 4.1 below. Due to a lag in the reverse transcription reaction at the site of cross linking, the HITS-CLIP protocol produces truncated cDNAs (Urlaub et al., 2002). The iCLIP protocol was developed and tested to address this issue (Konig et al., 2011; Huppertz et al., 2014). The idea behind iCLIP is to crosslink RNA- protein interactions in vivo and then immuno-precipitate the resulting complexes using an antibody that recognises RBMXL2 protein. SDS-PAGE is then used to separate the immunoprecipitated complexes, and a LI-COR machine is used to visualise these complexes after Western blotting. According to the protocol, the protein-RNA complexes must be taken off the membrane at size of between 150 and 400 nt. Complexes are then purified and subjected to reverse transcription in order to generate a cDNA library, which will then be amplified and submitted for sequencing to generate a transcriptome map of RBMXL2 protein RNA targets using the iMAps pipeline. This will allow the identification and characterisation of novel RNA targets of RBMXL2 protein in mouse testis on a broad transcriptional scale.

During my PhD, with the assistance of my colleague Chile Siachisumo, I optimised the iCLIP experiment to identify RBMXL2 global RNA targets, and to generate a list of genes that bind RBMXL2 protein to perform gene ontology (GO) analysis. Also, with Dr Ivailo Yonchev, I performed K-mer analysis to identify RNA sites highly enriched in RBMXL2 protein crosslinks. I then took RBMXL2 protein's direct RNA targets and combined them with RNA-seq data from

Rbmxl2 knockout testes (Ehrmann et al., 2019) to compare how splicing changes correlate with sites of RNA-protein interaction, and analyse how RBMXL2 protein controls splicing. I also utilised the minigene assay to investigate splicing regulation of some of these targets, and how these are controlled by RBMX and RBMXL2.

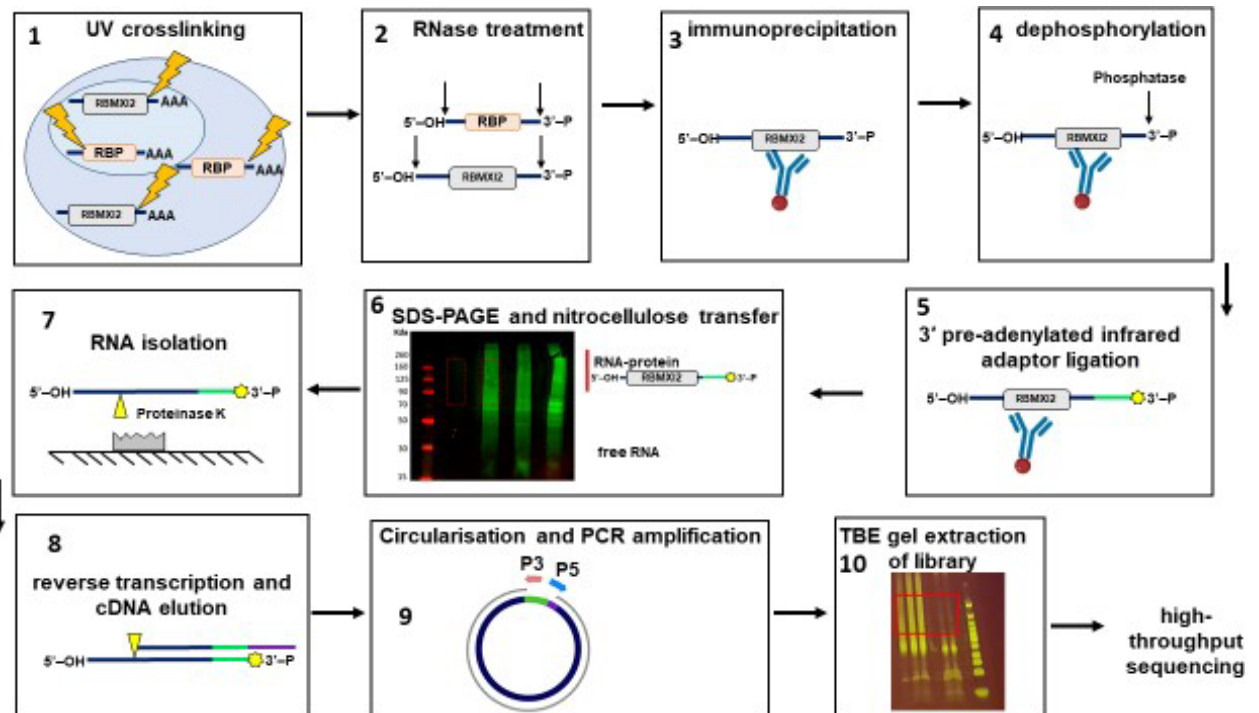


Figure 4. 1: Visual illustration of the iCLIP protocol (1) UV radiation was used to induce irreversible covalent RNA-protein crosslinks in mouse testes. (2) Testis extracts were digested with RNase after sonication. (3) RNA was immunoprecipitated with the anti-RBMXL2 antibody to form an RNA-protein complex. (4-5) RNA was dephosphorylated, and a pre-adenylated L3-IR adaptor 3' RNA linker was bound to the RNA complex. (6) The protein-RNA complexes were separated by size using the SDS-PAGE and nitrocellulose membrane. The area of the membrane with the protein-RNA complex was cut out with the help of a printed template of the area. (7) The protein-RNA region was fragmented with a scalpel and digested with proteinase K. (8) After purifying the RNA, a primer containing a barcode sequence was used to reverse-transcribe the RNA into cDNA. (9) AMPure XP beads were used to purify the cDNA, and then (10) Agarose gel was used to separate DNA in the appropriate size range. (150-400 nt) that was purified and sent for sequencing using the Illumina 500 machine.

4.2 Aims:

- 1-Identifying RNA targets of RBMXL2 in mouse testis using individual-nucleotide UV crosslinking and immunoprecipitation (iCLIP).
- 2-Investigate RBMXL2 mechanisms of splicing regulation by combining iCLIP results with RNA-seq data obtained previously in our lab (Ehrmann et al., 2019).
- 3-Investigate whether RBMX is involved with RBMXL2 in splicing and regulation of target exons using a minigene assay.

4.3 Results:

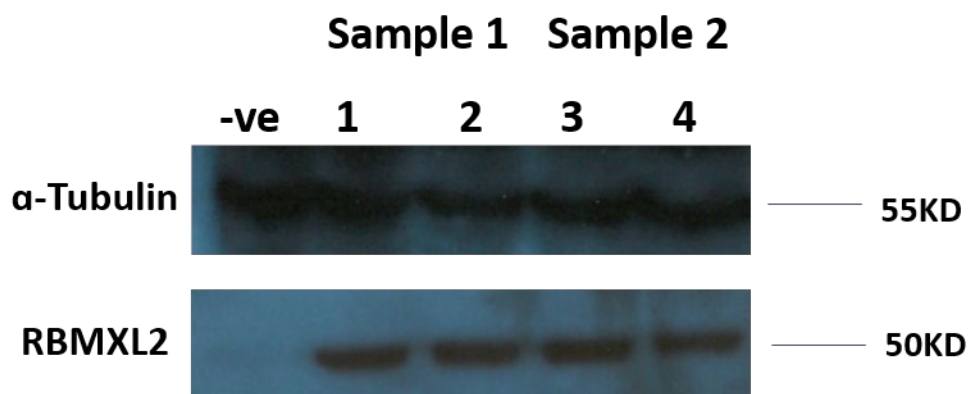


Figure 4. 2: Western blot to show the specificity of the RBMXL2 antibody. Lane 1 contains protein extracted from (MDA-MB-231 cells) as a negative control. Lanes 2 -5 contains protein extracted from mice testis (duplicate samples). A loading control (α-tubulin) was used to prove protein availability on the blot. However, the membrane is over-exposed which makes it difficult to distinguish the tubulin signals in some lanes.

4.3.1 Optimisation of the RBMXL2 iCLIP experiment

Optimising the iCLIP experiment was performed with Dr. Chile Siachisumo's assistance figure 4.1. I first examined how specific the RBMXL2 antibody was that we had made in our lab several years ago (Ehrmann et al., 2008, 2019). I did this by measuring the expression of RBMXL2 protein in a wild-type testis and negative control (breast cancer cell line, MDA-MB-231, that does not express RBMXL2) using the Western blot technique. I probed this western blot for RBMXL2 followed by a loading control (α -tubulin, sigma Cat# T6199). This demonstrated the expression of the testis specific RBMXL2 in all mouse testis samples but not in the breast cancer cell line which does not express RBMXL2, supporting the specificity of the antibody as seen in the figure 4.2.

I next set out to optimise the other steps in the iCLIP procedure. The iCLIP protocol was provided by the Ule Lab (unpublished). They substituted a non-radioactive IR dye for the radioactive dye in their previously published protocol (Huppertz et al., 2014). Also, they tweaked the elution, circularisation conditions, and beads-to-sample ratio during purification, but the core steps remained the same.

The steps used in the iCLIP procedure are shown in Figure 4.1. Mrs Caroline Dalglish provided wild-type mouse testes, which were diced into small pieces with cold-PBS and a protease inhibitor, and serial pipetting was performed prior to UV cross-linking. Mouse testis was UV cross linked 4 times at 254 nm with mixing in between. Following Chile's advice, I kept the Bioruptor[®] sonication cycles at 10 cycles. Then, the immunoprecipitation of RNA-protein complexes, observed by gel electrophoresis and detection of Infra-red labelled complexes, was next optimised. The iCLIP protocol needs to be optimised at various points to ensure accurate immunoprecipitation of RBMXL2. The RBMXL2 antibody was confirmed as specific, as tested before using Western blot and confirmed by IgG antibody immunoprecipitation samples that showed no signal. In contrast, only RBMXL2 precipitated samples showed an IR signal, as shown in Figure 4.3. The RNase amount used was also built in Chile's optimisation step, with RNase amounts of 0.8U/ml for high RNase I and 0.25 U/ml for low RNase. A high concentration of RNase I showed better RNA-protein complexes as seen below figure 4.3 in comparison to the low RNase I.

The last stage of the optimisation was the number of PCR cycles used for the purpose of amplifying the iCLIP cDNA library. Fewer cycles are generally preferred; however, sufficient cycles are required to yield PCR products for cloning. I tried 15, 19 and 30 cycles. This is shown in Figure 4.4. 19 cycles provided the highest level of specific amplification (bands were seen around 150 to 400 bp) with minimal to no non-specific amplification. 15 cycles showed non-specific bands around 10 to 150 bp, while 30 cycles showed smear amplification. Then, libraries for four biological replicates were produced. The samples were then sent to the Tapestation (bioanalyser) for analysis, and no primer dimer was detected as seen in the figure 4.5 below. Finally, the core genomics facility at Newcastle University sequenced the samples.

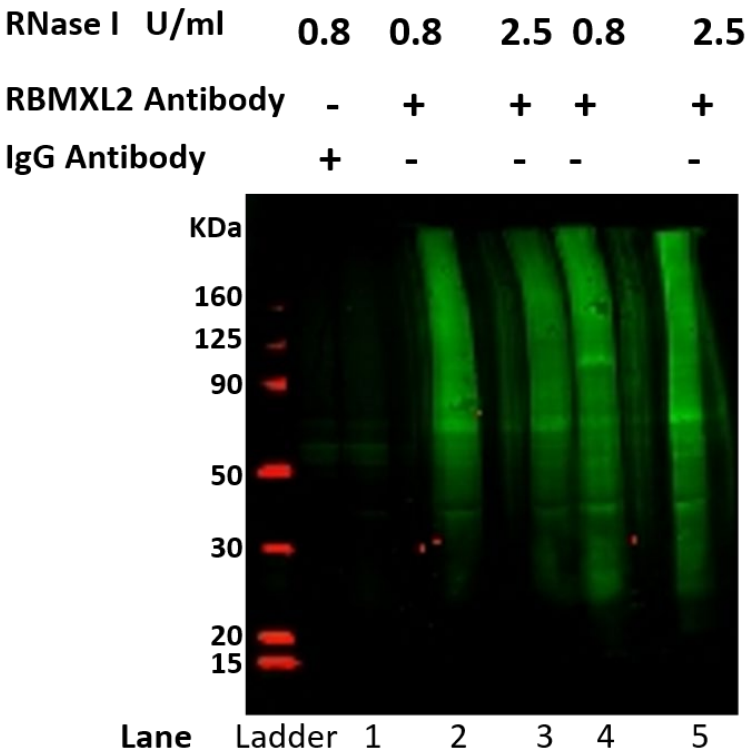


Figure 4. 3: iCLIP early optimisation of immunoprecipitations step. Samples on the blot membrane after SDS-PAGE are a ladder shown in Red, followed with immunoprecipitates made with IgG antibody (lane 1 as a negative control) and 2 replicates of high RBMXL2 antibody with low RNase (lanes 2 and 4), and 2 samples was high antibody and high RNase (lanes 3 and 5). The RNA-protein complexes appear as a smear above the 60 kDa RBMXL2 protein.

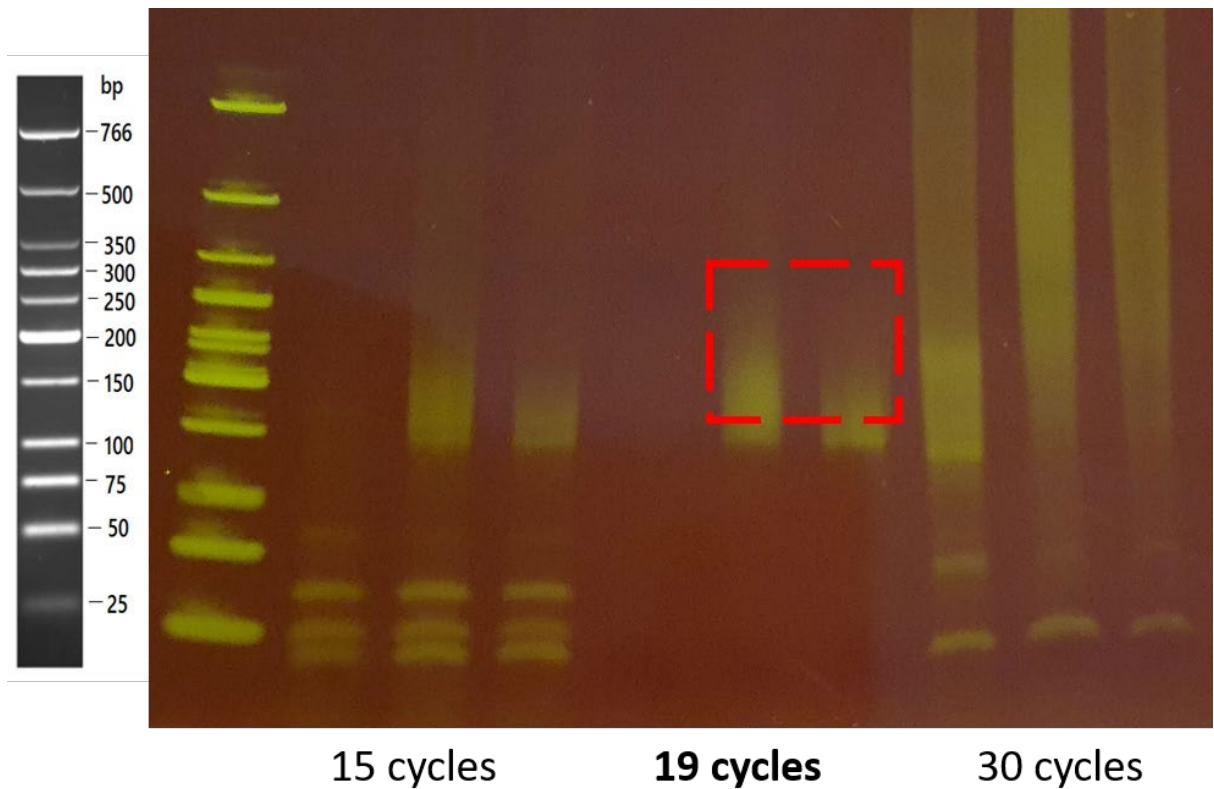


Figure 4. 4: TBE agarose gel showing experiments to optimise production of cDNA libraries at 15, 19, and 30 cycles with a DNA size marker (ladder). The red box designates the region excised (150–400 bp) following the iCLIP protocol for gel purification.

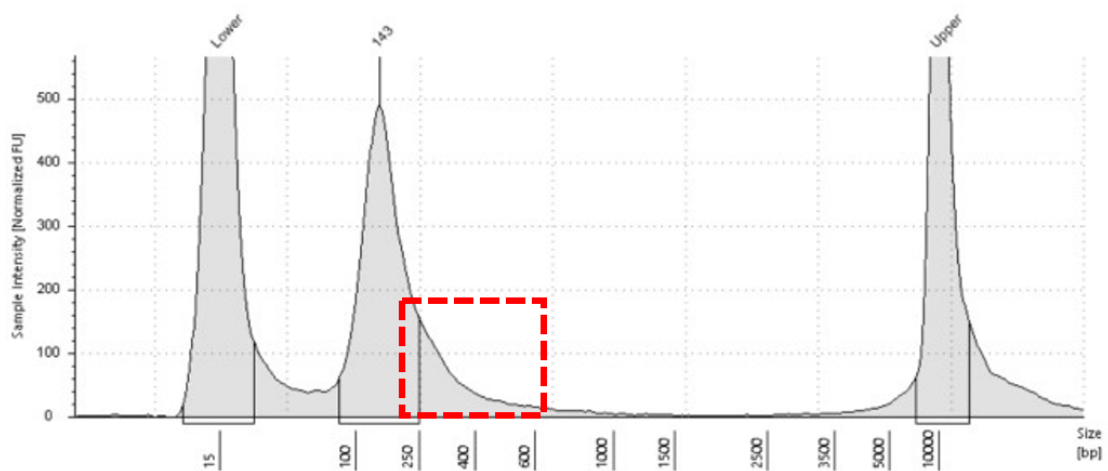


Figure 4. 5: Tapestation result for the cDNA library of one of iCLIP samples. The amplification product was around 143 bp (region designated in the dotted red box).

4.3.2 RBMXL2 iCLIP analysis of the genome-wide RNA targets

Following sequencing, raw data files were received, and Dr Graham Smith helped process the Fastq files for the four samples to be analysed with the iMaps pipeline (<https://imaps.goodwright.com>). The primary analysis pipeline of iMaps was used to analyse and demultiplex Fastq files. iMaps evaluates the reads' quality, eliminates duplicates, and demultiplexes the reads following the Unique Molecular Identifier (UMI). The iCLIP libraries were aligned to the mouse genome reference (GRCm39).

RBMXL2 was found to be bound to 30,945 genes, of which 14,048 were coding genes. Some of these coding genes were not yet well annotated, while other genes were non-coding, including Long Non-Coding RNAs (lncRNAs), Small Nucleolar RNAs (snoRNAs), and MicroRNAs (miRNAs). The bam files were used on the bioinformatics tool, Galaxy server (<https://usegalaxy.org>), to generate correlation analysis between the four biological replicate samples as seen in the figure 4.6 below. The correlation between samples were high (above 90%) as indicated in the blue boxes. The orange boxes show the no antibody sample.

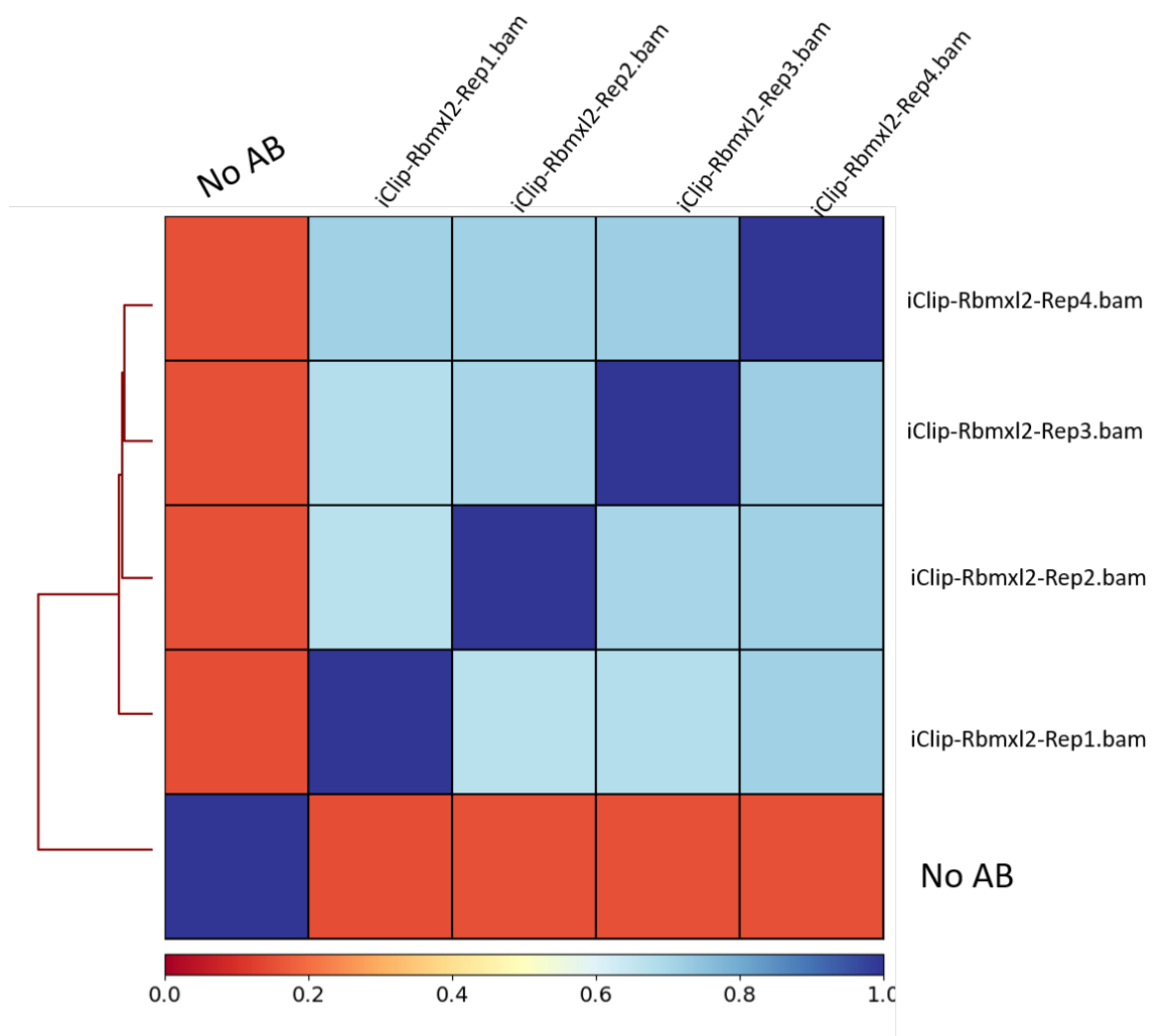


Figure 4. 6: A correlation analysis between the 4 biological replicates showing the positive correlation. Correlation matrix made with Dr. Chile Siachisumo's assistance on (<https://usegalaxy.org/>).

4.3.3 Location of RBMXL2 binding within the genome.

The downstream analysis of the iCLIP experiment was carried out using a merged file of all 4 replicates, after it was confirmed that the iCLIP libraries created from the 4 biological replicates had a positive correlation across the 4 samples with a slight variation. To establish where RBMXL2 protein binds to RNA at a global level, I looked at where its binding sites are in the transcript regions. The resulting figure 4.7 showed that most RBMXL2 binding sites were in the intronic regions, however, around 10 % are in the coding regions of the genome.

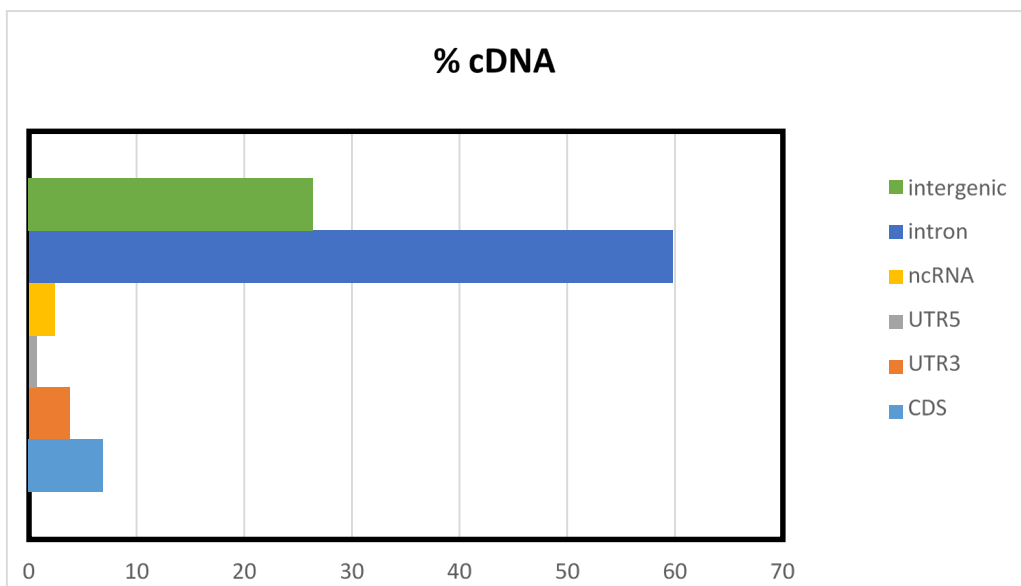


Figure 4. 7: Bar chart showing where RBMXL2 protein binds to different regions of the genome. RBMXL2 binds strongly in intergenic and intronic regions.

The genome is divided into exons (frequently protein-coding regions) and introns (non- coding regions between exons), while non-coding regions between genes are referred to as intergenic. The three types of exons are coding sequences (CDS), 3' untranslated regions (UTR), and 5' untranslated regions (UTR). As seen in figures 4.7 above and 4.8 below, the regions of transcripts that RBMXL2 protein prefers to bind to were identified using the iMAPS pipeline. Intron and intergenic regions contained the majority of RBMXL2 binding sites. This may be due to the fact that a significant proportion of the genome is made up of introns and intergenic regions, so these are quite a big target for crosslinking. However, RBMXL2's strong affinity for introns suggests a potential function in controlling pre-mRNA splicing. Around 10% of the RBMXL2-crosslink sites are within exonic regions (CDS and 3 UTR).

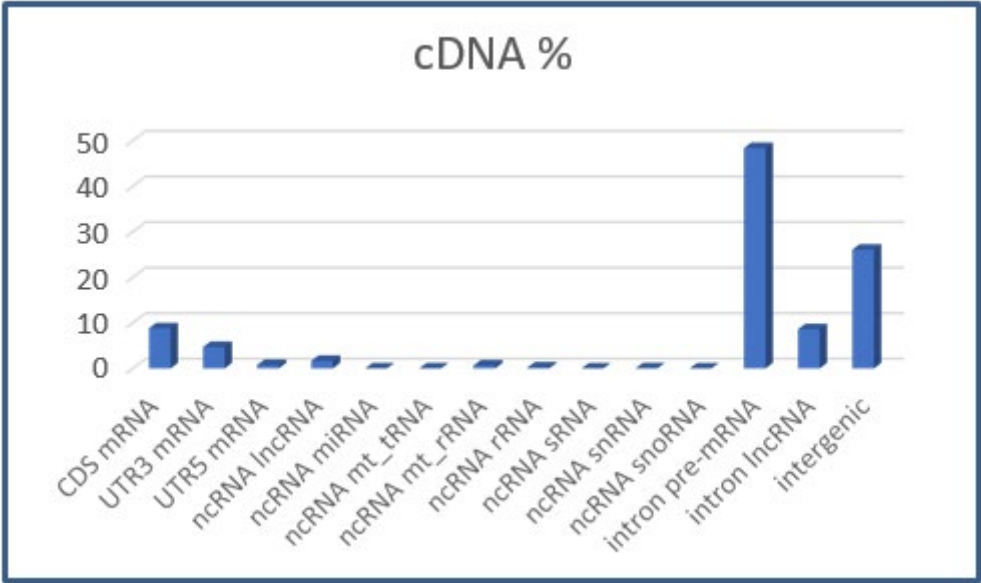


Figure 4. 8: Bar chart showing the subtypes of gene that are bound by RBMXL2 protein.

Gene ontology (GO) was used to determine the biological processes/pathways associated with genes that bind RBMXL2. The first top 500 genes that bind to RBMXL2 (when ordered by the number of tags) were employed from 2 biological replicates. Figure 4.9 shows that the majority of the 20 enriched pathways were related to spermatogenesis, sexual reproduction, and reproductive processes, with pathways involved in cilium (sperm tail) assembly and movement.

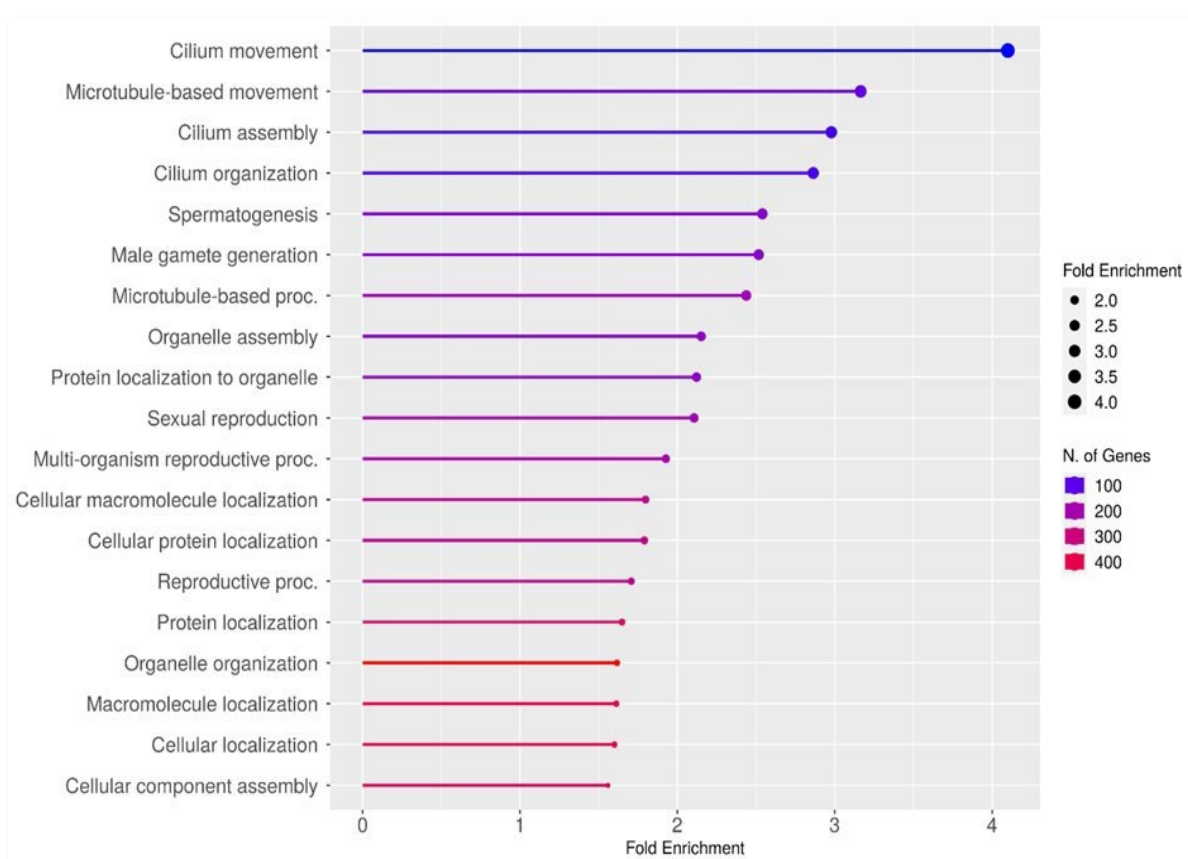


Figure 4. 9: A chart representing gene ontology enrichment analysis of RBMXL2 bound genes (Top 500 genes) identified by iCLIP. The chart was produced in ShinyGO 0.77 (<http://bioinformatics.sdstate.edu/go/>).

4.3.4 Analysis of known regulated targets of RBMXL2 protein for iCLIP RNA binding site

iCLIP provides single nucleotide resolution of RNA-protein binding sites that can be visualised on genome browsers. On the IGV (Integrative Genome Visualisation) browser, tags (where RBMXL2 binds to RNA) were clearly visible within interesting target genes. I also had access to RNAseq data from wild type and *Rbmxl2* knockout mice. In RNA-seq, peak calling involves identifying the number of tags bound to each gene. Therefore, I have chosen the *Meioc* gene to demonstrate the peak calling, as shown in Figure 4.10 below.

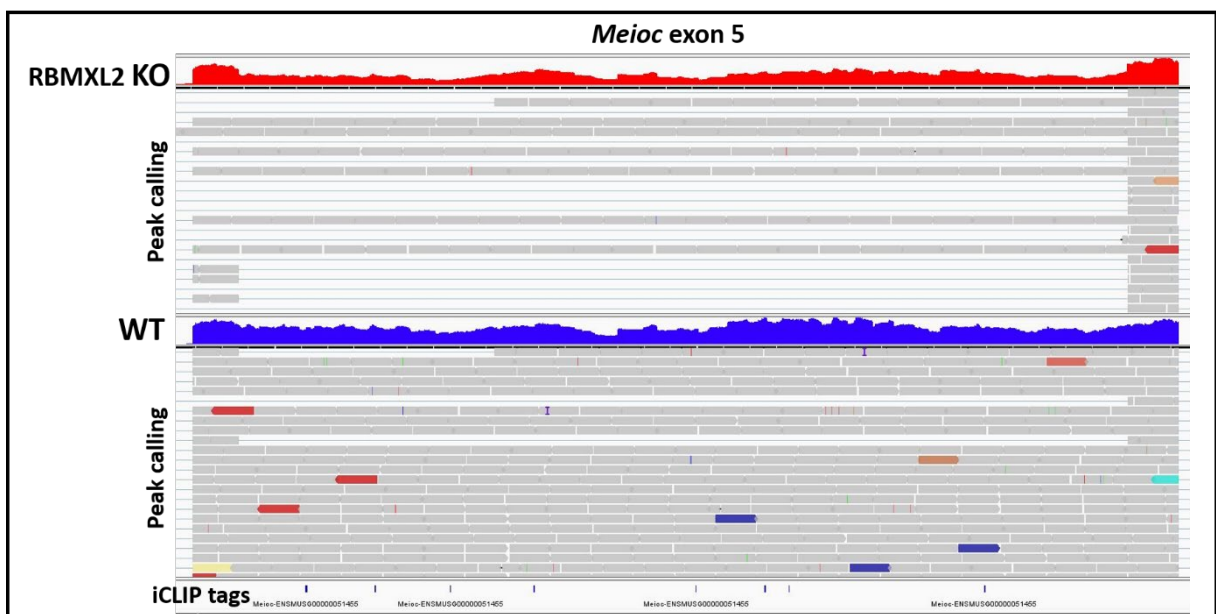


Figure 4. 10: A screenshot from IGV shows the effect on processing of *Meioc* exon 5 after knockout of the *Rbmxl2* gene versus the wild type. This Figure displays the number of iCLIP tags mapping to exon 5 of the *Meioc* gene. The knockout (KO) sample exhibits fewer peaks compared to the wild type (WT) sample. Picture taken as snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

Esco1, *Meioc*, *Kdm4d*, *Alms1* and *Brca2* were identified as important genes based on previous HITS-CLIP and RNA-Seq data (Ehrmann et al., 2019) so I particularly analysed these genes. Starting with *Esco1*, the 5' cryptic splice site was seen within the ultra-long exon 3, as shown in Figure 4.11 and 4.12. Tags were observed around 60 bp from the utilised cryptic splice site in ultra-long exon 3. Additionally, multiple tags can be found throughout the exon as seen in figure 4.12.

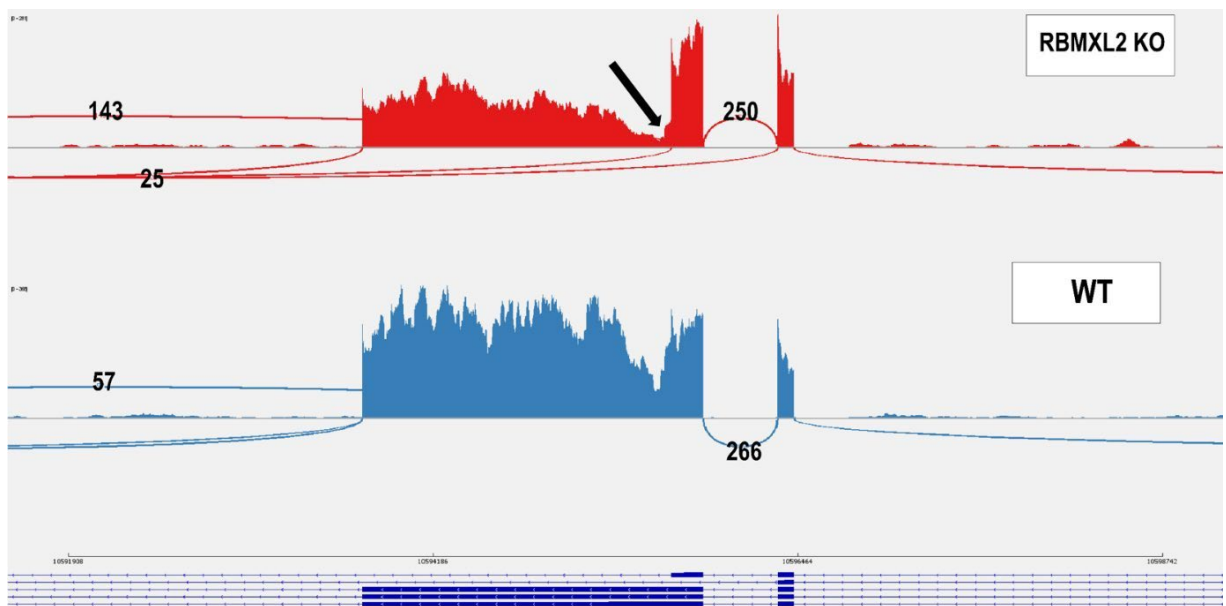


Figure 4. 7: Cryptic splicing of *Esco1*. Screenshots from Sashimi plot showed the utilisation of the 5' cryptic splice site in exon 3 of *Esco1* of the *Rbmxl2* knockout that is not seen in the wild type testis. The arrow shows position of the cryptic splice site. This is RNA-seq data from 18 days mouse consist of *Rbmxl2* knockout and Wild type testis. Numbers indicates the splicing junction reads. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

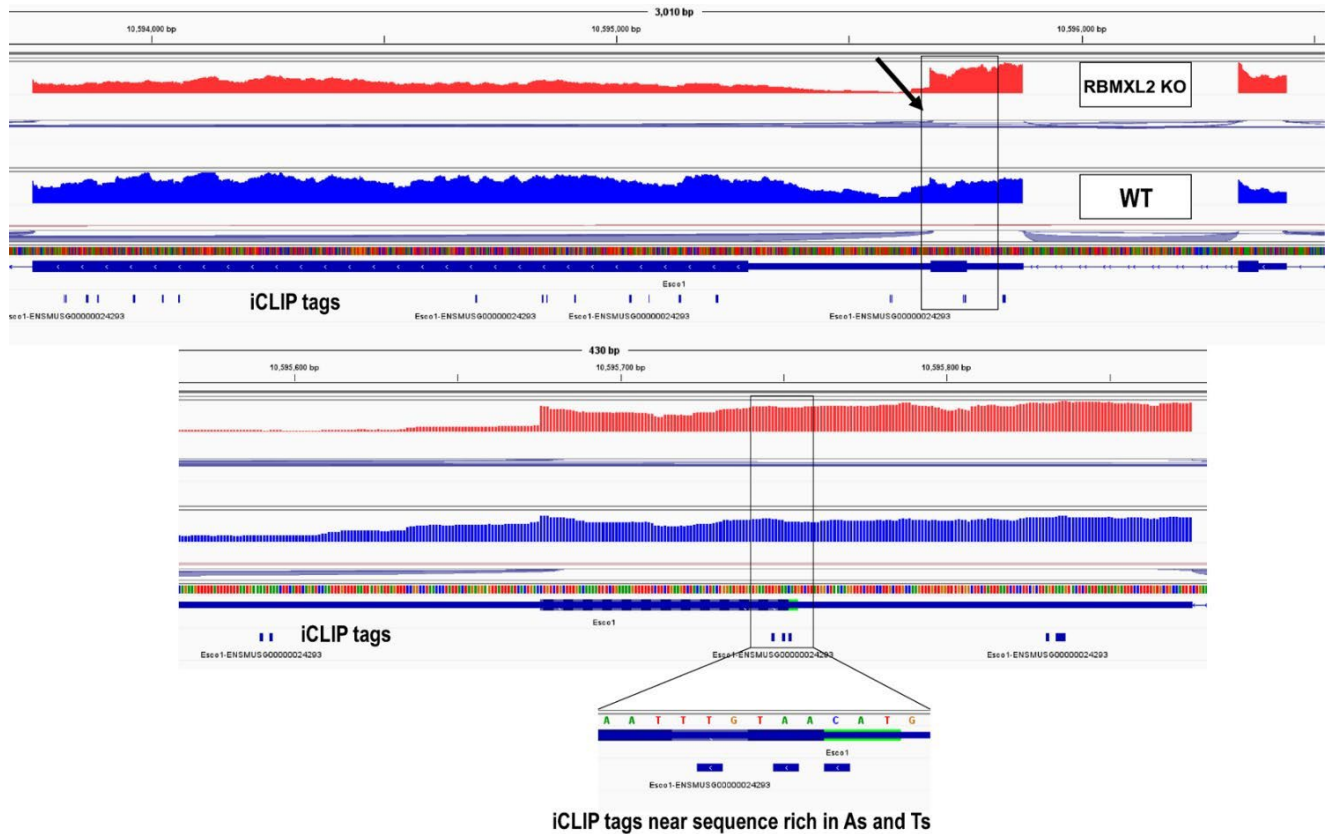


Figure 4. 8: RBMXL2 binding to *Esco1*. Screenshots from IGV showed several iCLIP tags bound to exon 3 of *Esco1*. iCLIP tags were rich with As and Ts were detected close to the cryptic splice site used in the *Rbmxl2* knockout mouse. The arrow showed the position of the cryptic 5' splice site within exon 3. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

Another example of RBMXL2 regulation of cryptic splicing is *Meioc*, which is known to be expressed during meiosis and was seen to form an exon in exon 5 if *Rbmxl2* was deleted. Figure 4.13 of the sashimi blot demonstrates the utilisation of the weak cryptic 3' and 5' splice sites within *Meioc* gene that were not seen in the wild type testis. In addition, iCLIP tags were detected close to the cryptic splice sites in the 3' (within 150 bp) and 5' (within 236 bp) splice sites, as shown in Figure 4.14. Tags were identified in regions rich in As and Ts sequences.

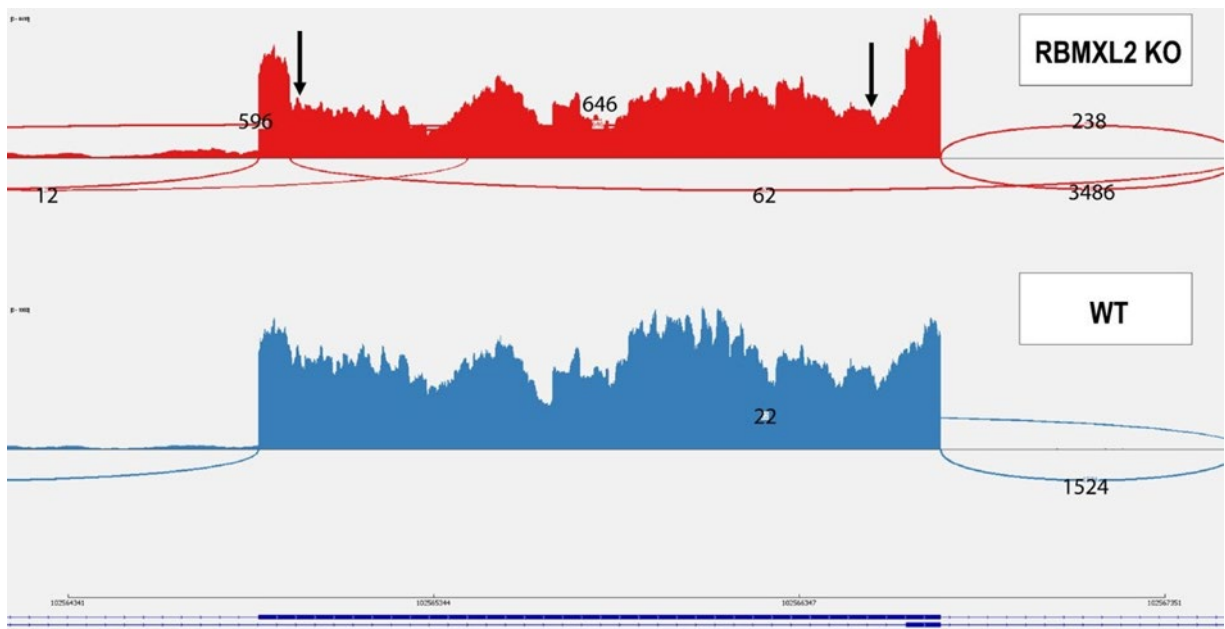


Figure 4. 9: An exon within the *Meioc* gene. Screenshots from Sashimi blot showed the formation of an exon within exon 5 of the *Rbmxl2* knockout that is not seen in the wild type testis. RNA-seq data from 18 days mouse (*Rbmxl2* knockout and wild type). Arrows show the utilisation of the weak 3' and 5' splice sites. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

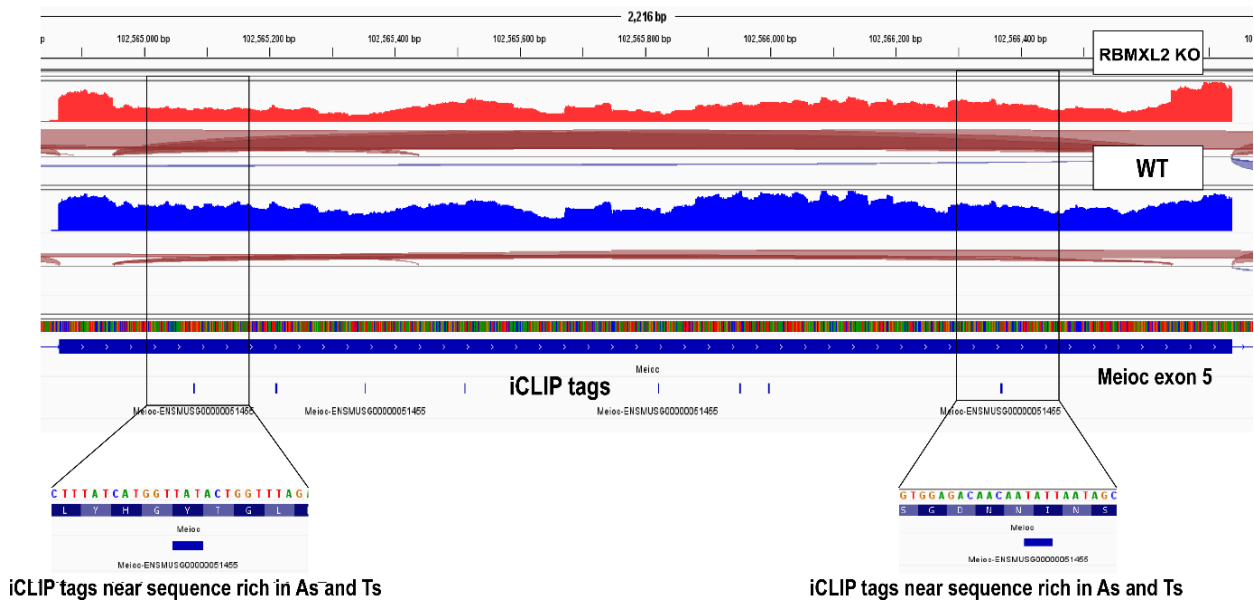


Figure 4. 10: iCLIP mapping of RBMXL2 RNA protein interactions within Meioc exon 5.

Screenshots from IGV showed several iCLIP tags bound to exon 5 of *Meioc*. iCLIP tags were rich with As and Ts were detected close to the 5' and 3' cryptic splice sites used in the *Rbmxl2* knockout mouse. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

The third gene I checked was *Kdm4d*, which has a cryptic exon in the *Rbmxl2* knockout mouse testis that is not present in the wild type. The figure below 4.15 illustrates the inclusion of the cryptic exon after knockout of *Rbmxl2*. This cryptic exon was repressed in the wild type testis when RBMXL2 protein was present. Also, inclusion of an already annotated alternative spliced exon was observed. In addition, multiple tags are present throughout the exon, as depicted in Figure 4.16. Moreover, an iCLIP tag was detected near the cryptic splice site (less than 20 bp) and bound to an A-rich sequence region.

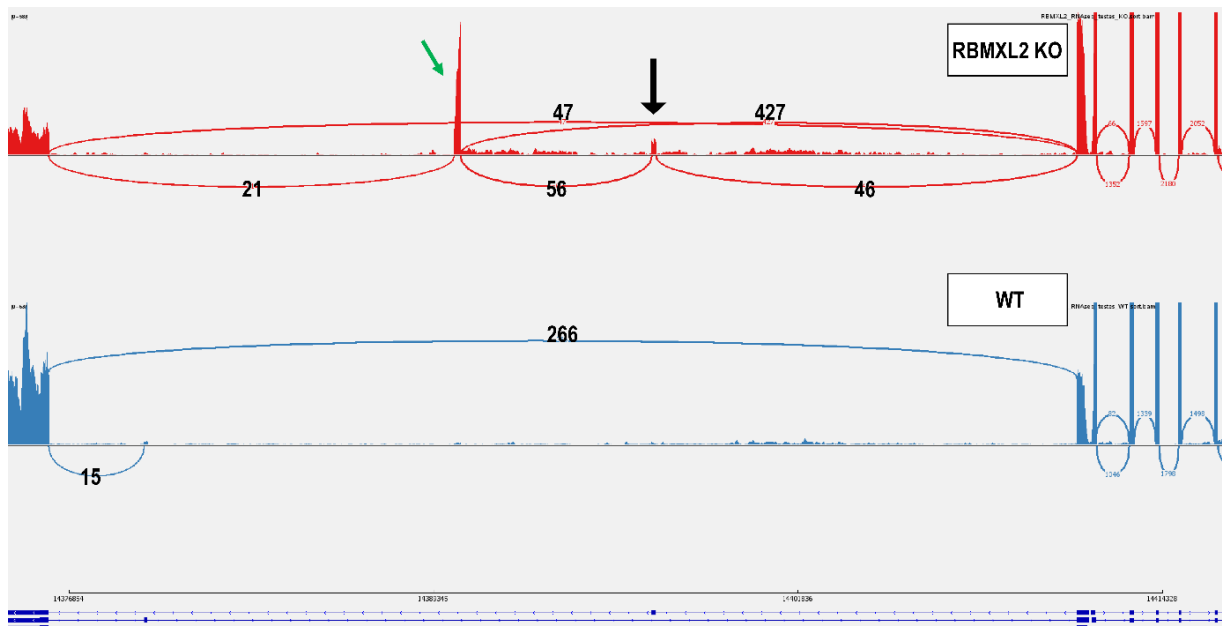


Figure 4. 11: Cryptic splicing within *Kdm4d*. Screenshots from Sashimi blot showed the inclusion of cryptic exon in *Rbmxl2* knockout that is not seen in the wild type testis. RNA-seq data from 18 days mouse (*Rbmxl2* knockout and wild type). Black arrow showed the inclusion of the known alternative exon while the green arrow showed the inclusion of the cryptic exon. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

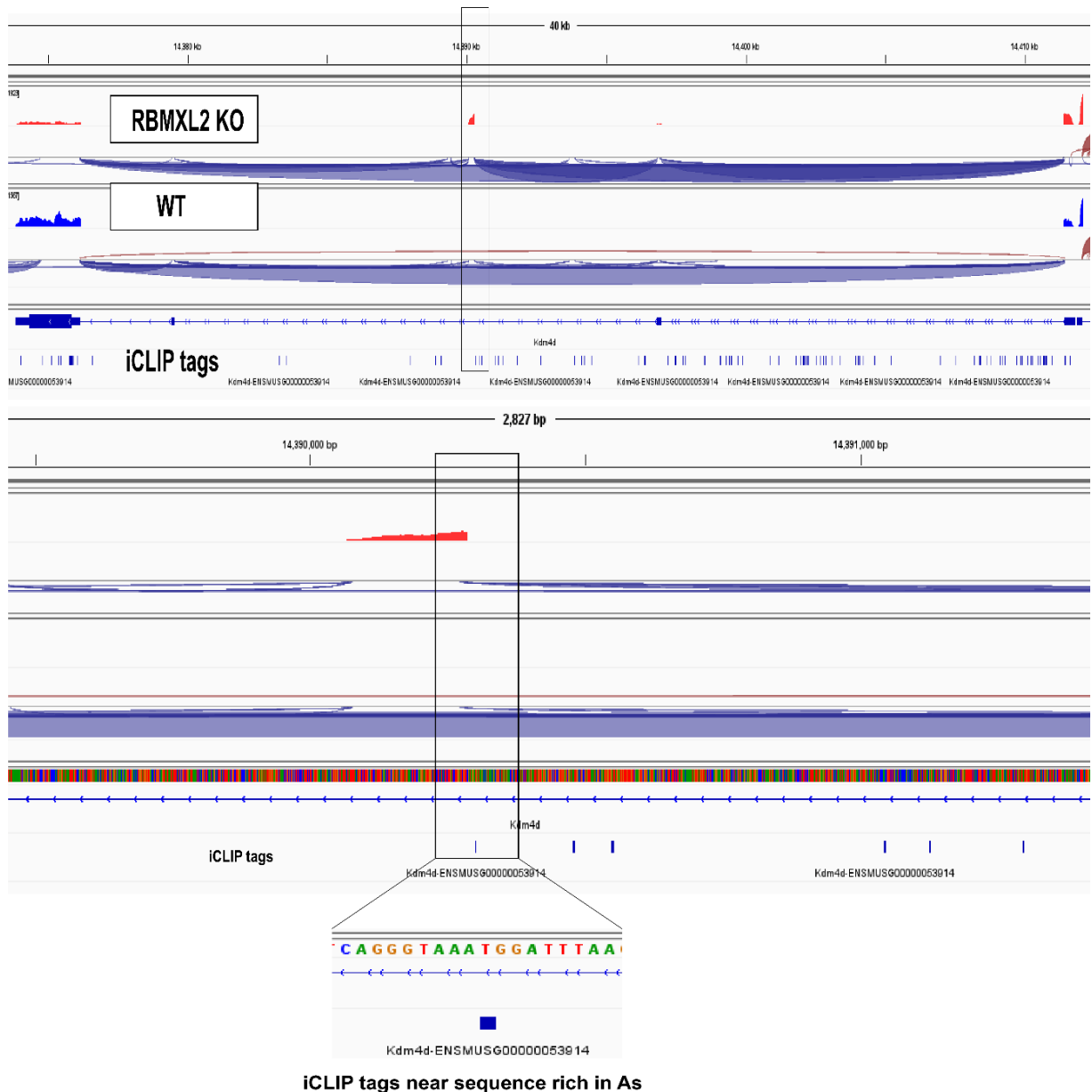


Figure 4. 12: Mapping of iCLIP tags within the *Kdm4d* gene. Screenshots from IGV showed several iCLIP tags bound to Kdm4d. iCLIP tags were rich with As and Ts were detected close to the cryptic exon in the *Rbmxl2* knockout mouse. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

Alms1 is another gene implicated in cell division and cilia formation (Arsov et al., 2006). The *Rbmxl2* knockout led to the utilisation of a cryptic 5' splice site within ultra-long exon 5. Figure 4.17 shows that the cryptic 5' splice site was repressed in the wild type when RBMXL2 was expressed. Additionally, several iCLIP tags were detected in exon 5. Figure 4.18 shows that an iCLIP tag was detected close (around 70 bp) to the 5' cryptic splice site.

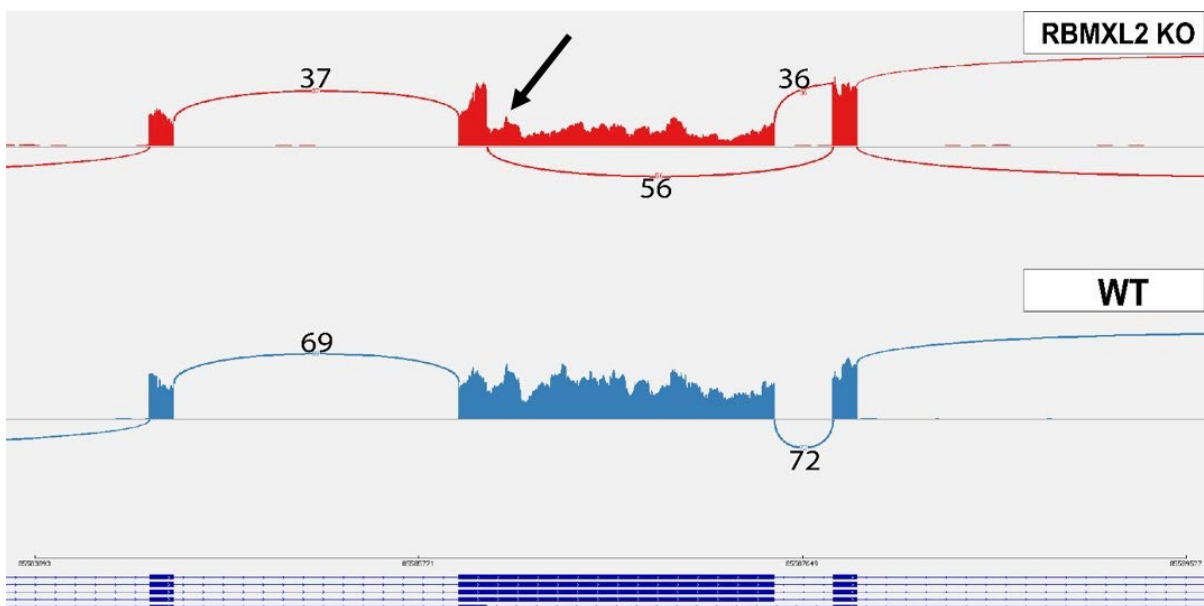
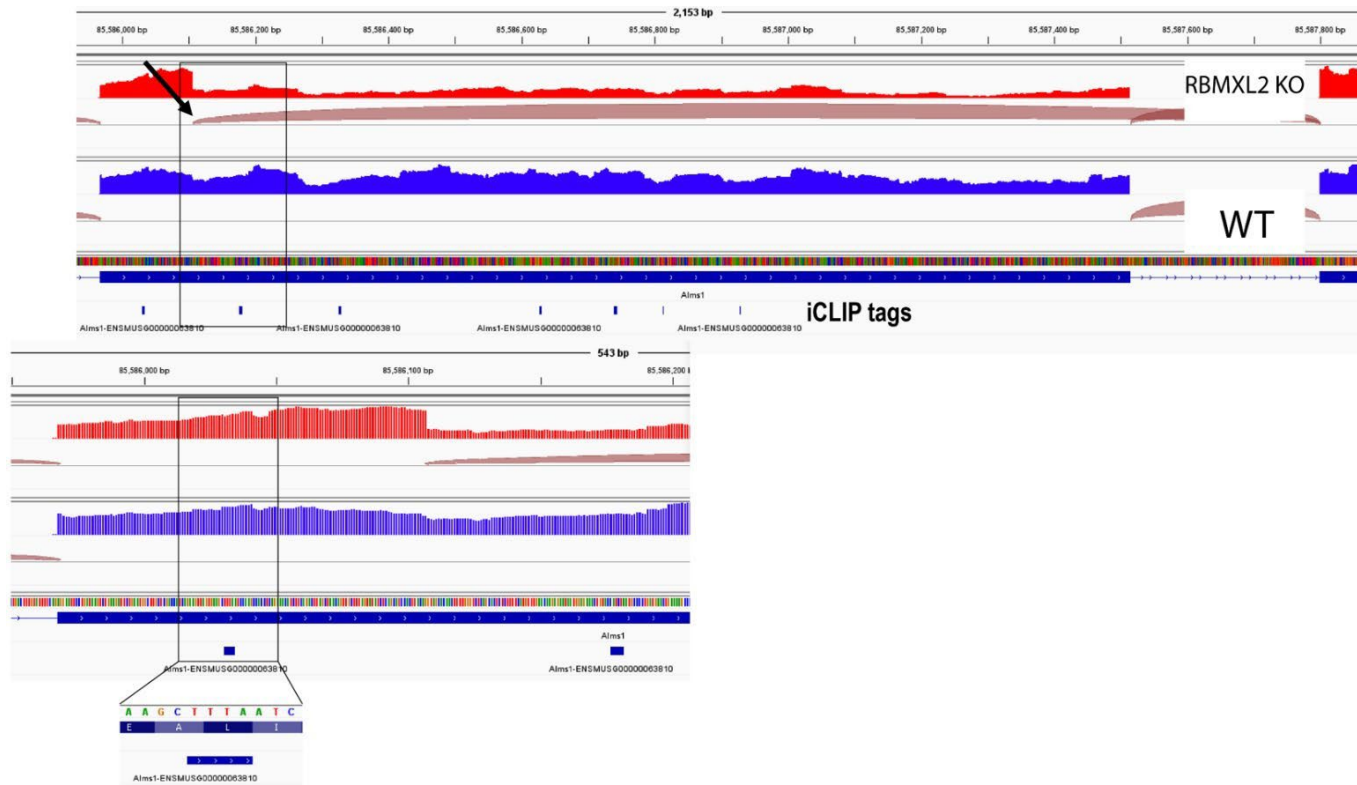


Figure 4. 13: Cryptic splicing in *Alms1*. Screenshots from Sashimi blot showed the utilisation of the 5' cryptic splice site within exon 5 in *Rbmxl2* knockout that is repressed in the wild type testis. Arrows showed the 5' cryptic SS. RNA-seq data from 18 days mouse (*Rbmxl2* knockout and wild type). Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).



iCLIP tag near a sequence rich in Ts and As

Figure 4. 14: Mapping of iCLIP tags near to the cryptic splice site used in *Alms1*.

Screenshots from IGV showed several iCLIP tags bound to *Alms1*. iCLIP tag was rich with As and Ts and was seen close to the 5' cryptic splice site of *Rbmxl2* knockout mouse testis. The arrow showed the position of the cryptic 5' splice site within exon 5. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

Based on RNA-seq data, the *Rbmxl2* knockout demonstrated usage of the 5' cryptic splice site within exon 11 of the *Brca2* gene, as shown in Figures 4.19 (Ehrmann et al., 2019). Figure 4.20 reveals that no iCLIP tags were detected in the *Brca2* gene, which suggests that either (1) RBMXL2 protein may regulate this cryptic splice site independent of RNA binding, or (2) the iCLIP mapping might not be totally comprehensive.

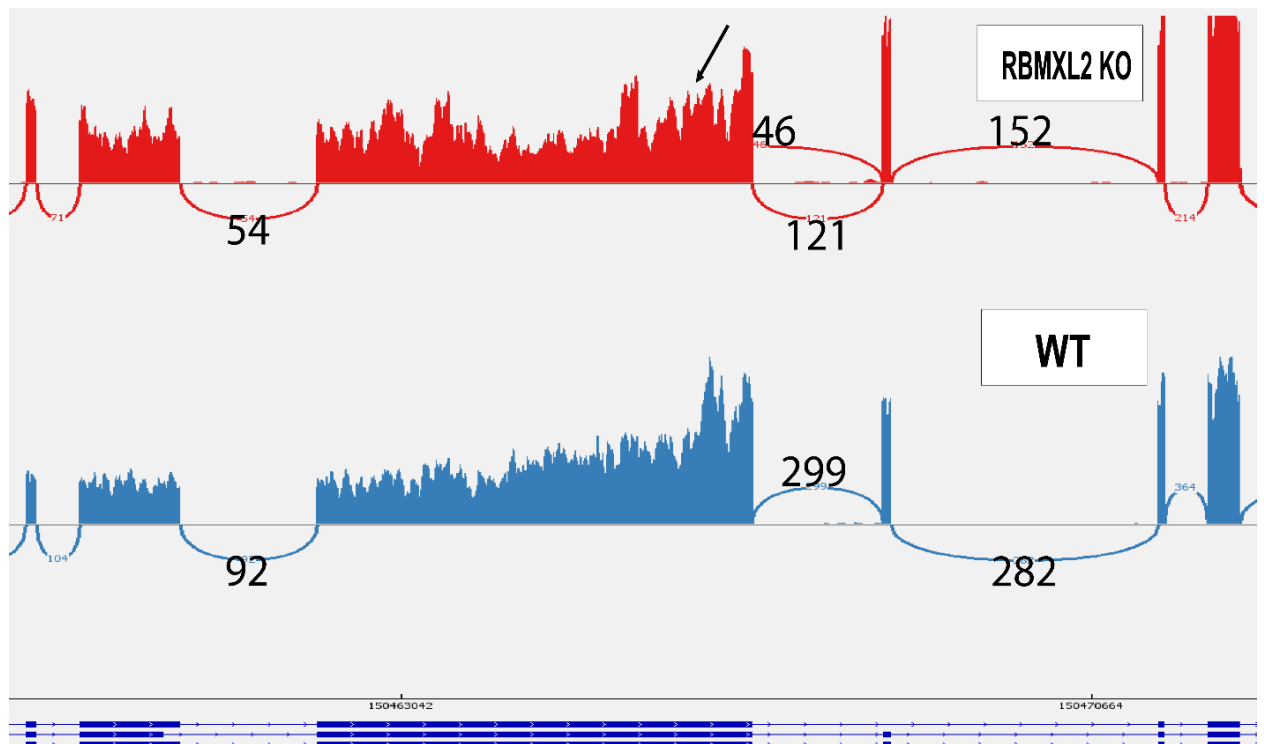


Figure 4. 15: Cryptic splicing in *Brca2*. Screenshots from Sashimi blot showed the utilisation of the 5' cryptic splice site within *Brca2* gene exon 11 in *Rbmxl2* knockout that is repressed in the wild type testis. The arrow shows the position of the cryptic 5' splice site. RNA-seq data from 18 days mouse (*Rbmxl2* knockout and wild type). Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

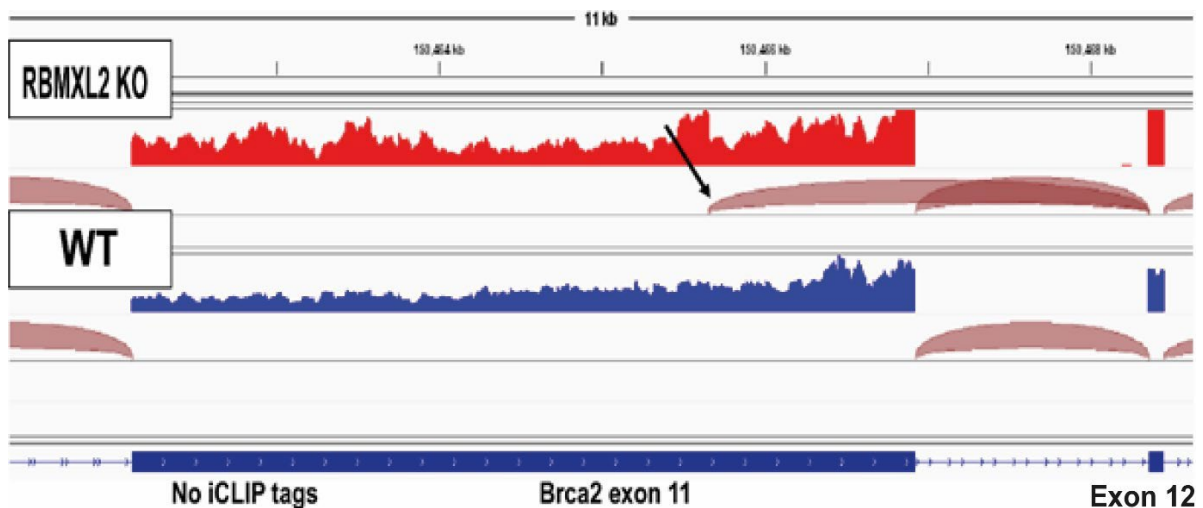


Figure 4. 16: Cryptic splicing of *Brca2*. Screenshots from IGV showed no iCLIP tags mapped to the *Brca2* gene exon 11. The arrow shows the cryptic 5' splice site. RNA-seq data from 18 days mouse (*Rbmxl2* knockout and wild type testis). Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

4.3.5 Identification of RBMXL2 binding site sequences (this analysis was done with collaborators mentioned below).

RNA-binding proteins typically bind to motif sequences of 6 to 8 nucleotide long. I collaborated with Dr. Ivaylo Yonchev (Sheffield University) to use my iCLIP data to identify binding sites preferentially bound by RBMXL2. Ivaylo conducted enriched motif analysis (kmer analysis) and metagene analysis of my iCLIP data. First, analysis of the top 50 motif sequences from Figure 4.21 showed that RBMXL2 binding was enriched in TTT-rich sequences.

This was different than the HITS-CLIP results where we identified RBMXL2 binding was enriched in AAA sequences (Ehrmann et al., 2019).

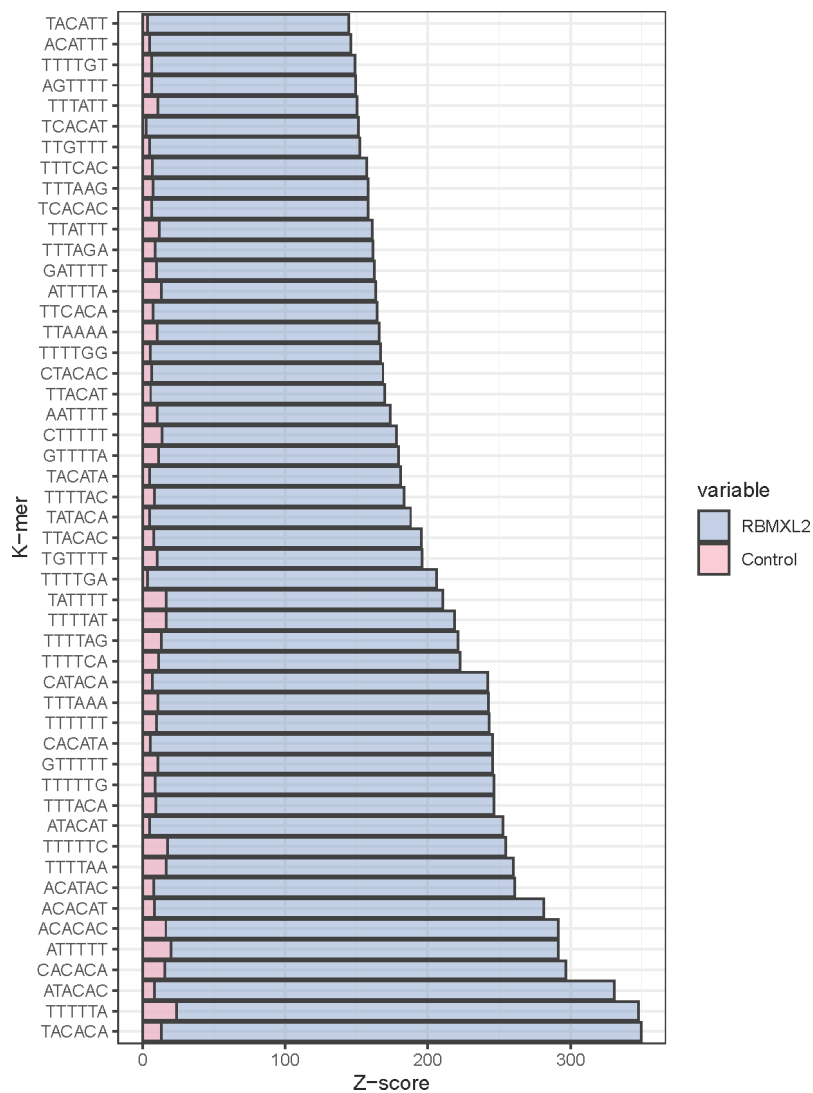


Figure 4. 17: Top 50 K-mer enrichment determined from analysis of my iCLIP data. Pink is the control (iCLIP done with IgG antibody) while grey is RBMXL2 protein binding to the motifs. The most enriched motif was TACATT.

While analysing RBMXL2 RNA processing, we detected that RBMXL2 regulates splicing events within ultra-long exons such as *Meioc*, *Esco1*, and *Kdm4d* among others. These particular genes are longer than the average size of mouse exons, which is around 120 bp. Dr. Sara Luzzi performed additional analysis to indicate that RBMXL2-regulated and bound exons were significantly longer than all mouse genome exons. The analysis is included in the figure 4.22 below.

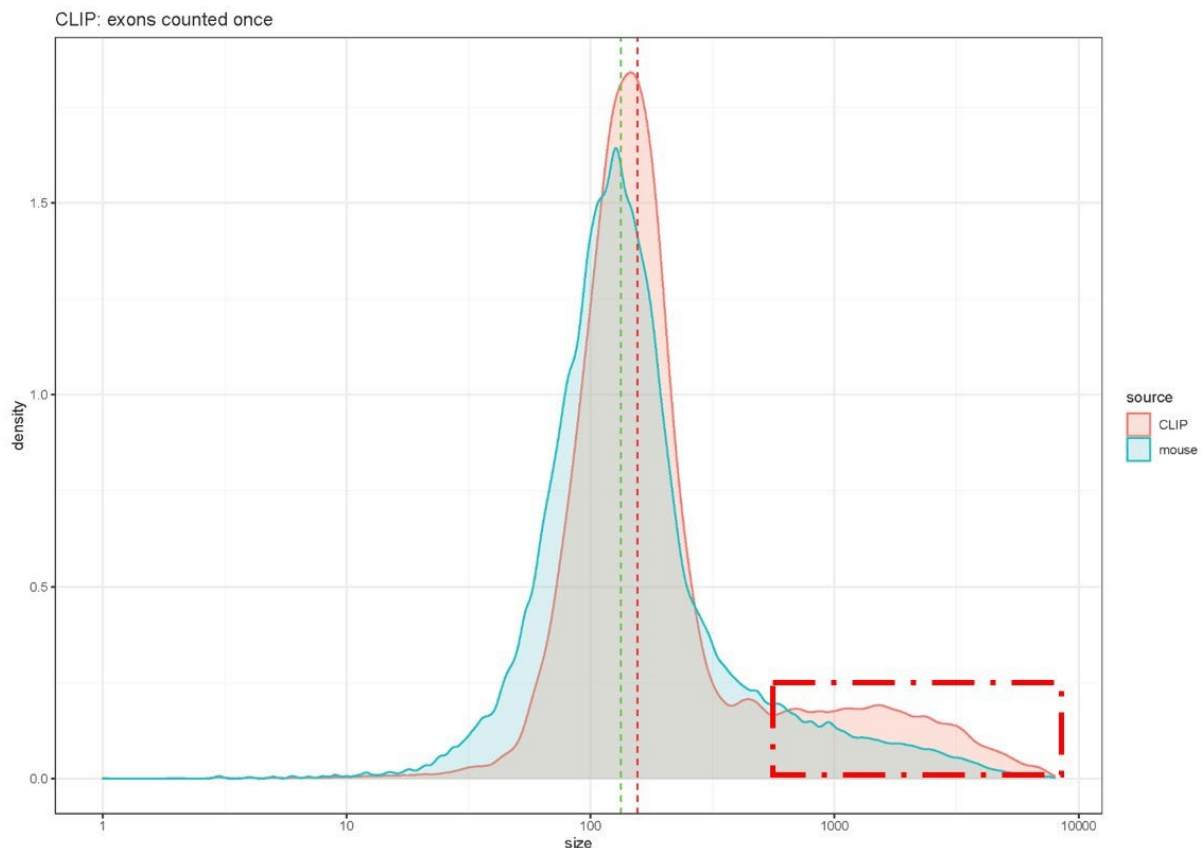


Figure 4. 18: RBMXL2 binds and regulates RNA processing events in ultra-long exons. Light blue colour demonstrating the distribution plots of all mouse exons; Red colour representing all RBMXL2 bounds genes identified by iCLIP (counted once). The dotted green line indicates the median mouse exon size and the dotted red line represents the iCLIP median exons size bound genes. Red dotted box shows the longer exons bound by RBMXL2. This analysis was done by Dr. Sara Luzzi, Newcastle University.

4.3.6 RBMXL2 Binds the 3' end of the transcript

Dr Ivaylo Yonchev (Sheffield University) also identified RBMXL2 binding sites across a typical gene in a metagene analysis profile 4.22. Metagene analysis is a tool for analysing gene expression across multiple samples by aligning the genes against a common reference point, such as the transcription start site, to generate an average expression profile from which we can identify gene expression patterns. Ivaylo generated a gene list using Ensembl annotations for the mouse genome (<http://mart.ensembl.org/info/genome/genebuild/index.html>). The RBMXL2 binding signal is highest toward the end of the whole transcript. Most sites where RBMXL2 binds are at the 3' end of the transcript, while the signal at the 5' end is weaker as seen in the figure 4.23.

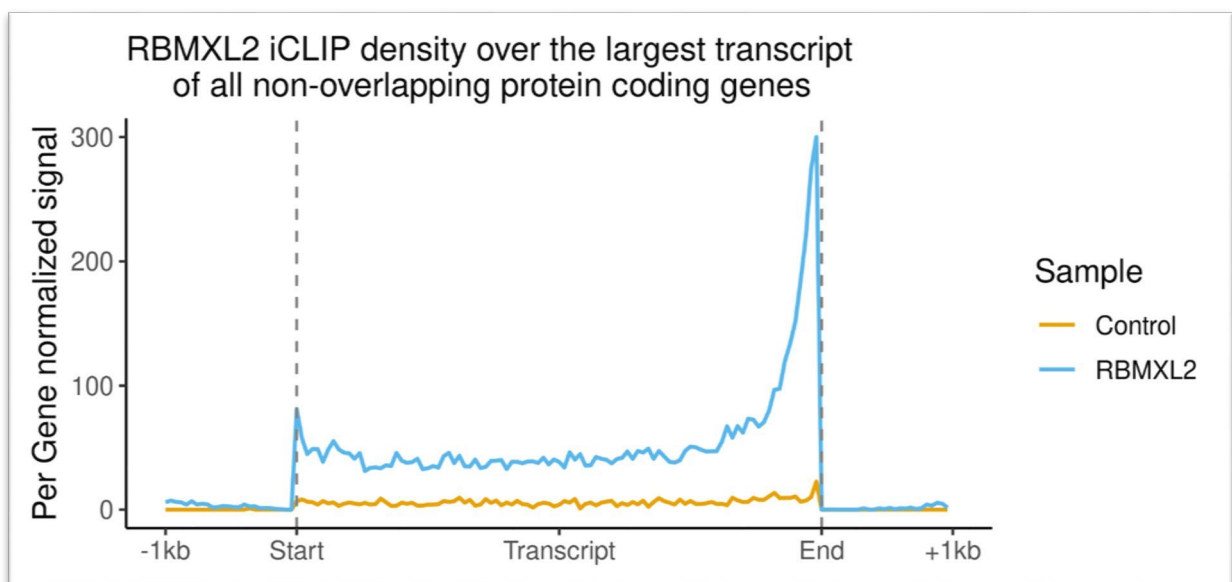


Figure 4. 19: Metagene analysis of RBMXL2 iCLIP tracks from mouse testes shown full transcript regions normalized per gene. Highest RBMXL2 binding density were seen towards the 3' end of the gene.

4.3.7 Investigation of RBMXL2, RBMX mechanism to regulate mouse *Esco1* gene using a Minigene assay

Having identified RBMXL2 binding sites within total testis RNA using iCLIP, I next wanted to use this information to test the importance of specific binding sites on RNA processing control. Splicing regulation in vivo is frequently studied using the minigene system (Mardon et al., 1987). The minigene system allows for mutational changes to binding sites within plasmids. In addition, several studies (Liu et al., 2009; Ehrmann et al., 2019) have examined the regulation of splicing by RBMXL2 protein using minigenes.

Minigene construction involves cloning the target region from the genomic DNA into an exon trap vector. Our exon trap vector (pXJ41) comprises two beta globin exons, downstream of a promoter. Minigene studies are suitable for investigating cis- and trans- acting factors that affect a particular area of a gene or to introduce new sequence (Anna & Monika, 2018). I attempted to characterize the splicing of RBMXL2-controlled target exons using minigenes. From our previous RNA-seq data and HITS-CLIP data (Ehrmann et al., 2019), the *Esco1* gene was identified as target for RBMXL2 RNA processing control. Exon 3 of the *Esco1* gene utilises a cryptic splice site if *Rbmxl2* is depleted. This switch can be observed on the IGV genome browser, comparing RNA-seq signals from wild type versus *Rbmxl2* knockout mouse testis figure 4.24. ESCO1 protein is essential for chromatid cohesion, especially on DNA repair, chromosome loop formation, controlling transcription, and stabilisation (Alomer et al., 2017).

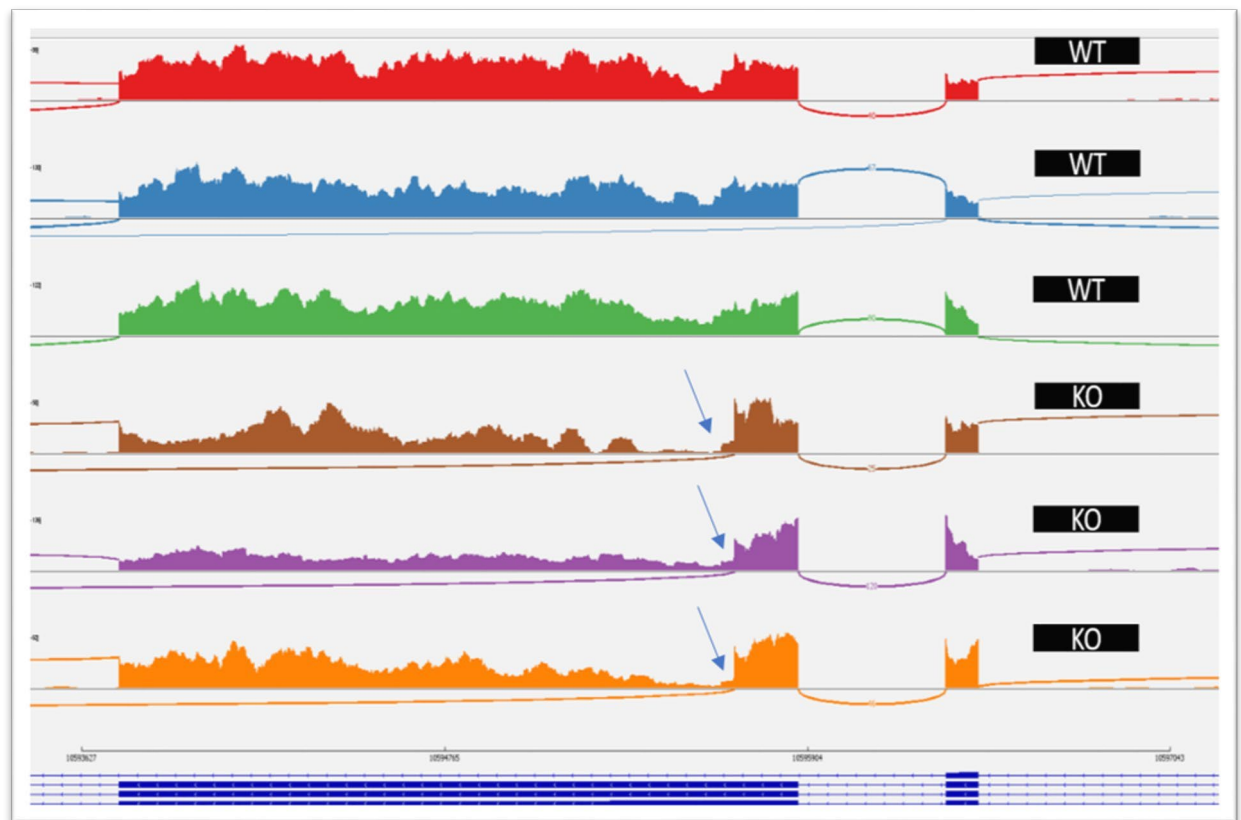


Figure 4. 20: A screenshot of a Sashimi plot shows RNA-seq data from day 18 mouse testis, three wild type samples then three *Rbm12* knockout samples. In the *Rbm12* knockout samples there is a clear utilisation of the cryptic splice site in exon 3 of the *Esco1* gene (annotated by arrows) which is not seen in the wild type testis. The Sashimi plot is a snapshot from the IGV genome browser, Integrative Genomics Viewer (Robinson et al., 2011).

After cloning and verification, the minigene was transfected into HEK-293 cells for splicing analysis. There were 3 possible products: (1) Full exon inclusion of exon 3 of *Esco1*, although this is too long to be amplified by RT-PCR (*Esco1* exon 3 is an ultra-long exon) so should not be detected by RT-PCR. (2) Exon 3 skipping, joining the beta-globin exons. (3) Splicing of the cryptic 5' splice site between the beta globin exons. Using RT-PCR I should only be able to detect the skipping of exon 3 or the use of the cryptic 5' splice site (the use of cryptic splice site of exon 3) as seen in the figure below 4.25.

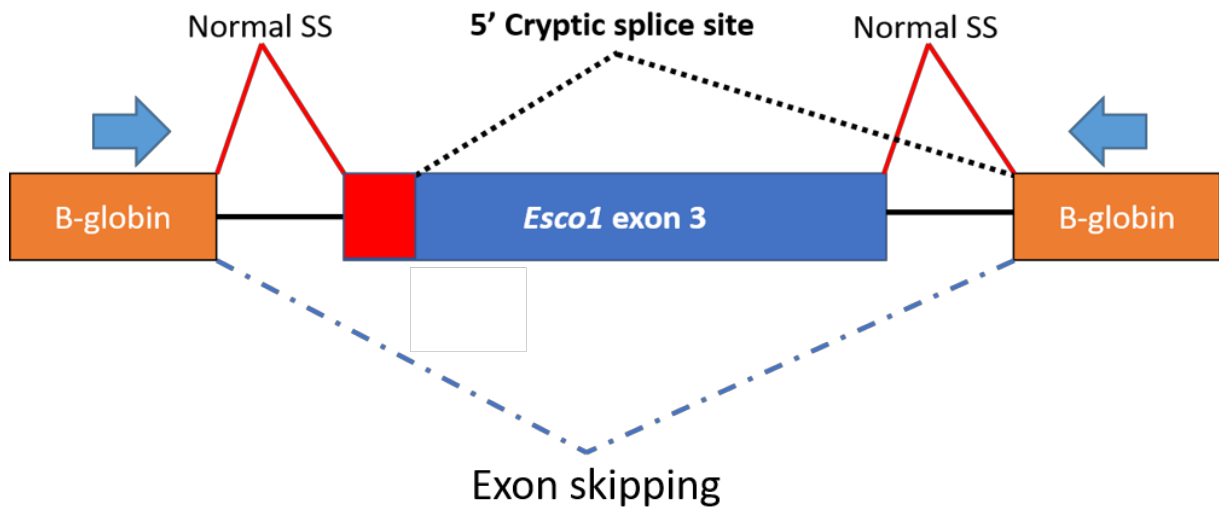


Figure 4. 21: A schematic of the *Esco1* exon 3 minigene design. Exon 3 with flanking introns was cloned between two B-globin exons within pXJ41 plasmid. The three products would be (1) Full exon inclusion of exon 3 of *Esco1*, although this is too long to be amplified by RT-PCR. (2) Exon 3 skipping, joining the beta-globin exons. (3) Or finally, the use of the cryptic 5' splice site. Boxes indicates exons while black lines represents introns.

To investigate splicing control, I then transfected cells using different expression vectors. These vectors expressed GFP (as an internal control), RBMXL2-GFP, RBMXL2 Δ RRM-GFP (RBMXL2 without the RNA recognition motif), and RBMX- GFP as seen in the figure 4.26. Transfections were done in triplicate to ensure that any changes in splicing could be replicated and were statistically significant. RNA was extracted from the transfected HEK-293 cells, analysed by RT-PCR using PXJF and PXJR primers as shown in the figures below, and analysed by Qiaxcel capillary gel electrophoresis figure 4.27.

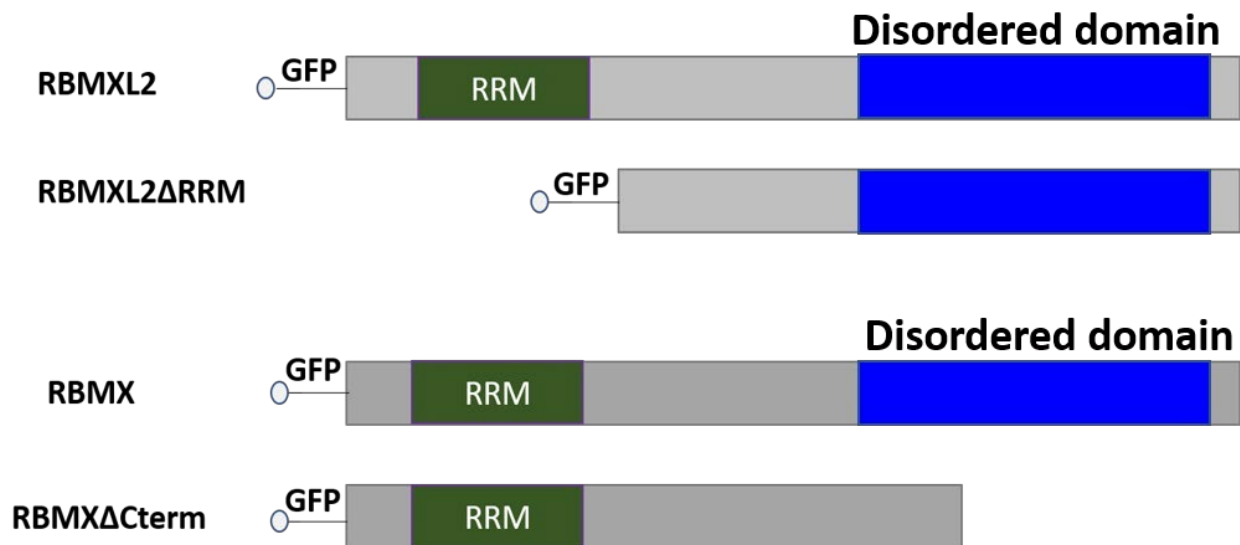


Figure 4. 22: A schematic figure showing domain structures of RBMXL2 and RBMX proteins and proteins construct missing domains of RBMXL2 and RBMX proteins tagged with GFP. Green boxes show the RNA-Recognition motifs while blue boxes represents the disordered domains.

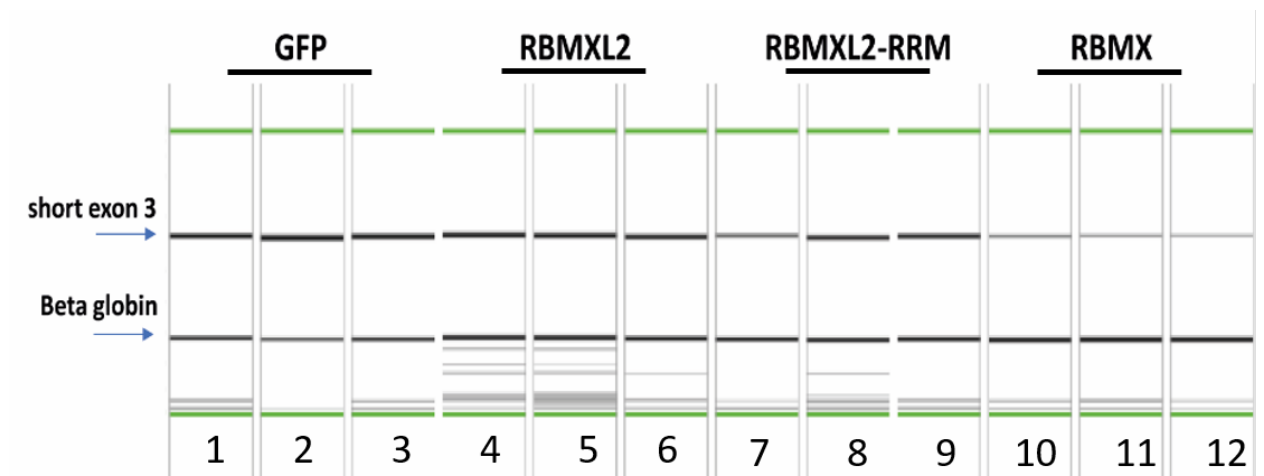


Figure 4. 23: QIAxcel electrophoretogram displaying splicing patterns of the *Esco1* minigene. RT-PCR for samples from different cell transfection were analysed in triplicate (GFP (lanes 1-3), RBMXL2-GFP (lanes 4-6), RBMXL2ΔRRM-GFP (lanes 7-9), RBMX (lanes 10-12) using PXJF and PXJR as RT-PCR primers, and then analysed by capillary gel electrophoresis.

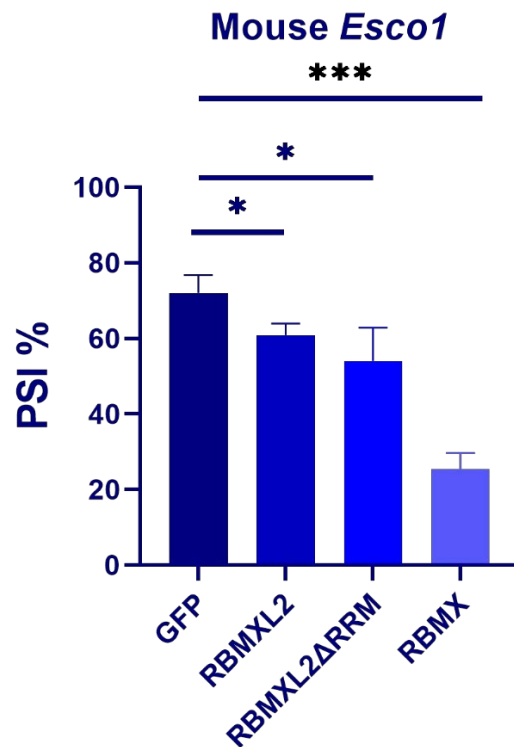


Figure 4. 24: RBMX weakly suppresses the use of the cryptic 3' splice site in *Esco1* exon 3 (n=3). Statistical analysis presented the PSI and P value associated for each condition on *Esco1* exon3 splicing. RBMX showed greater splicing repressor affect in comparison to RBMXL2 and RBMXL2ΔRRM. The PSI was calculated using data from three biological replicates. Lines in the bar plots represent the mean \pm SEM. P-values were calculated using an unpaired t-test on GraphPad Prism v 9.31 (*, **, ***, p-value 0.05).

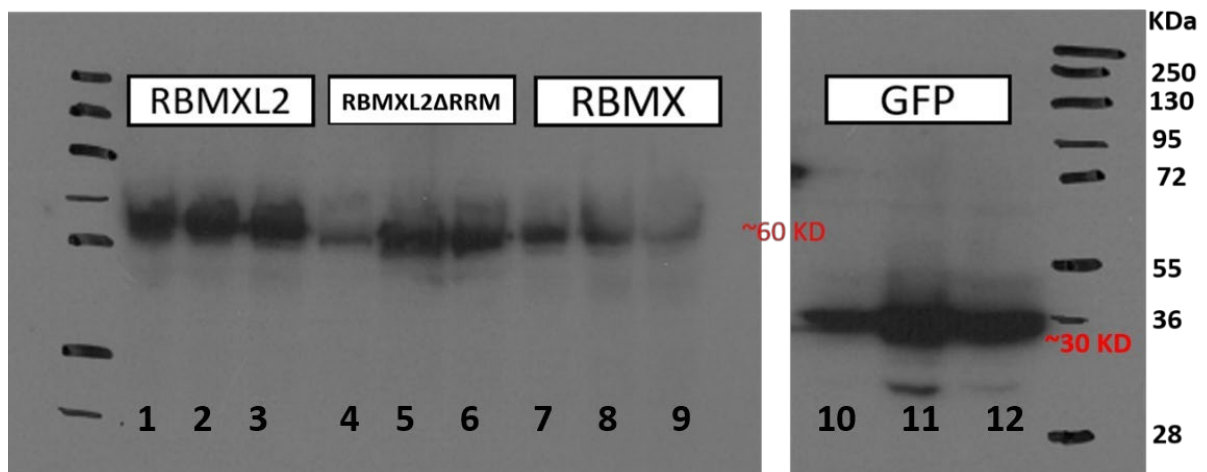


Figure 4. 25: Western blot shows plasmids were expressing similar levels of RNA binding proteins in triplicate transfections. Western blots were probed with antibodies specific for GFP. No loading control was performed for this experiment.

After RNA was extracted from transfected HEK293 cells, the PXJF and PXJR primers were used in a one-step PCR reaction to map the splicing patterns of the *Esco1* minigene.

In Figure 4.27, I analysed the RT-PCR products using capillary electrophoresis. Then I calculated the percentage splicing inclusion of the short form of *Esco1* exon 3 Figure 4.28 using the equation described in the 2nd chapter (materials and methods 2.6.10). Firstly, in the HEK293 cells transfected with the GFP expression plasmid (lanes 1-3), I observed two amplification products, the lower band representing the exclusion of the insert (exon 3) and the upper band indicating the use of the (cryptic) weak 5' splice site. In the cells transfected with RBMXL2 (lanes 4-6), I detected two products representing beta-globin splicing (the skipping of the (insert) and the splicing inclusion of the cryptic splice site of exon 3. This analysis showed that RBMXL2 weakly represses the cryptic 5' splice site. HEK293 cells transfected with RBMXL2 Δ RRM (lanes 7-9) also showed similar results to RBMXL2 (lanes 4-6). Both RBMXL2 and RBMXL2 Δ RRM transfections were statistically significant compared to GFP on the transfection shown in (lanes 1-3). HEK293 cells transfected with RBMX (lanes 10-12) displayed two amplification products: a strong band representing the skipping of the insert and a faint band indicating the utilisation of the cryptic splice site of exon 3. The result from transfection with RBMX was significantly different from GFP transfection (lanes 1-3), indicating that RBMX can more strongly suppress the selection of the cryptic splice site in *Esco1* exon 3 compared to RBMXL2.

Finally, I did a Western blot to confirm that the transfection had worked and that the amount of GFP-Fusion protein in each condition was equal, as indicated above in Figure 4.29. The amount of protein in RBMX-transfected cells was less than the amount in RBMXL2 and RBMXL2 Δ RRM-transfected cells. However, it shows the higher impact of RBMX on *Esco1* minigene splicing in comparison to the other conditions. In addition, it seems that the N-terminal RNA Recognition Motif in RBMXL2 did not affect the inclusion or the exclusion in exon 3 of *Esco1* gene.

These experiments showed that expression of RBMX prevents the use of the cryptic splice site of exon 3 of *Esco1* more strongly than expression of RBMXL2 and RBMXL2 Δ RRM. As mentioned earlier exon 3 of *Esco1* gene is ultra-long exon, meaning it cannot be easily visualised in the RT-PCR reaction. To further study the splicing activation patterns of the full exon in the *Esco1* minigene I designed a new set of primers that should be able to detect the longer version of the *Esco1* exon 3 figure 4.30. To get around this, a combination of PXJF and PXJR with an internal *Esco1* exon 3 primer was used as seen below figure 4.31.

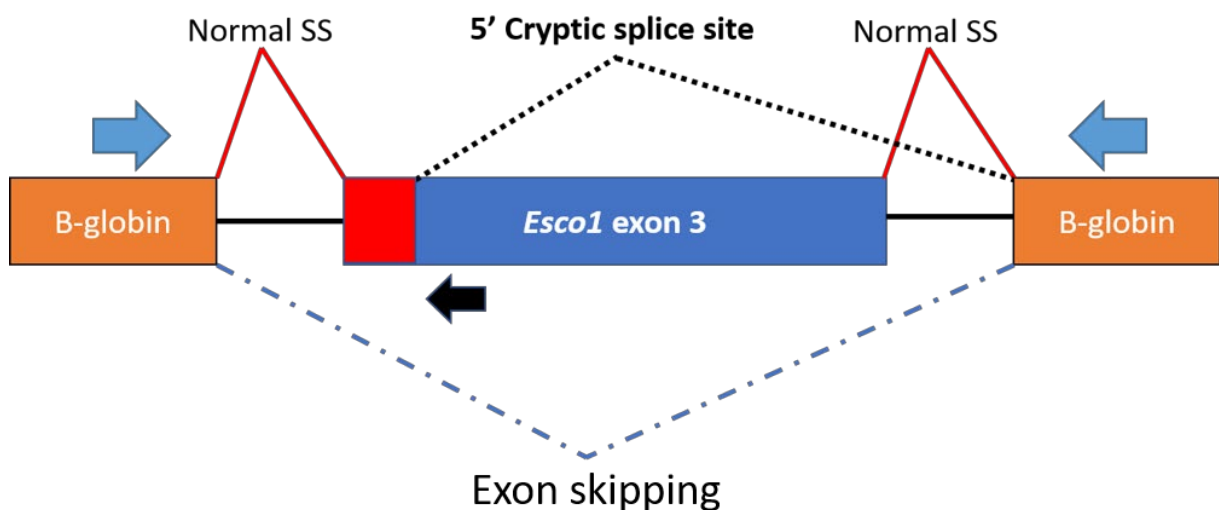


Figure 4. 26: Schematic figure of *Esco1* gene exon 3 with internal primer to detect the full length, inserted within pXJ41 plasmid. The orange blue boxes indicate the b-globin exons. The blue box in the middle is representing the 2.4 kb of exon 3. The black arrow represents the *Esco1* internal primer and blue arrows represent PXJF and PXJR.

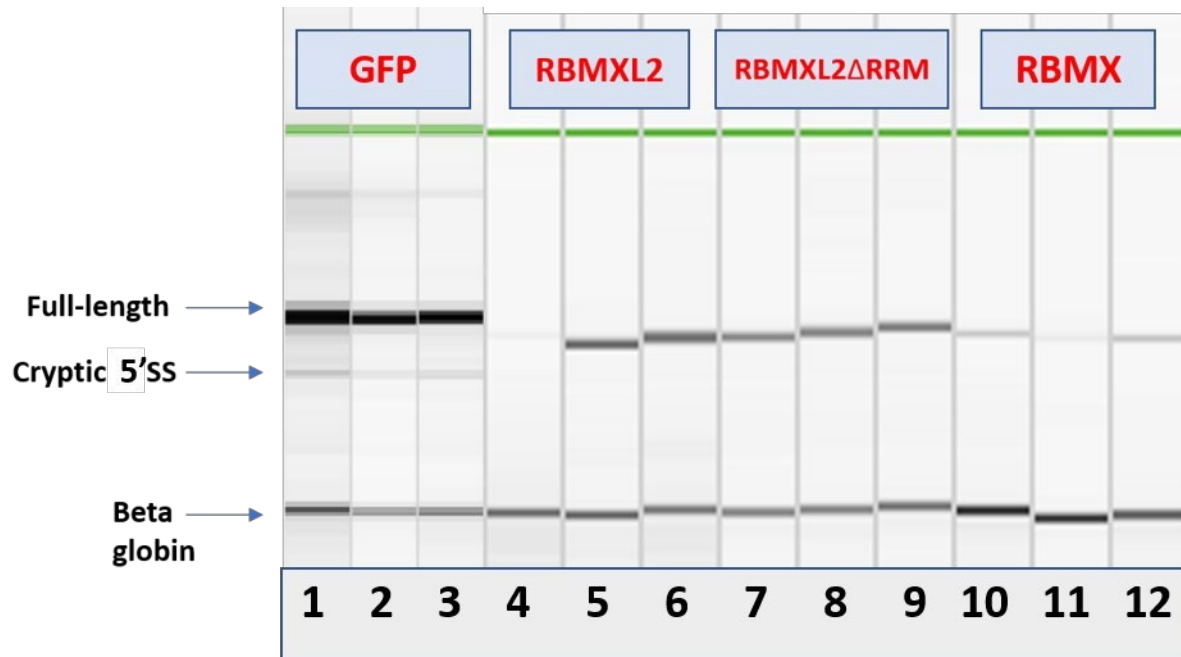


Figure 4. 27: QIAxcel electrophoretogram displaying the full-length splicing of exon 3 of the *Esco1* gene (n=3). RT-PCR for RNA extraction samples from different cell transfection carried out in triplicates: GFP (lanes 1-3), RBMXL2-GFP (lanes 4-6), RBMXL2 Δ RRM-GFP (lanes 7-9), and RBMX- GFP (lanes 10-12). RT-PCR was carried out using PXJF and PXJR primers with internal *Esco1* exon3 primer. However, the bands are not consistent in all lanes. GFP shows full-length expression along with the cryptic 5'ss, and exon skipping (to give the beta globin band). In the remaining lanes, there are no full-length bands; only the cryptic 5'ss and beta globin products are visible.

The new primer combinations enabled me to detect the splicing inclusion of full-length *Esco1* exon 3 figure 4.30. Transfection with GFP (lanes 1-3) revealed three bands in Figure (4-31), representing the skipping exon, the full length splicing of exon 3, and the use of the cryptic splice site. Then, transfection with RBMXL2 (lanes 4-6) revealed two bands: the insert skipping and the full length of exon 3, indicating that RBMXL2 (lanes 4-6) clearly suppressed both the use of the cryptic splice site and the full length exon. This means that RBMXL2 is not only suppressing the cryptic splice site, but also the full length exon inclusion. It's worth mentioning here that 2 out of the three replicates worked (4-6), which may affect the accuracy of the statistical test. Like RBMXL2, RBMXL2RRM-GFP (lanes 7-9) inhibits the use of the cryptic splice site and the full length exon. A stronger result was obtained for cells co-transfected with RBMX lanes (10-12). Very strong repression of both the full length and cryptic exon 3. RBMX was able to prevent exon inclusion. The above data showed that RBMX and RBMXL2 did suppress selection of a cryptic 5' splice site in *Esco1* exon 3, but surprisingly also suppressed selection of the whole exon 3 as well.

4.3.8 Investigation of RBMXL2, RBMX mechanism to regulate human *ESCO1* gene using Minigene assay

To investigate this further, I examined the splicing regulation of human *ESCO1* exon 4. This exon was selected because it was identified as an activated exon by RBMX from RNA-seq of MDA-MB-231 cells depleted for *RBMX* (Chile Siachisumo thesis, 2022). Using *RBMX* Knockdown RNA-seq we identified skipping of exon 4 in human *ESCO1* figure 4.32 (Chile Siachisumo thesis, 2022). Therefore, I wanted to test *ESCO1* splicing in human, specifically if RBMX expression will activate the inclusion of exon 4 within the minigene context. To test this I designed the cloning primers using similar steps to those used for mouse *Esco1*. Figure 4.33 shows the process.

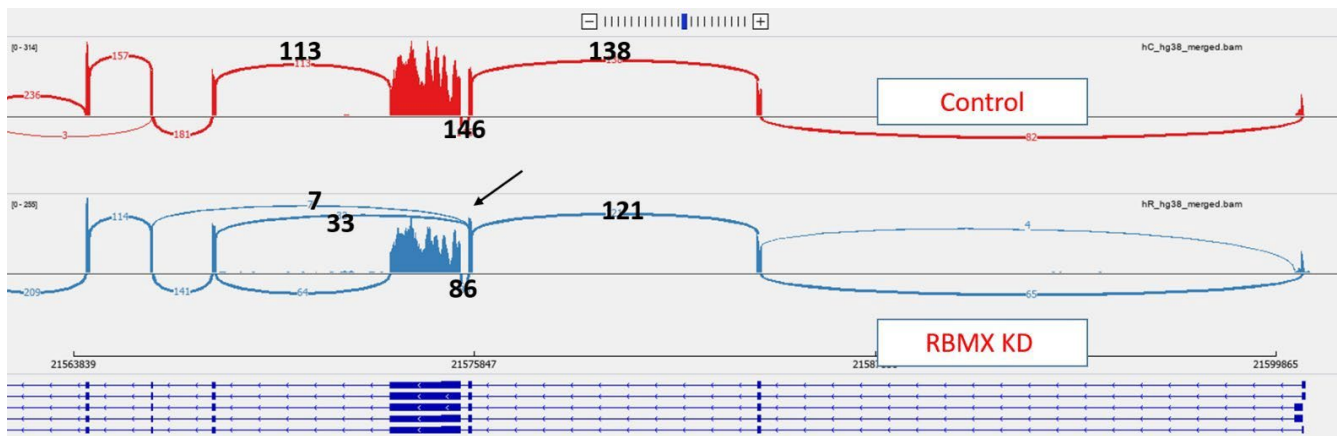


Figure 4. 28: Exon 4 of the *ESCO1* gene is activated by RBMX. A screenshot of a Sashimi plot shows control sample (normal expression of RBMX) and RBMX knockdown sample. The arrow shows clear a skipping of exon 4 of the *ESCO1* gene. The junction reads show more skipping after depleting *RBMX*, which is not seen in the control. This means that RBMX normally activates *ESCO1* exon 4. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

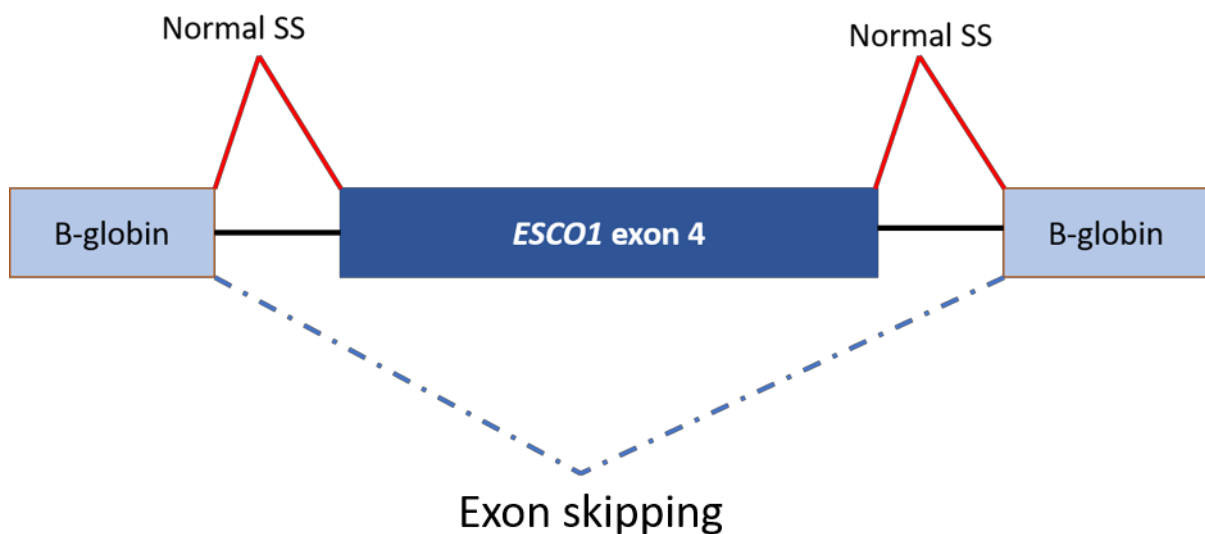


Figure 4. 29: A schematic of the *ESCO1* exon 4 minigene design. Exon 4 with flanking introns was cloned between two B-globin exons within the pXJ41 plasmid. Predicted two products would be 1) activation of exon 4 of *ESCO1*; (2) exon 4 skipping, joining the B- globin exons. Boxes indicates exons while black lines represents introns.

To investigate splicing regulation of *ESCO1* exon 4, similar to the previous minigene, I transfected cells using different expression vectors. These vectors expressed GFP (as an internal control), RBMXL2-GFP, RBMXL2 Δ RRM-GFP (RBMXL2 without the RNA recognition motif), RBMX- GFP and RBMX Δ Cterm- GFP as seen in the figure 4.27.

Triplicate transfections were performed to ensure that any changes in splicing could be replicated and were statistically significant. PXJF and PXJR primers were used in RT-PCR to analyse the RNA extracted from the transfected HEK-293 cells, which was then analysed by Qiaxcel capillary gel electrophoresis figure 4.34.

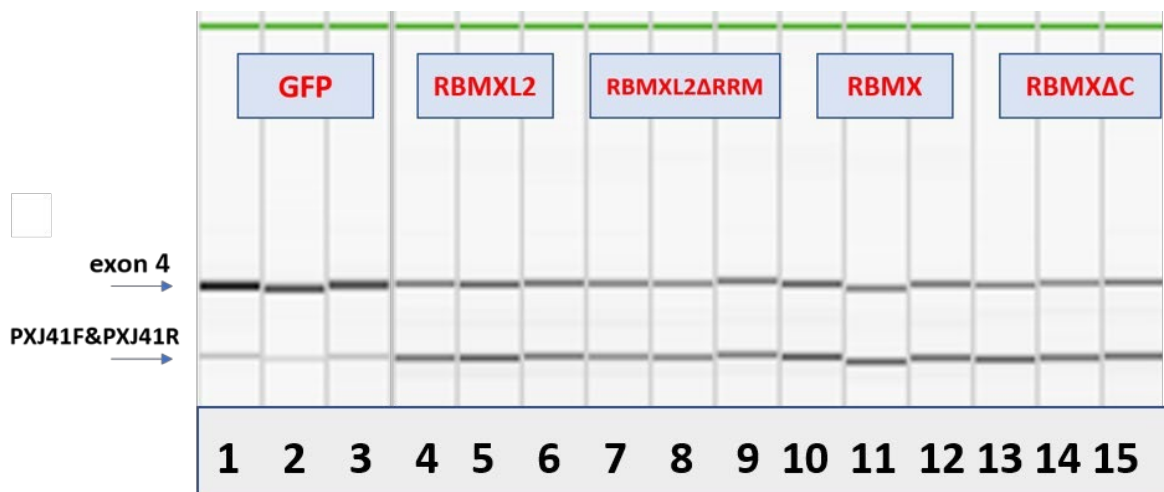


Figure 4. 30: RBMXL2 and RBMX suppress the splicing of exon 4 of human *ESCO1* gene (n=3).

Automated electrophoretic separation from QIAxcel machine, showing RT-PCR products from transfection of the human *ESCO1* minigene into HEK293 cells along with different expression vectors (encoding GFP (lanes 1-3), RBMXL2-GFP (lanes 4-6), RBMXL2 Δ RRM-GFP (lanes 7-9), RBMX-GFP (lanes 10-12), and RBMX Δ Cterm-GFP (lanes 13-15)). RT-PCR analysis was done similar to the previous minigene analysis 4.30.

Beginning with GFP (lanes 1-3) in Figure 4.34, two bands were observed: a prominent band representing exon 4 inclusion and a fainter band representing exon 4 skipping. RBMXL2 transfected cells (lanes 4-6) displayed two identical bands, one for the inclusion of exon 4 and the other for exon 4 skipping. There was reproducibly more of the lower band, representing exon skipping. RBMXL2 Δ RRM-GFP (lanes 7-9) showed similar results to RBMXL2. Similar to RBMXL2, RBMX (lanes 10-12) and RBMX Δ Cterm-GFP (lanes 13-15) produced similar results to

RBMXL2. Exon 4 splicing inclusion was reduced in all transfections except for cells expressing GFP (lanes 1-3) (control).

A PSI calculation was done using the equation mentioned earlier in 2nd chapter (materials and methods 2.6.10), and then statistical analysis was performed in Graphpad Prism, as seen in the figure below 4.35A. Finally, I did a Western blot to check that plasmids are expressing similar level of protein in each condition, as indicated in figure 4.35B.

Since RBMX activates human *ESCO1* exon 4, I originally hypothesised that by introducing RBMXL2 and RBMXL2 Δ RRM-GFP, RBMX and RBMX without the C-terminus, I would see more exon 4 inclusion. However, with the exception of GFP (control), we saw less exon 4 splicing inclusion when cells were transfected with RBMX or RBMXL2 proteins compared with GFP. This means that RBMX is repressing the splicing inclusion of human *ESCO1* exon 4. No difference was detected between RBMXL2 and RBMXL2 Δ RRM, which might suggest that the RRM is not essential for the inclusion of *ESCO1*. These experiment suggest that RBMX and RBMXL2 are mainly splicing suppressers in minigene experiments, but are also not reproducing how the RBMXL2 protein works in mouse testis.

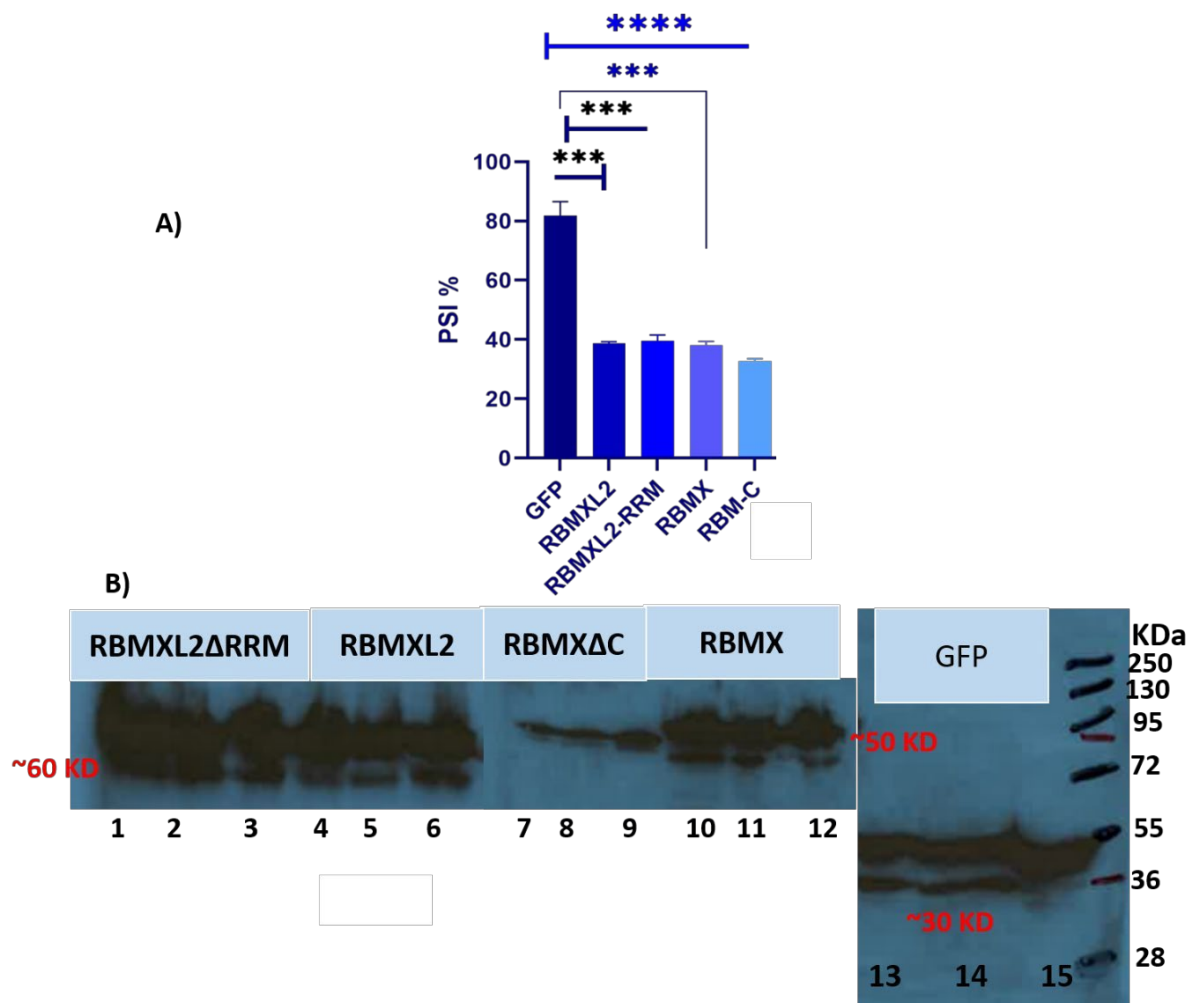


Figure 4. 31 :RBMXL2 and RBMX supresses the activation of exon 4 in human ESCO1 gene (n=3). A) Statistical analysis showed the P value for each condition on *ESCO1* exon 4 expression. Data from three biological replicates were used to calculate the PSI. The mean SEM is shown as lines in bar plots. Furthermore, P-values were calculated using an unpaired t-test on GraphPad Prism v. 9.31 (*, **, ****, p-value 0.05). RBMXL2, RBMXL2ΔRRM, RBMX and RBMXΔCterm shows reversed affect comparing to the control GFP. B) Shows the amount of protein in each condition in triplicates. Lanes 1-6 are from one membrane, lanes 7-12 from a different membrane, lanes 13-15 from a third membrane. However, RBMXL2 and RBMXL2ΔRRM are not clearly visible (because the membrane is over-exposed), and no loading control was performed for this experiment.

4.4 Discussion

In earlier work RBMXL2 was identified to work mainly as a cryptic splice site suppressor (Ehrmann et al., 2019). In this chapter, I used the iCLIP protocol to identify RBMXL2 protein-RNA binding sites in the mouse genome down to the single nucleotide. I then carried out minigene experiments with the aim of using this to dissect the function of specific binding sites in splicing repression.

4.4.1 RBMXL2 iCLIP in mouse testes

After a preliminary result was produced from one replicate using HITS-CLIP previously in Professor Elliott's lab (Ehrmann et al., 2019), one of the main goals of my project was to conduct a global analysis of RBMXL2 in mouse testes. First, I identified RBMXL2's genome-wide RNA targets using the iCLIP method. This required some troubleshooting. The Western blot technique was used to test the RBMXL2 antibody, and the iCLIP experiment was then adjusted for RBMXL2 immunoprecipitation and cDNA amplification. SDS-PAGE and transfer showed that the RBMXL2 antibody was specific and could accurately detect an RBMXL2-RNA crosslink. Additionally, barcoded sequences were used in the computational analysis to distinguish unique cDNA. The generated iCLIP data is consistent across 4 testis samples, and there was a strong correlation between all four biological replicates.

4.4.2 RBMXL2 binding within the genome

I analysed the genome-wide distribution of RBMXL2 binding sites using the iCLIP technique. Based on where RBMXL2 binds in the genome, I found that it binds most often in intergenic and intronic regions. Also, enrichment binding was seen in pre-mRNA introns, indicating a role in splicing. I also looked at candidate genes that we had originally identified as targets for RBMXL2 processing (Ehrmann et al., 2019). Many iCLIP tags were detected close to cryptic splice sites. *Meioc*, for example, shows many iCLIP tags within exon 5; some iCLIP tags were observed close to the cryptic 5' and 3' splice sites, which form the exitron. However, no iCLIP tags were found in *Brca2*, which might mean that RBMXL2 indirectly regulates cryptic splice sites or that iCLIP is perhaps not fully comprehensive. Finally, from iCLIP data RBMXL2 was found to bind preferentially and regulate longer exons.

4.4.3 Identifying RBMXL2 binding motifs

A Kmer analysis performed by Ivaylo Yonchev with my data identified that RBMXL2 binds to regions rich in TTT nucleotides. This may suggest a role in binding at branch points or the 3' splice site of introns. A metagene analysis of the iCLIP tags showed high enrichment of RBMXL2 binding in the 3' UTR of the transcript, which might be involved in alternative 3' end formation.

4.4.4 RBMXL2-bound genes that have a role in spermatogenesis and meiosis

Gene Ontology analysis showed that RBMXL2-bound genes are involved with spermatogenesis, tail assembly, and movement. The testis-specific *Meioc*, for example, is described in the literature as essential for meiosis and mice lacking *Meioc* are infertile (Soh et al., 2017). Another example is *Esco1*, which is implicated in sister-chromatid cohesion in cells (Alomer et al., 2017). The *Alms1* gene is involved in cell division and cilia formation; *Alms1*-deficient mice are sterile (Arsov et al., 2006), indicating this gene's significance in spermatogenesis. In addition, another tumour suppressor gene, *Brca2*, is involved in homologous recombination; deletion of *Brca2* rendered mice infertile with no sperm observed (Sharan et al., 2004).

4.4.5 Mouse and human *ESCO1* minigene assays did not follow the expected pattern of splicing regulation by RBMX and RBMXL2

In this chapter, I attempted to study the regulation of splicing by different RNA-binding proteins, such as RBMXL2 and RBMX in the *Esco1* gene. However, the result using minigene differed from what we expected from the RNA-seq. For mouse *Esco1*, the RBMXL2 and RBMX proteins suppressed the cryptic 5' splice site on exon 3; however, they also suppressed the inclusion of the full-length exon 3. This led to the investigation of human *ESCO1*, in which the activation of exon 4 was observed when RBMX was expressed within cells. However, after transfection with different proteins (RBMX and RBMXL2), the inclusion of exon 4 was suppressed. Previous data had suggested that RBMX binds to A-rich sequences, based on NMR which showed that RBMX binds to an AAN regions (Moursy et al., 2014). Also, Professor Elliott's lab showed that RBMX binds to A-rich regions via iCLIP (Chile Siachisumo thesis, 2022).

However, the minigene experiment did not replicate the splicing regulation of RBMXL2 and RBMX. Moreover, there are no differences between RBMXL2 and RBMXL2 without the RRM, suggesting that the RRM is not essential for splicing control. This finding is also consistent with the findings of Liu et al. (2017), who showed that the disordered domain of RBMX, not the RRM, is responsible for RNA binding (Liu et al., 2017).

A possible reason for these discrepancies is that the minigene is an artificial system that does not replicate endogenous gene regulation in vitro. Plasmid DNA in minigene assays is within a double helix, but not organised into chromatin. Chromatin conformation has been found to influence splicing. Modifications to histones, like H3K36me3, play a role in transcription by influencing splicing activity. In addition, these modifications have been found in weakly expressed and alternatively spliced exons, indicating that they regulate the splicing outcome (Wilhelm et al., 2011; Braunschweig et al., 2013).

Another possible reason why the minigene experiments did not work is that we might have missed some intronic regions within our minigene constructs that affect alternative splicing. To study the splicing regulation of RBMXL2, we need to use a stable cell line to mimic the endogenous genes in vitro. These experiments are described in the next Chapter. Using an overexpression approach, where a gene is introduced into cells in excess, comes with complexities when compared to knockdown methods that reduce gene activity.

Overexpression of a gene can disrupt normal cell processes, making it difficult to understand its effects. It may also result in the formation of non-functional protein complexes and even cellular damage, complicating the determination of the gene's precise function.

Overexpression can also unintentionally affect other genes, making data interpretation difficult.

In contrast, knockdown methods such as RNA interference or CRISPR provide more precise control over gene activity with fewer side effects. They generally excel at avoiding unintended consequences for unrelated genes or processes. However, it is essential to note that knockdown methods have their own limitations, such as the possibility of off-target effects and incomplete gene suppression (Zhang et al., 2003; Prelich, 2012; Tavleeva et al., 2022).

4.5 Chapter summary:

I identified genome-wide RBMXL2-RNA targets and RBMXL2 binding sites in mouse testis using iCLIP. The global analysis revealed that RBMXL2 binds to TTT-rich regions. Furthermore, a metagenome analysis of the iCLIP tags revealed that the transcript's 3' UTR enriched RBMXL2 binding. Finally, a gene ontology analysis revealed that RBMXL2-bound genes are involved in spermatogenesis and reproduction.

Chapter 5 : RBMXL2 is functionally interchangeable with RBMX in somatic cells.

5.1 Introduction:

RBMXL2, RBMX and RBMY are a family of RNA-binding proteins containing an N-terminal RNA recognition motif (RRM) and a C-terminal disordered domain (including a small region for recruitment of nascent RNA and a C-terminal region used for mediating RNA-Protein interactions, Figure 1.8 (Elliott et al., 2019). The *RBMX* and *RBMY* genes are located on the X and Y chromosome respectively. *RBMX* and *RBMY* are ancient genes that evolved around 300 million years-ago (Lahn et al., 1997). RBMX is ubiquitously expressed in all tissues. *RBMY* has a testis-specific expression which suggests a link to spermatogenesis (Delbridge et al., 1999). *RBMXL2* is intron-less, a retrogene derived from *RBMX* that evolve 65 million years ago (Elliott et al., 2000). RBMXL2 is expressed during meiosis when the X and Y are inactivated (Delbridge et al., 1999). For many years it was not known what biological processes these RNA binding proteins were involved in. Then the Elliott lab showed RBMXL2 is able to suppress cryptic splice site selection in some long exons (sometimes more than 1 KB in length) within genes implicated in genome stability (Ehrmann et al., 2019). RBMXL2 and RBMX proteins have 73% protein similarity, and they share similar structure as aforementioned. *RBMX* and *RBMXL2* are both closely related to a gene on the Y chromosomes long arm named *RBMY*, which was found mutated or deleted in some infertile men (Ma et al., 1993; Elliott et al., 1997). Similar to *RBMXL2*, recent data to which I contributed (Siachisumo et al, submitted 2023), showed that RBMX is functioning as a splicing repressor in human cell lines. Moreover, RBMX is important in repressing cryptic exons in ultra-long exons. Splicing changes were identified in the *Rbmxl2* knockout mouse compared to the wild type using day 18 testis transcriptomes (Ehrmann et al., 2019). Many splice events were repressed, but some were predicted to be activated by RBMXL2. RBMX activates splicing of human *ESCO1* exon 4.

In the previous chapter, I described experiments that attempted to recapitulate these splicing patterns using minigenes, including cloning exon 3 of *Esco1* into an exon trap vector and co-transfecting into HEK293 cells with RBMXL2. The prediction was that the RBMXL2 protein, should repress the cryptic splice site of *Esco1* exon 3, but we observed repression of the whole exon (which was not predicted). The same experiment was repeated for human *ESCO1* exon 4, and again the result was not as expected: instead of human *ESCO1* exon 4 being activated by RBMX and RBMXL2, within minigenes exon repression was seen. So, these results show the minigenes are not working to recapitulate what is going on in vivo. This suggested that minigenes were not an appropriate route to investigate the RBMXL2 function. In this chapter, I attempted an alternative approach using stable cell lines to replace *RBMX* function after *RBMX* had been depleted using siRNA.

Once established, I used this approach to test if RBMXL2, RBMX, and RBMY performed the same functions. In addition, I tested whether splicing regulation in stable cell lines depended on the RRM or the disordered domain of RBMXL2. An advantage of using a stable cell line approach will allow for the analysis of splicing patterns in the endogenous genes that would have their normal chromatin structure. RBMX has also been found to be associated with chromatin (Matsunaga et al., 2012). RBMX was found to be enriched within heterochromatin and was involved with a repressed chromatin marker (H3K9me3) to prevent gene transcription (Becker et al., 2017). However, a disadvantage is that functional testing by making mutations for genes in situ on chromosomes is more challenging than making mutations for minigenes.

To test whether the rescue approach was working, I looked at both activated and repressed exons. In the minigene context, both these exons were repressed by RBMX and RBMXL2. However, if the rescue experiments were really working, I should be able to induce activation and repression as appropriate of exons within endogenous genes. I further hypothesised that if RBMXL2 and RBMY are performing the same functions as RBMX, they should be able to rescue the effects of *RBMX* knockdown in HEK293 cells. This hypothesis was tested using a set of inducible cell lines. For example, I should see the activation (inclusion) of exon 4 in *ESCO1* by RBMXL2 over-expression. But if RBMXL2 is not replacing RBMX (restoring its function), I would expect exon 4 to be repressed or left out. The same principle applies to the rest of the targets.

Chileleko Siachisumo (former PhD student from our lab) provided some of the basis for the experiments this Chapter. Chile depleted *RBMX* in MDAMB-231 cells and using RNAseq identified some genes controlled by *RBMX* that are implicated in meiosis, sister chromatid separation, and replication. Below chord plot for Gene Ontology figure 5.1 and a table 5.1 shows some important *RBMX* targets with their functions (Siachisumo et al, submitted 2023). For example, *ETAA1* (Ewings Tumour Associated Antigen 1) gene, requires *RBMX* to prevent a cryptic 3' splice site that reduces the size of its exon 5. *REV3L* contains a similar cryptic splice event to *ETAA1* within an ultra-long exon, and encodes a protein essential in chromosome break repair (Ben Yamin et al., 2021). *RBMX* knockdown caused an activation or recognition of a 3' cryptic splice site normally suppressed by *RBMX*. *ATRX* (alpha thalassemia patients with a mental retardation syndrome X-linked) is another example of an ultra-long exon (3074 bp). According to Nandakumar et al. (2017), the *ATRX* protein is crucial for maintaining genome stability, cell cycle regulation, and chromatin remodelling (Nandakumar et al., 2017). The depletion of *RBMX* boosts the use of an exitron in *ATRX* exon 9. The term "exitron" refers to a region inside the exon that develops into an intron. In contrast *ESCO1* exon 4 is activated by *RBMX*. The *ESCO1* gene encodes an essential protein for sister chromatid cohesion, DNA repair and formation of the chromosome loop as well as transcriptional control (Alomer et al., 2017). Depletion of *RBMX* was found to increase the skipping of ultra-long exon 4 (longer than 2000 bp), and *ESCO1* was enriched with *RBMX* tags. However, in the presence of *RBMX*, the inclusion of that exon was seen.

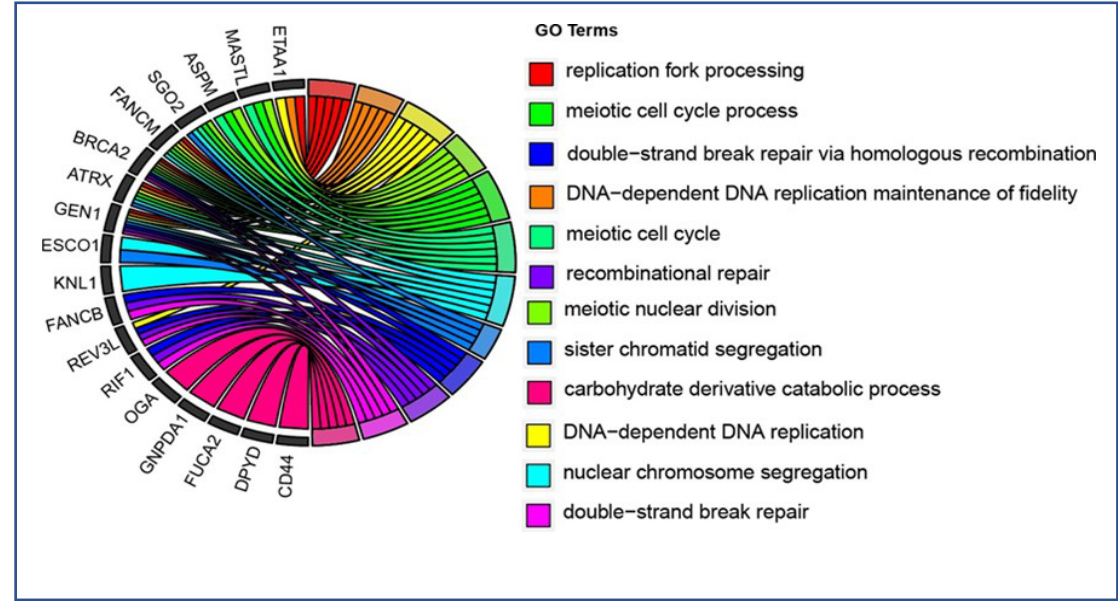


Figure 5. 1: Chord plot showing Gene Ontology of enriched genes involved in cell cycle and DNA repair within *RBMX* regulated exons.

Gene	Function	RNA processing event	Exon length (nt)
<i>ETAA1</i>	(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
<i>LRIF1</i>	LRIF1 (Ligand Dependent Nuclear Receptor Interacting Factor 1) is a retinoic acid receptor binding. LRIF1-HP11 α interaction is important for accurate cell division and mitotic cell division (Akram et al., 2018).	Skipped alternative exon activated by RBMX	1527
<i>ATRX</i>	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
<i>REV3L</i>	reversionless 3-like (REV3L) Protein ensures progression of replication forks and prevents spontaneous chromosome break formation (Ben Yamin et al., 2021).	3'ss repressed by RBMX	4162
<i>ESCO1</i>	ESCO1 is involved in cohesion functions during cell cycle progression (Alomer et al., 2017).	Skipped alternative exon activated by RBMX	2117

Table 5.1: functionally important RBMX RNA processing targets.

5.2 Hypothesis:

After the failure of the minigene to reproduce the splicing changes that were predicted from study of the *Rbmxl2* knockout mouse compared to the wild type using day 18 testis transcriptomes, I hypothesised that a stable cell line might work better. This chapter's primary aim is to determine whether the inducible cell line approach can actually be used to determine whether the RBMX family, (which consists of RBMXL2, RBMX, and RBMY) is capable of carrying out the same functions in a somatic cell. I hypothesised that RBMXL2 and RBMY are replacing the function of RBMX in somatic cells. Additionally, I investigated whether the disordered domain or the RRM of RBMXL2 are required for rescue of RBMXL2 splicing functions.

5.3 Aims:

- 1- To establish an experimental system using stable tetracycline-inducible cell lines.
- 2- 2- Investigate whether RBMX family proteins can substitute for each other in somatic cells.

5.4 Results:

5.4.1 Establishment of a new strategy to investigate RBMXL2 using an inducible stable cell line approach

To first establish this protocol I used a stable cell line from our lab that can overexpress RBMXL2-FLAG. A stable cell line (Flp-In HEK293) was used.

The rescue protocol starts with plating two 6-well plates—one as a control (without expression) and another for overexpression. The main difference between the two plates is the addition of tetracycline. First, cells were seeded at 100,000 cells per well. Then, I performed *RBMX* depletion in the first 3 wells using an siRNA directed against *RBMX* using Lipofectamine™ RNAiMAX Transfection Reagent (Invitrogen). The other wells were transfected with a control siRNA to compare with the knockdown of *RBMX* (section 2.8) contains the siRNA. The 6 well plates were kept in incubators at 37 °C with 5% CO₂ for 48 hrs. Then, the experimental plate was induced with tetracycline to overexpress the RBMXL2-FLAG. After adding the tetracycline (1mg/ml) I continued the incubation for 24 hrs before cells were harvested for RNA

extraction (to be used for RT-PCR mentioned in section 2.6.12) and protein extraction (for a western blot mentioned in 2.6.14). Details of the experiment are shown in the figure below 5.2. I used Western blot analysis to confirm the overexpression of RBMXL2, and to confirm the depletion of *RBMX* compared to a house keeping protein (tubulin) as seen in the figure below 5.2. However, Western blot showed that RBMX knockdown was modest. Despite the low knockdown level achieved, it significantly impacted RBMX overall activity and functionality.

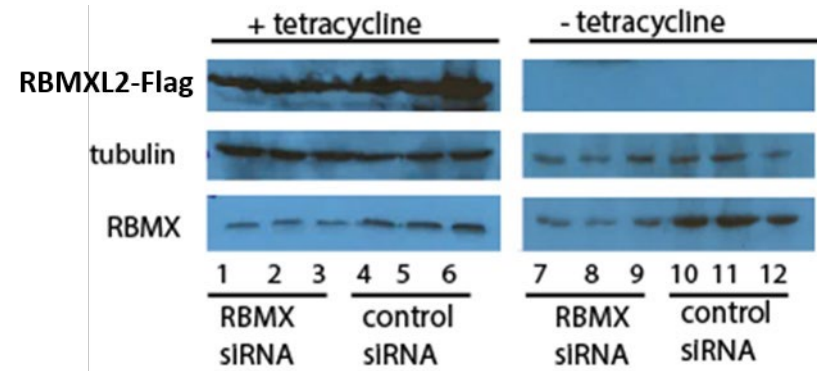


Figure 5. 2: A western blot showing the successful knockdown of RBMX compared to the siRNA control and tubulin as a control. This western confirms the knockdown of *RBMX*. Detection of the FLAG-fusion protein indicates that the tetracycline-induced HEK293 cells express RBMXL2.

5.4.2 RT-PCR analysis of some initial targets of RBMX to establish if the rescue experiment was working:

Initially, we were not sure if *RBMXL2* would be able to rescue splicing defects after *RBMX* depletion. To test this, I followed the experimental strategy of depleting *RBMX*, then inducing *RBMXL2* by tetracycline for 24 hours, followed by RNA extraction. I first tested some candidate genes using RT-PCR. These candidate genes had predicted splicing defects after *RBMX* depletion (Siachisumo et al, submitted 2023), Therefore, I selected an initial group of genes to test splicing activation and repression patterns by RBMX from the chord plot mentioned above 5.1. I used the primers from my colleague Chile's thesis as seen in the table below 5.2, and I analysed splicing patterns of the genes shown in figure 5.3. I used the one-step PCR kit from Qiagen following the manufacturer's protocol, and samples were analysed using capillary gel electrophoresis figure 5.4.

PCR primer sequences			
Gene	Forward primer	Reverse primer	Internal primer
<i>ETAA1</i>	5'-GCTGGACATGTGGA TTGGTG	5'GTGGGAGCTGCATTT ACAGATG	5'-GTGCTCCAAAAAG CCTCTGG
<i>REV3L</i>	5'-TCACTGTGCAGAAA TACCCAC	5'-AGGCCACGTCTACAA GTTCA	5'-ACATGGGAAGAAA GGGCACT
<i>ATRX</i>	5'TGAAACTTCATTTTC AACCAAATGCTC	5'ATCAAGGGGATGGCA GCAG	—
<i>ESCO1</i>	5'TGCTGTGTGATGTGT TAAGAGC	5'ACTGAGCAGAGTCGA AGCAG	5'AAAACCACAGAAAG GGCTCC
<i>LRIF1</i>	5'GCAACTCCAGCCCA ACATTC	5'TTGTCATTTCTTCCTT GGGCAC	—

Table 5.2: list of PCR primer sequences used for RT-PCR analysis of *RBMX*-regulated exons.

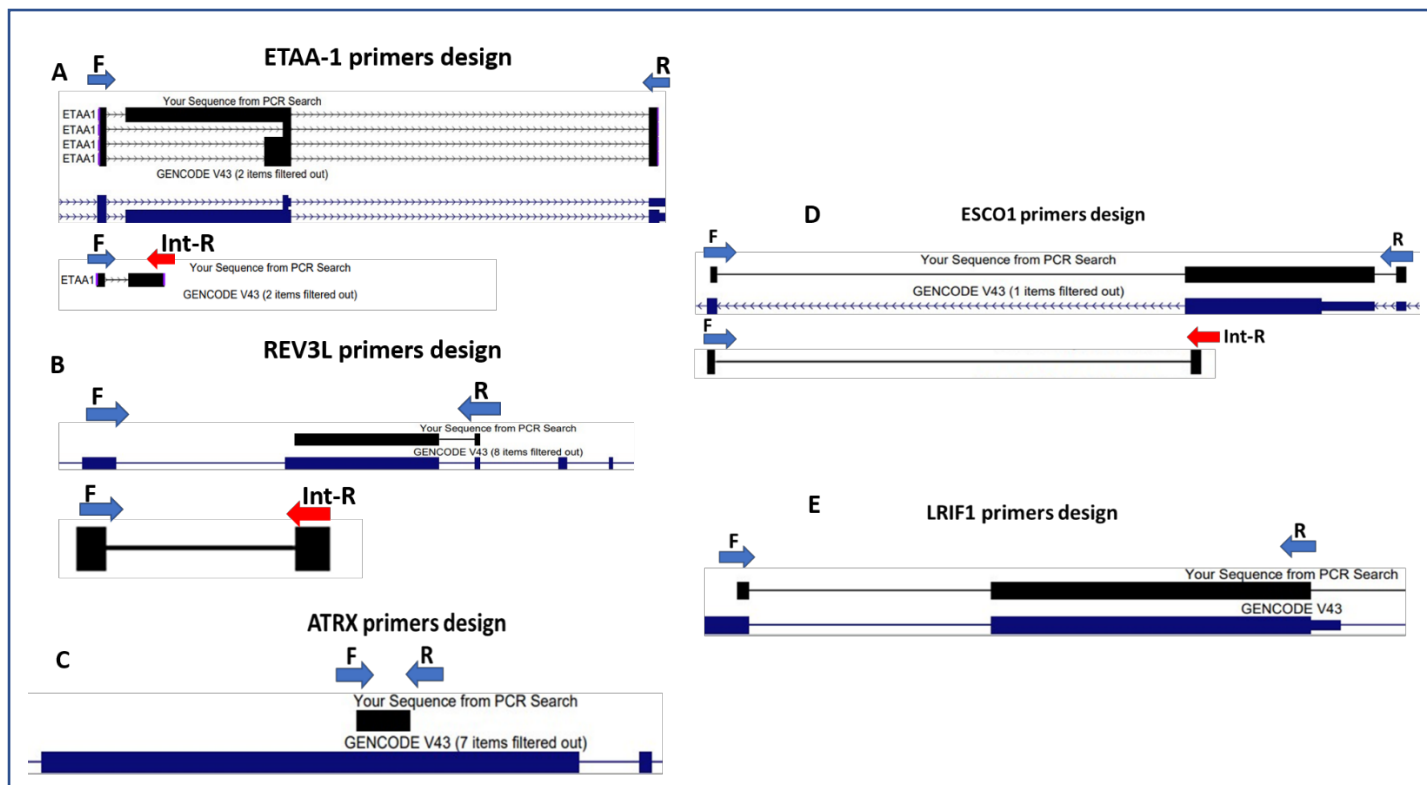


Figure 5. 3: A snapshot of primers designed from the UCSC website

(<http://genome.ucsc.edu/>). Some of these primers were designed by my colleague Chileleko Siachisumo. A) Set of *ETAA-1* primers. Reverse transcription-PCR using *ETAA-1F* & *ETAA-1R* will detect the alternative exon (short product only), as the full length exon is longer than 1 KB and can't be efficiently amplified. Therefore, another internal *Int-R* is used to detect the expression of the long exon. B) RT-PCR analysis of *REV3L*. Similar to *ETAA-1*, the first set of primers *REV3LF* & *REV3LR* will only detect the alternative exon as the long exon is longer than 1 KB and can't be efficiently amplified. An internal *REV3L* primer called *int-R* is utilised to show the expression of the long exon. C) The *ATRX* gene formed an exitron within exon 9 if *RBMX* was depleted. Therefore, I used *ATRXF* and *ATRXR* to analyse the exitron region. If there are two bands, that means the exitron is being removed. D) A set of primers to detect the activation of the *ESCO1* gene exon 4. The first set containing *ESCO1F* and *ESCO1R* will show the skipping of exon 4 as the forward primer on exon 3, and the reverse primer is on exon 5. The size of the skipped exon 4 means that it is too big to amplify. Therefore, an internal-R is needed to show the activation of exon 4. So, the inclusion will show as one band, while the skipping will show as two bands. E) *LRIF1F* & *LRIFR* detect the expression of *LRIF* exon 2. *RBMX* activates this exon. If *RBMX* is depleted, one band will be detected, which means exon 2 is not activated. However, if *RBMX* were expressed, two bands would be present, showing the splicing of exon 2.

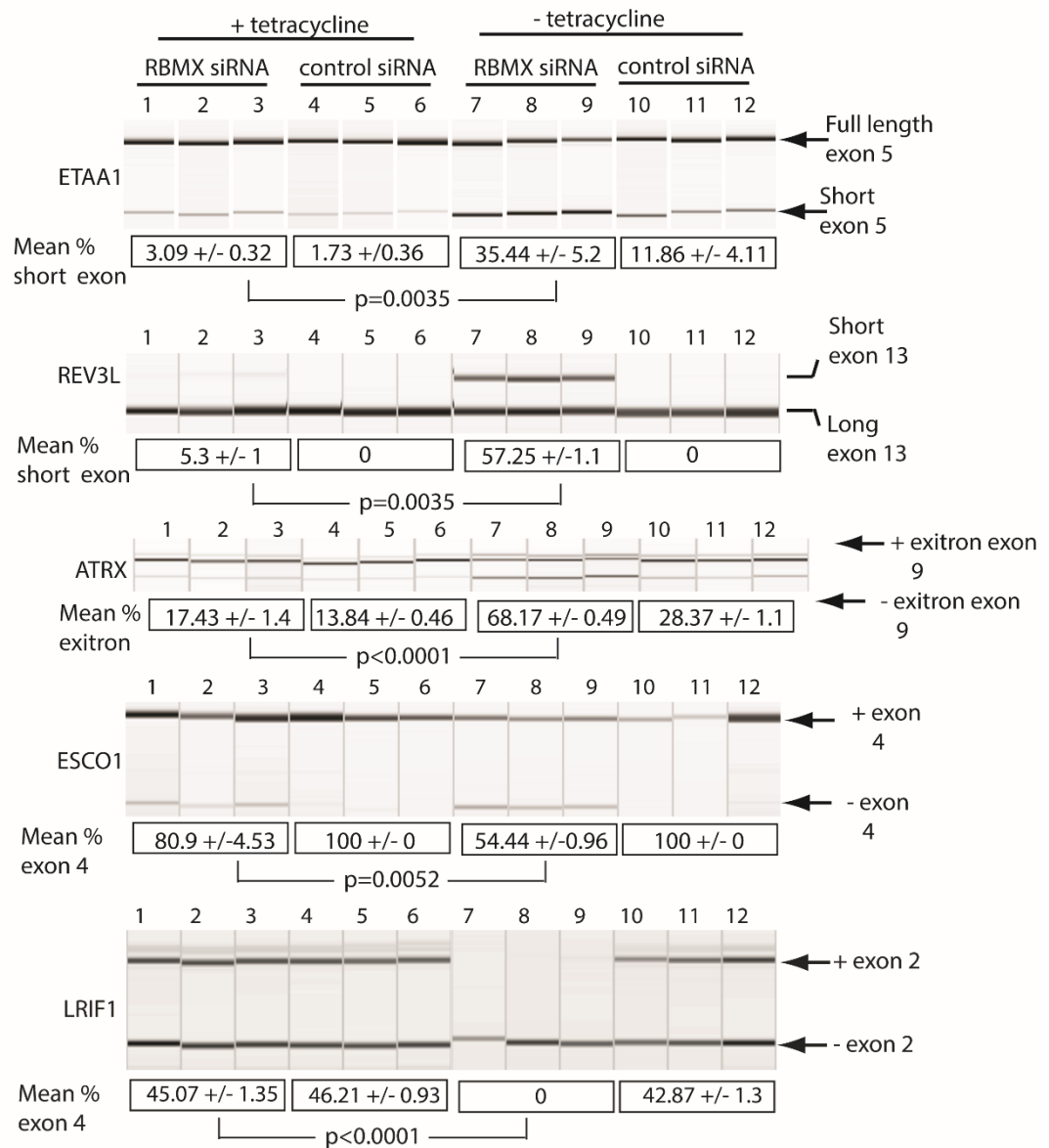


Figure 5. 4: Capillary gel electrophoretograms show RBMXL2's ability to substitute RBMX in somatic cells and ensure the proper splicing inclusion of ultra-long exons on different genes (Siachisumo et al., 2023). RT-PCR analysis following different experimental conditions for each gene was run in biological triplicate, comparing the *RBMX* siRNA to the siRNA control, and when tetracycline was added to induce *RBMXL2*, to check if *RBMXL2* is rescuing the use of the cryptic splice site, thus possibly replacing *RBMX* function. Boxes below each triplicate gel electrophoresis show the mean after calculating the PSI percentage. In addition, the P value was calculated for each gene to show the significance between *RBMX* Knockdown and *RBMX* Knockdown with *RBMXL2* overexpression (the P value was calculated using t-test by Graphpad). These experiments were performed by me and are part of the Siachisumo et al 2023 manuscript of which I am joint first author.

Our data demonstrated that RBMX facilitates the inclusion of the full-length exon 5 of the *ETAA-1* gene. *ETAA-1* is involved in genome stability, and the knockdown of *RBMX* caused an increased selection of the weak 3' splice site of exon 5, producing a shorter defective mRNA. I tested the impact of RBMXL2 on *ETAA-1* expression using RT-PCR from an inducible cell line that overexpressed RBMXL2 if Tetracycline was added. I found (lanes 7-9), where *RBMX* is depleted, showed an increase in the inclusion of the shorter isoform of exon 5, compared to (lanes 10-12) that were transfected with control siRNA. In parallel I analysed RNA samples from the other plate that was induced with tetracycline to overexpress RBMXL2. Samples (lanes 4-6) that were treated with the control siRNA were similar to samples (lanes 1-3), where *RBMX* was depleted and RBMXL2 was overexpressed. This showed restoration of RBMX function, as less inclusion of the short ETAA1 exon 5 was seen. In each case the PSI was calculated, and Graphpad prism was used to get the P value and to compare the *RBMX* knockdown in the control plate (without overexpression) with the *RBMX* knockdown and RBMXL2 overexpression. This analysis showed that result was significant, and indicating that RBMXL2 successfully replaced the function of RBMX.

Another example I tested was the *REV3L* gene, which is also implicated in cell cycle regulation and genome stability. Use of the 3' cryptic splice site of exon 13 was analysed after *RBMX* depletion. To test splicing of this gene using the stable cell line, I started by analysing the RT-PCR of the control plate. On (lanes 10-12) treated with siRNA control, amplification products from the long exon 13 were the only ones detected. However, in (lanes 7-9) where *RBMX* was knocked down using siRNA, the long isoform and the short isoform of exon 13 were seen, which indicates the increased use of the 3' cryptic splice site in the absence of RBMX protein. For the experimental plate where RBMXL2 was overexpressed after the induction of tetracycline, (lanes 1-3) show more of the long exon 13 and only a slight expression of the short exon 13. PSI values were calculated, and Graphpad prism was used to get the P value in order to compare the *RBMX* knockdown in the control plate (without overexpression) with the *RBMX* knockdown and RBMXL2 overexpression. The result was significant, showing that RBMXL2 successfully replaced the function of RBMX. Finally, (lanes 4-6), treated with the control siRNA, detected the long exon 13 only.

The third example I tested is *ATRX*. Our data had showed that when *RBMX* is depleted, the ultra-long exon 9 of the *ATRX* gene has an exitron that is usually not seen when *RBMX* is

present. This exon is formed due to the selection of cryptic 5' and 3' splice sites in exon 9. I used the same experimental approach to test if *RBMXL2* can restore the function of *RBMX* and prevent the formation of this exon. Two bands were seen on (lanes 10-12) made from cells treated with siRNA control. The expression of full length exon 9, and a slight expression of exon (a shorter product) were both detected. However, these two amplification products were more equal where *RBMX* was knocked down using siRNA, as seen on (lanes 7-9). This means depleting *RBMX* induces the exon. On the experimental plate where *RBMXL2* was overexpressed after tetracycline induction, the expression of the long version of *ATR*X exon 9 was strong, and only a faint band representing the exon was also detected, shown in (lanes 1-3). PSI values were calculated, and Graphpad prism was used to apply the t-test to get the P value in order to compare the *RBMX* knockdown in the control plate (without *RBMXL2* overexpression) with the *RBMX* knockdown and *RBMXL2* overexpression. The result was significant, showing that *RBMXL2* successfully replaced the function of *RBMX*. In fact, use of the exon in lanes 1-3 was very similar to that in cells treated with the control siRNA in (lanes 4-6) as seen in the figure 5.4.

RBMX activates splicing of exon 4 within the *ESCO1* gene. Therefore, the depletion of *RBMX* leads to skipping the ultra-long exon 4 (around 2.4KB). Multiplex primers were designed to allow the detection of exon 4. To test splicing of this gene using the stable cell line, I started by analysing the RT-PCR of the control plate. On (lanes 10-12) prepared from cells treated with siRNA control, the ultra-long exon 4 was included. However, two bands were detected in (lanes 7-9), where *RBMX* was depleted using siRNA. The upper bands represents exon 4 inclusion and the lower band increased skipping of exon 4 that was not seen in the presence of *RBMX*. On the experimental plate, *RBMXL2* was overexpressed after the induction of tetracycline. These samples (lanes 4-6) showed the inclusion of exon 4 only, while (lanes 1-3) showed even more of the inclusion of exon 4 and a slight expression of skipping of exon 4. The PSI was calculated, and Graphpad prism was used to apply the t-test to get the P value in order to compare the *RBMX* Knockdown in the control plate (without overexpression) with the *RBMX* Knockdown and *RBMXL2* overexpression. The result was significant, showing that *RBMXL2* partially replaced the function of *RBMX* as seen in the figure 5.4.

Another gene that contains an exon that is activated by RBMX is *LRIF1*. Depletion of *RBMX* leads to the skipping of the ultra-long exon 2. To test this gene using the stable cell line, I started by RT-PCR analysis of the control plate. On (lanes 10-12) from cells treated with siRNA control, two bands were detected: these are from inclusion and the skipping of exon 2.

However, in (lanes 7-9), when *RBMX* was depleted using siRNA, the skipping of exon 2 was the only amplification product detected. This shows that *RBMX* depletion prevents inclusion of *LRIF1* exon 2. On the experimental plate, *RBMXL2* was overexpressed after the induction of tetracycline. Samples from (lanes 4-6) that were from cells treated with the siRNA control and with overexpression of *RBMXL2* showed amplification products corresponding to the inclusion of exon 2, and also from the skipping of the same exon. While (lanes 1-3) that have the depletion of *RBMX* and the overexpression of *RBMXL2* showed more inclusion of exon 2 and, a slight expression of skipping of exon 2, with a pattern that is similar to the siRNA control. The PSI values were calculated, and Graphpad prism was used to get the P value in order to compare the *RBMX* knockdown in the control plate (without *RBMXL2* overexpression) with the *RBMX* Knockdown and *RBMXL2* overexpression as seen in the figure 5.4. The result was significant, indicating that *RBMXL2* replaced the function of *RBMX*.

5.4.3 RNA-seq data from knockdown of *RBMX* with and without overexpression of RBMXL2:

The above experiments using RT-PCR to analyse splicing patterns of specific target genes clearly showed that tetracycline-induced RBMXL2 expression was sufficient to rescue depletion of *RBMX*. In order to enable more global analysis of this I next used RNAseq to monitor gene expression changes at the transcriptome level. This global data would mean that I would be able to monitor splicing effects on any gene on the IGV genome browser. RNA extraction was performed on all samples (overexpressing RBMXL2 and those without overexpression) and then sent for sequencing in the Newcastle University Genomics Core Facility. I analysed a sample for each of the four conditions (O/E *RBMX* knockdown, O/E control, *RBMX* knockdown, and control). Below are Sashimi plot snapshots of genes of interest to check for splicing changes figures 5.5 to 5.9.

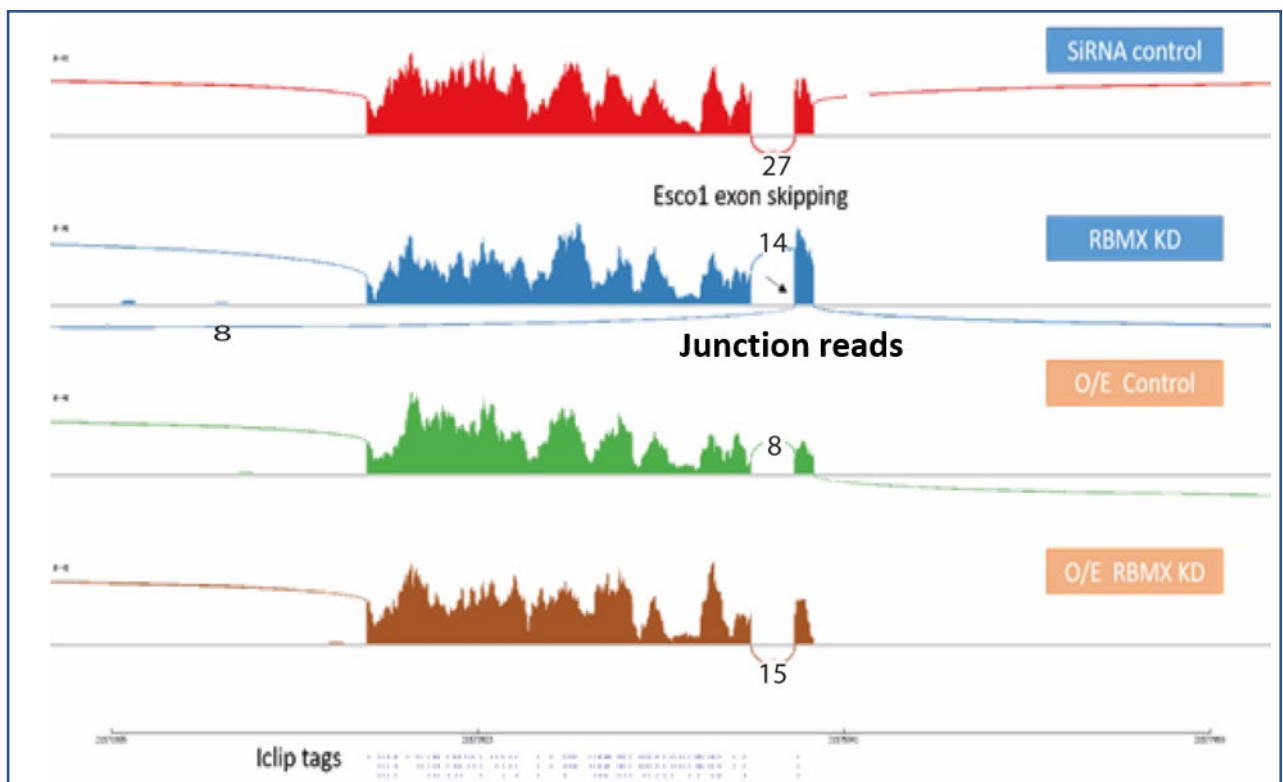


Figure 5. 5: RNAseq data used to analyse the effects of the rescue experiment on ESCO1 gene. The RNAseq data was visualised using a Sashimi plot. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

The Sashimi plot shows that RBMX activates the inclusion of exon 4 in the *ESCO1* gene. A snapshot from the IGV genome browser shows that the ultra-long *ESCO1* exon 4 (around 2.4KB) is included in the mock depletion, when the siRNA control has been used. However, the depletion of *RBMX* leads to skipping of the ultra-long *ESCO1* exon 4 (notice exon skipping junction reads). The third track is the siRNA control when tetracycline induction overexpresses *RBMXL2*. Again, the inclusion of exon 4 is seen. Tetracycline overexpressed *RBMXL2* is shown in the final condition, where siRNA has been used to deplete *RBMX*. No exon 4 skipping was seen, so the fact that exon 4 was included confirms that *RBMXL2* can replace the function of *RBMX* in somatic cells. These RNAseq results thus confirmed the RT-PCR results that were seen before.

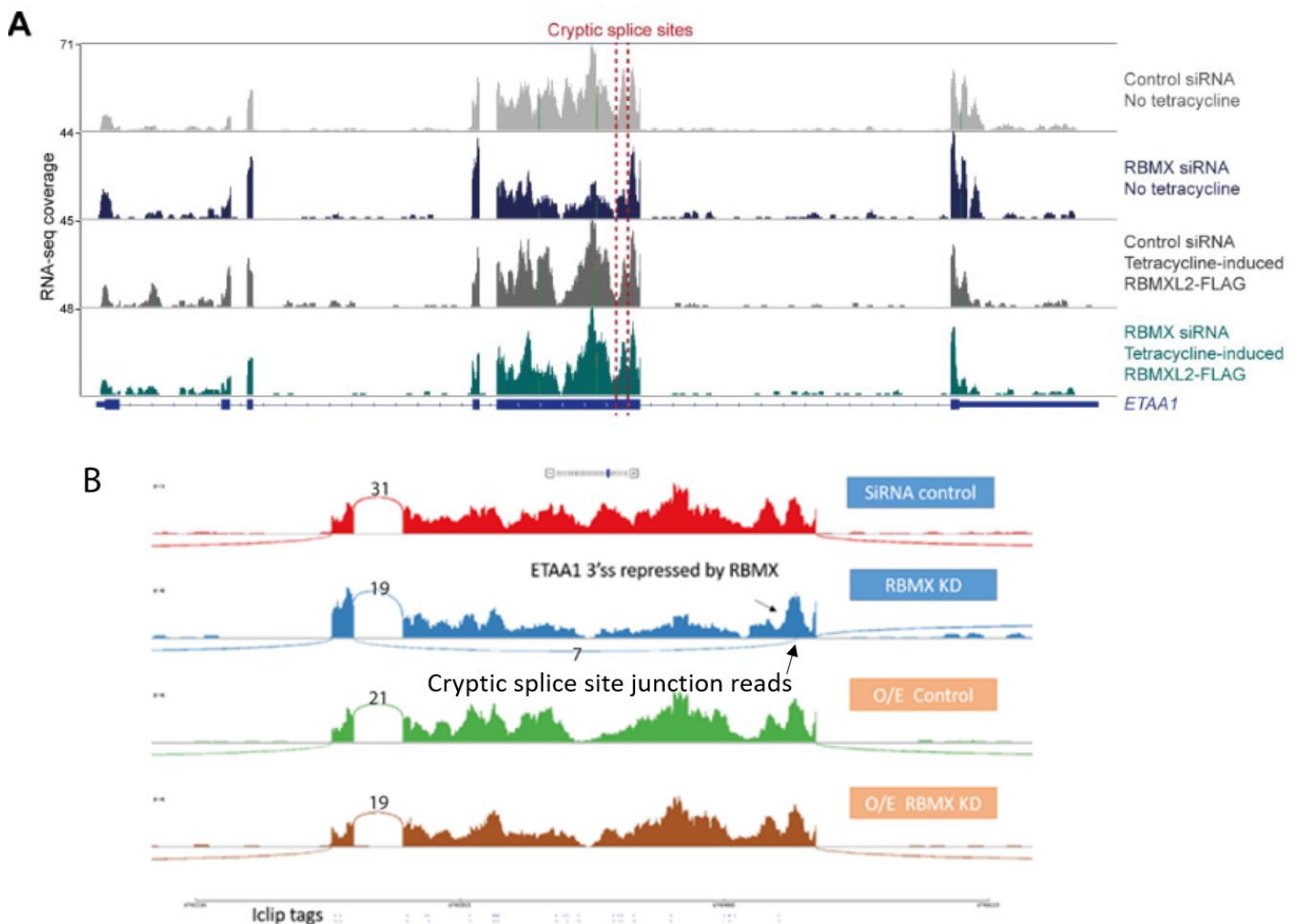


Figure 5. 6:RBMXL2 can substitute the function of RBMX in ensuring appropriate splicing and inclusion of ETAA1 exon 5.

RNAseq of *RBMXL2*-FLAG cells showing: Cells treated with control siRNA with no tetracycline; Cells treated with *RBMX* siRNA with no tetracycline (that should show the effect of the *RBMX* knockdown); Cells treated with control siRNA and tetracycline; Cells treated with *RBMX* siRNA and tetracycline (that should show the effect of the rescue). (A) A snapshot from IGV for *ETAA1* showing the use of a cryptic splice site in the *RBMX* knockdown that was not used after the induction of *RBMXL2*. (B) Sashimi plot of RNAseq data used to analyse the effects of the rescue experiment on *ETAA1*. The RNAseq data was visualised using a Sashimi plot. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

Another example I analysed is the *ETAA1* gene. A snapshot from the IGV genome browser shows that *RBMX* is repressing the use of the cryptic 3' splice site of exon 5. So, the full exon 5 was expressed normally in the first siRNA control condition. However, when *RBMX* is depleted by siRNA, this second condition uses the 3' cryptic splice site of exon 5, leading to a defective, shorter mRNA that was not seen if *RBMX* was expressed. The third condition represents siRNA with overexpression of *RBMXL2*, showing the normal inclusion of the full-length exon 5. Finally, the last condition is where *RBMX* is depleted using siRNA, and *RBMXL2* is overexpressed after tetracycline induction. Again, the full-length exon 5 is included, which means that *RBMXL2* is substituting for *RBMX* by suppressing the use of the cryptic splice site. The conclusion is that *RBMXL2* successfully replaced the function of *RBMX* that was seen in both RT-PCR and RNAseq.

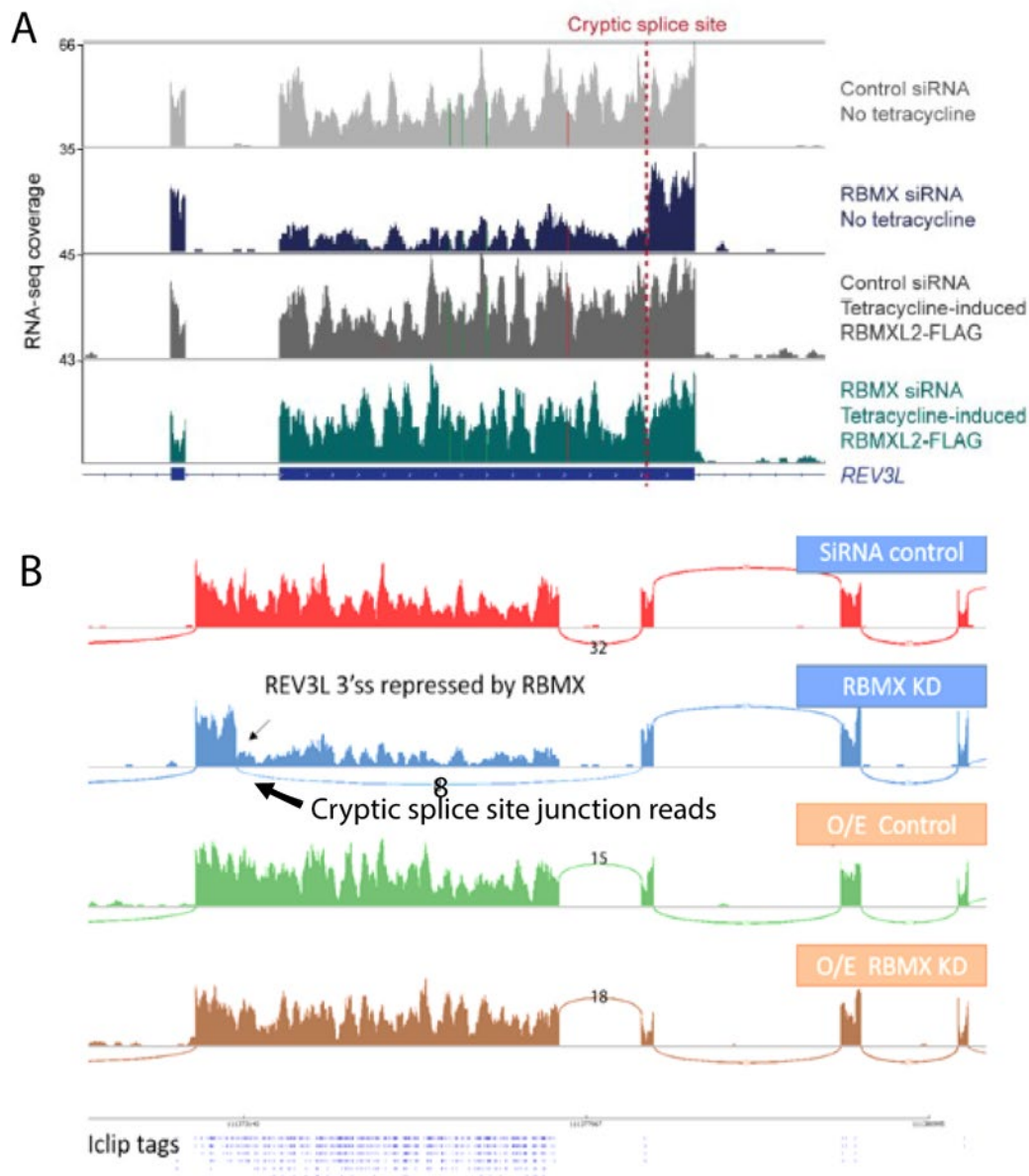


Figure 5. 7: RBMXL2 can substitute the function of RBMX in ensuring appropriate splicing and inclusion of REV3L. RNAseq of RBMXL2-FLAG cells showing: Cells treated with control siRNA with no tetracycline; Cells treated with RBMX siRNA with no tetracycline (that should show the effect of the *RBMX* knockdown); Cells treated with control siRNA and tetracycline; Cells treated with *RBMX* siRNA and tetracycline (that should show the effect of the rescue). (A) A snapshot from IGV for *REV3L* showing the use of a cryptic splice site in the *RBMX* knockdown that was not used after the induction of RBMXL2. (B) RNAseq data used to analyse the effects of the rescue experiment on *REV3L*. The RNAseq data was visualised using a Sashimi plot. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011). The numbers of exon junction reads are shown.

Another example is the *REV3L* gene, which is shown in figure 5.8, in a snapshot from the IGV genome browser. This shows that *RBMX* is repressing the use of the cryptic 3' splice site of exon 13. So, the full exon 13 was expressed normally in the first siRNA control condition. However, when *RBMX* is depleted by siRNA, the result is use of the 3' cryptic splice site of exon 13, leading to a defective, shorter mRNA that was not seen if *RBMX* was expressed. The third condition represents control siRNA with overexpression of *RBMXL2*, showing the normal inclusion of the full-length exon 13. Finally, the last condition is where *RBMX* is knocked down using siRNA, and *RBMXL2* is overexpressed after tetracycline induction. Again, the full-length exon 13 is included, which means that *RBMXL2* is substituting for *RBMX* by suppressing the use of the cryptic splice site. The conclusion is that *RBMXL2* successfully replaced the function of *RBMX* that was seen in RT-PCR and RNAseq.

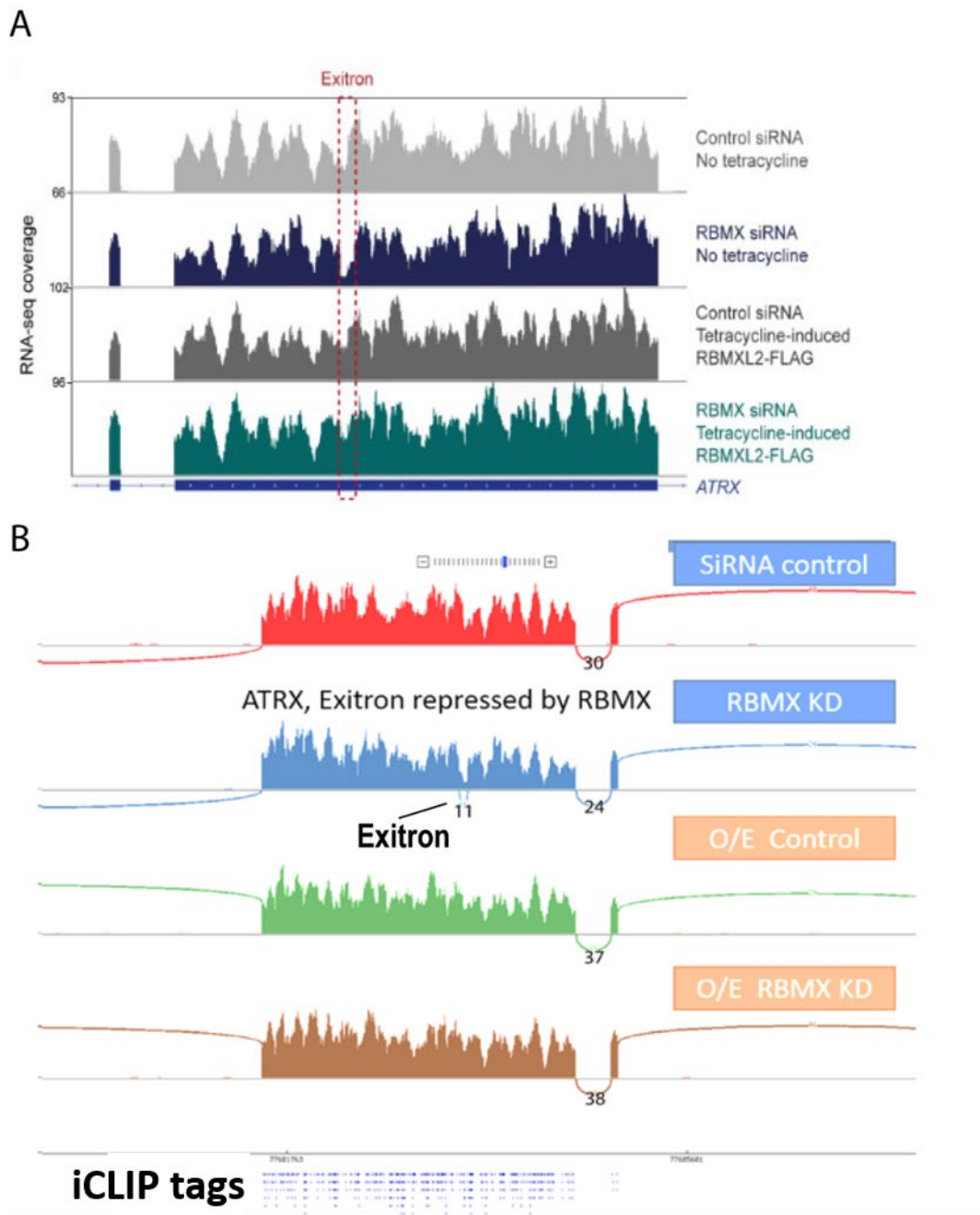


Figure 5. 8: RBMXL2 can substitute the function of RBMX in ensuring appropriate splicing and inclusion of ATRX exon 9. RNAseq of RBMXL2-FLAG cells showing: Cells treated with control siRNA with no tetracycline; Cells treated with *RBMX* siRNA with no tetracycline (that should show the effect of the *RBMX* knockdown); Cells treated with control siRNA and tetracycline; Cells treated with *RBMX* siRNA and tetracycline (that should show the effect of the rescue). (A) A snapshot from the IGV genome browser for the *ATRX* gene showing the use of an exon in the *RBMX* knockdown sample that was not used after the induction of RBMXL2. (B) Sashimi plot of RNAseq data used to analyse the effects of the rescue experiment on *ATRX*. The RNAseq data was visualised using a Sashimi plot. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011).

A snapshot from the IGV genome browser shows that *RBMX* is repressing exon formation within exon 9 of the *ATRX* gene. The full *ATRX* gene exon 9 was expressed normally in the first siRNA control condition. However, when *RBMX* is depleted by siRNA, in the second condition, the formation of exon, is seen, leading to a defective, shorter RNA that is not seen if *RBMX* is expressed. The third condition is control siRNA with overexpression of *RBMXL2*, which shows the full-length exon 9 as it is usually expressed. Finally, the last condition is where *RBMX* is knocked down using siRNA, and *RBMXL2* is overexpressed after tetracycline induction. Again, the full-length exon 9 is included, which means that *RBMXL2* is substituting for *RBMX* by suppressing the exon formation. The conclusion is that *RBMXL2* successfully replaced the function of *RBMX*.

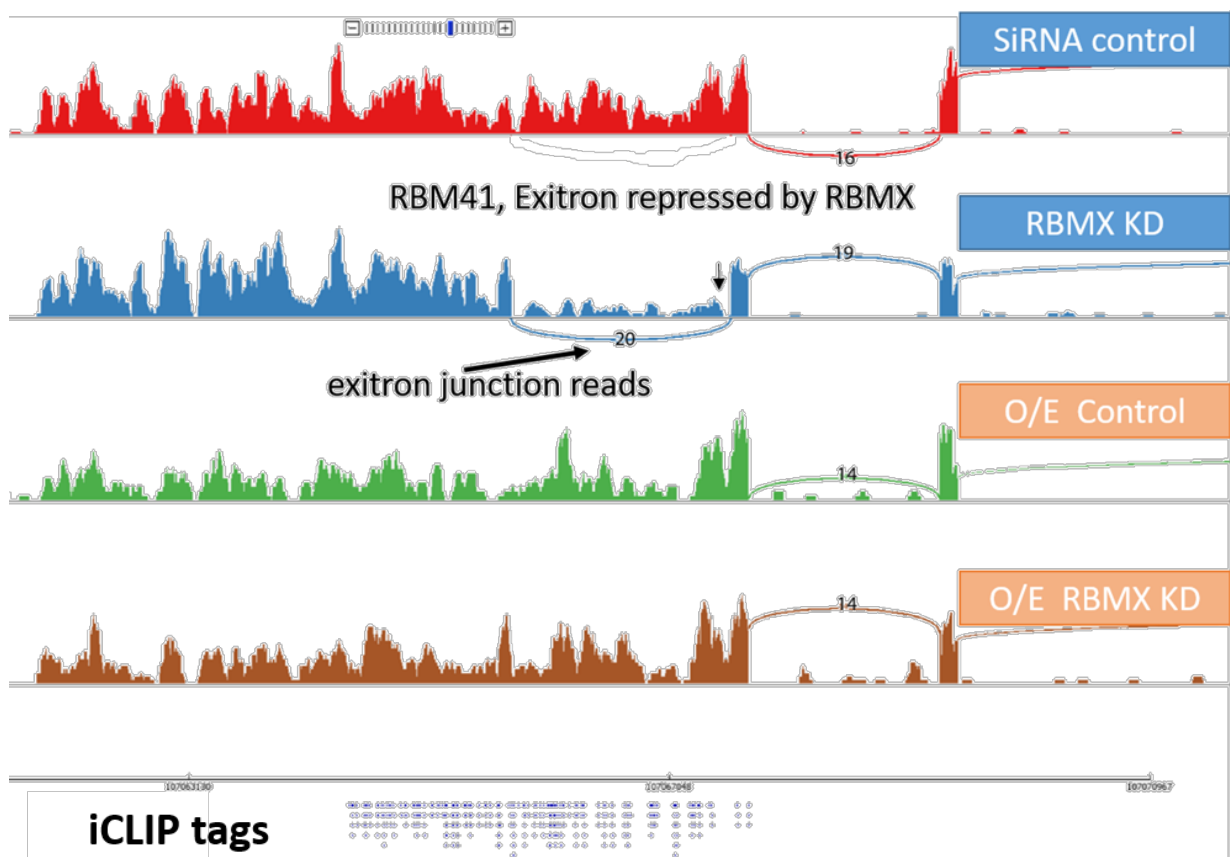


Figure 5. 9: RNAseq data used to analyse the effects of the rescue experiment on RBM41 gene. RNAseq data was visualised using a Sashimi plot. Picture taken as a browser snapshot from the Integrative Genomics Viewer (Robinson et al., 2011). A snapshot of the IGV genome browser shows that RBMX represses exitron formation within exon 7 of the *RBM41* gene. The full exon 7 was expressed normally in the control siRNA condition. However, when *RBMX* is depleted by siRNA, the *RBMX* KD condition, the formation of exitron is seen, leading to a defective, shorter mRNA that is not seen if *RBMX* is expressed. The third condition is siRNA with overexpression of RBMXL2 (O/E control), which shows the full-length exon 7 as it is usually expressed. Finally, the last condition (O/E *RBMX* KD) is where *RBMX* is knocked down using siRNA, and RBMXL2 is overexpressed after tetracycline induction. Again, the full-length exon 7 is included, which means that *RBMXL2* is substituting for *RBMX* by suppressing the exitron formation. The conclusion is that RBMXL2 successfully replaced the function of RBMX.

5.4.4 Experiment to test if RBMY is able to replace RBMX function in somatic cells:

From the previous experiment, I demonstrated that RBMXL2 is able to provide a like-to-like replacement for RBMX. Therefore, I wanted to replicate this same kind of experiment using the more distantly related RBMY protein. RBMY has been reported as a splicing regulator (Venables et al., 2000; Elliott, 2000). However, the mechanism of this regulation and whether RBMY is similar to RBMX still needs to be better understood.

Therefore, I used a stable cell line (Flp-In HEK293) from our lab that can overexpress RBMY-FLAG. Therefore, adding tetracycline will allow the overexpression of *RBMY* in those cells. The protocol closely follows the one outlined in Chapter 2, Section 2.11. Afterward, cells were harvested for RNA extraction (to be used in RT-PCR) and protein extraction, as demonstrated in Figure 5.10 below.

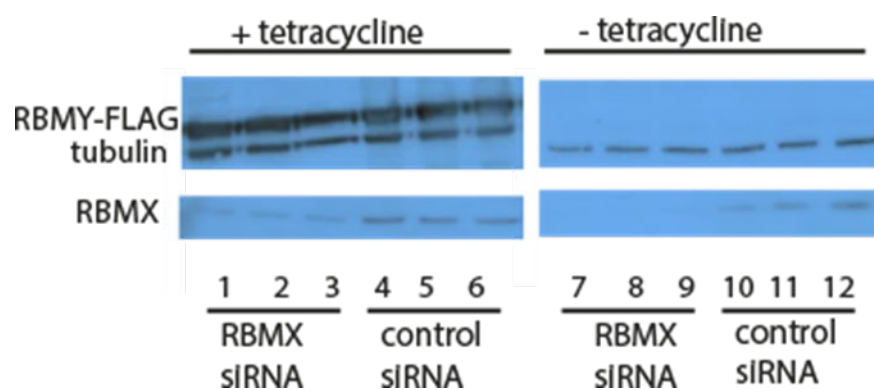


Figure 5. 10: Western analysis of RBMY rescue experiment. The western blot shows the successful knock down of *RBMY* against the siRNA control and tubulin as a control. (even across the membrane). Finally, the signal detected using the flag antibody indicates that the tetracycline induced HEK293 cells express *RBMY*.

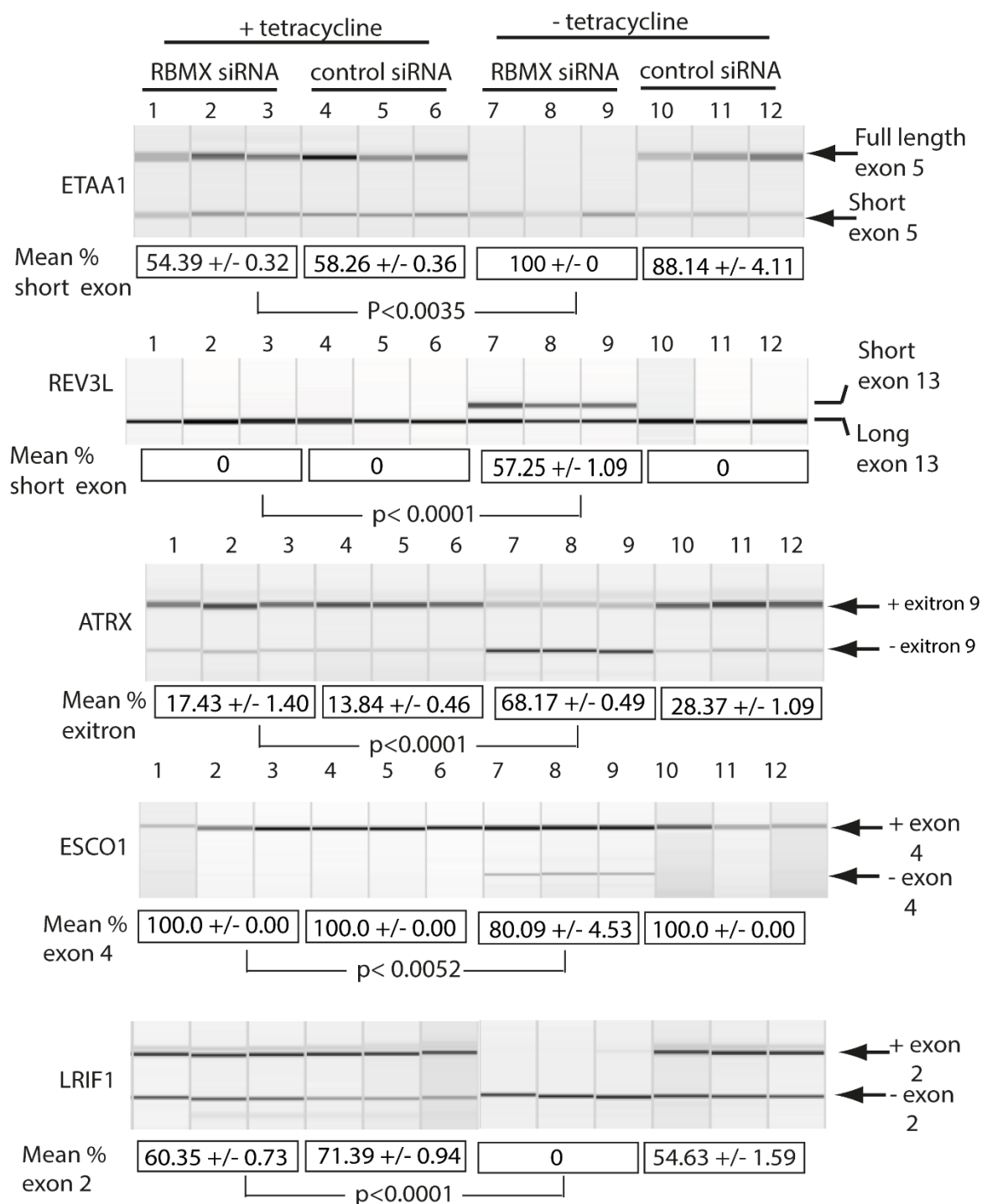


Figure 5. 11: Capillary gel electrophoretograms show RBMY's ability to substitute for RBMX in somatic cells and ensure the proper splicing inclusion of ultra-long exons on different genes (Siachisumo et al., 2023). Each experimental condition was run in triplicate, comparing the *RBMX* siRNA to the siRNA control, and when tetracycline was added to induce RBMY, to check if RBMY is preventing the use of the cryptic splice sites, thus replacing RBMX function. Boxes below each triplicate gel electrophoresis show the mean of PSI values. The P value was calculated for each gene to show the significance between *RBMX* knockdown and *RBMX* knockdown with RBMY overexpression. These experiments were performed by me and are part of the Siachisumo et al 2023 manuscript of which I am joint first author.

I followed a similar kind of “rescue” experiment to those described above for *RBMXL2*. This experiment was analysed using RT-PCR for the same ultra-long exon targets of *RBMX*: *ETAA1*, *REV3L*, *ATRX*, *ESCO1*, and *LIRF1*. Rather than doing RNAseq, I only analysed the experiment using a one-step RT-PCR (Qiagen) and capillary gel electrophoresis figure 5.11.

Our previous data show that the depletion of *RBMX* caused an increased selection of the weak cryptic 3' splice site within exon 5. Here, I tested the effect of *RBMX* depletion on *ETAA-1* cryptic splicing using RT-PCR analysis of RNA from an inducible cell line that overexpressed *RBMX* if Tetracycline was added. I depleted *RBMX* using siRNA, or used an siRNA control, in three biological replicates. In (lanes 7-9), where *RBMX* was depleted, the inclusion of the shorter isoform of exon 5 was only detected, while (lanes 10-12) that were treated with control siRNA showed expression of full length of exon 5. On the experimental plate, cells were induced with tetracycline to overexpress *RBMX*. Similar to (lanes 10-12), which had the control siRNA transfection, (lanes 4-6) showed expression of full length of exon 5 and a small amount of the shorter isoform. Finally, (lanes 1-3), where *RBMX* was depleted and *RBMX* was overexpressed, showed more inclusion of the full-length *ETAA1* exon 5 compared to when *RBMX* was depleted without *RBMX* overexpression (compare lanes (1-3) with lanes (7-9)). The PSI was calculated, and Graphpad prism was used to perform t test to calculate the P value and to compare splicing patterns after the *RBMX* knockdown from the control plate (without overexpression) with the *RBMX* knockdown and *RBMX* overexpression. The result was significant, showing that *RBMX* successfully replaced the function of *RBMX*.

The second example I tested is the *REV3L* gene, which is also implicated in cell cycle regulation and genome stability. The 3' cryptic splice site of exon 13 was used after *RBMX* depletion. To test cryptic splicing of this gene using the stable cell line, I started by analysing the RT-PCR of the control plate. On (lanes 10-12) treated with siRNA control, the long version of exon 13 was the only one detected. However, in (lanes 7-9), where *RBMX* was knocked down using siRNA, the long isoform and the short isoform of exon 13 were seen, which indicates the increased use of the 3' cryptic splice site. However, on the experimental plate where *RBMX* was overexpressed after the induction of tetracycline, (lanes 4-6) showed production of the long exon 13 isoform only. Finally, (lanes 1-3) were similar to the control siRNA lanes, showing the long exon 13 only. The PSI values were calculated, and Graphpad prism was used to calculate

the P value in order to compare the *RBMX* Knockdown in the control plate (without overexpression) with the *RBMX* Knockdown and *RBMX* overexpression. The result was significant, showing that *RBMX* successfully replaced the function of *RBMX*.

The third example is *ATRX*. Our data showed that when *RBMX* is depleted, the ultra-long exon 9 of the *ATRX* gene has an exon that is usually not seen when *RBMX* is present. I used the same approach to test if *RBMX* can restore the function of *RBMX* and prevent the formation of this exon. Two amplification products were seen on (lanes 10-12) from cells treated with the siRNA control. There was strong expression of long exon 9 and a slight expression of exon 9 after exon removal. However, the levels of these two products changed on (lanes 7-9) where *RBMX* was knocked down using siRNA, indicating increased exon use. On the experimental plate where *RBMX* was overexpressed after tetracycline induction, the expression of the long exon 9 was strong, and a faint band representing the exon was also detected. Finally, in (lanes 1-3), where *RBMX* was depleted and *RBMX* was overexpressed, again, two bands were seen. The expression of the long exon 9 was strong, and a faint band representing the exon was also seen. The PSI was calculated, and Graphpad prism was used to get the t test and P value in order to compare the *RBMX* knockdown in the control plate (without overexpression) with the *RBMX* Knockdown and *RBMX* overexpression. The result was significant, showing that *RBMX* successfully replaced the function of *RBMX*.

RBMX activates the splicing of exon 4 in the *ESCO1* gene. Therefore, the depletion of *RBMX* leads to skipping the ultra-long exon 4 (around 2.4KB in length). To test this gene using the stable cell line, I used RT-PCR. On (lanes 10-12) with siRNA control, RT-PCR showed that the ultra-long exon 4 was included. However, two bands were detected in (lanes 7-9), where *RBMX* was depleted using siRNA. On the experimental plate, *RBMX* was overexpressed after the addition of tetracycline. (Lanes 4-6) showed the inclusion of exon 4 after treatment with the control siRNA. Amplification products in (lanes 1-3), from cells depleted for *RBMX* were similar to (lanes 4-6) treated with control siRNA, both showing inclusion of exon 4. The PSI was calculated, and Graphpad prism was used to get the t test and P value in order to compare the *RBMX* Knockdown in the control plate (without overexpression) with the *RBMX* Knockdown and *RBMX* overexpression. The result was significant, showing *RBMX* was successfully substituted for *RBMX* in activation of *ESCO1* exon 4.

The last gene tested in these experiments was *LRIF1*, for which RBMX activates exon 2. On (lanes 10-12) from cells treated with siRNA control, two bands were detected: the inclusion and the skipping of exon 2. When *RBMY* was depleted with siRNA, only the skipping of exon 2 was detected in (lanes 7-9). On the experimental plate, *RBMY* was overexpressed after the induction of tetracycline. (Lanes 4-6) that were treated with the siRNA control and induced for overexpression of *RBMY* showed strong inclusion of exon 2, as well as another band that showed the skipping of the same exon. While (lanes 1-3) that were depleted for *RBMX* and also had overexpression of *RBMY* induced by tetracycline showed more inclusion of exon 2 similar to the siRNA control (lanes 4-6). The PSI was calculated, and Graphpad prism was used to get the P value in order to compare the *RBMX* depletion in the control plate (without overexpression) with the *RBMX* depletion and *RBMY* overexpression. The result was significant, indicating that *RBMY* replaced *RBMX*'s function.

5.4.5 Investigating the importance of the RBMXL2 (RRM) in RNA binding:

After proving that RBMXL2 and RBMY are able to replace RBMX function in somatic cells, I wanted to study the importance of the RRM and the disordered in RBMXL2 protein. I hypothesised that removing the RRM will affect the RNA-protein interactions. Thus if these RRM-RNA interactions are important for splicing control, then the RBMXL2 Δ RRM protein will not replace the function of RBMX efficiently as the full length RBMXL2. Note that the C-terminal disordered region can also interact with RNA (Ottoz & Berchowitz, 2020), so if the C-terminal disordered region is important for splicing regulation, then the RBMXL2 Δ RRM protein will still rescue.

The rescue same experiment was repeated, using RT-PCR for some ultra-long exon targets, such as *ETAA1*, *REV3L*, *ATRX*, *LIRF1* and *ESCO1*. One-step RT-PCR (Qiagen) was used, and samples were run on the Qiaxcel machine (Qiagen) following the manufacturer's protocols, as seen in the figure below 5.13. PSI and P values were calculated using Prism GraphPad software. Finally, a Western blot was used to confirm the overexpression of the RBMXL2 Δ RRM-FLAG and the knockdown of *RBMX*.

I first used a stable cell line from our lab that Mrs Caroline Dalglish constructed to establish this protocol. This was a stable cell line that overexpressed the RBMXL2 without the RRM (RBMXL2 Δ RRM-FLAG, as seen in the figure below 5.12).

Like the previous rescue experiment, the protocol starts with plating two 6-well plates—one as a control (without overexpression) and another for overexpression. The main difference between the two plates is the addition of tetracycline. Following the same protocol in chapter 2nd cells were harvested for RNA extraction (to be used for RT-PCR) and protein extraction (for a western blot). Lastly, I used Western blot to confirm the overexpression of RBMXL2 Δ RRM and the depletion of *RBMX* against a housekeeping protein (tubulin), as seen in the figure below 5.14.

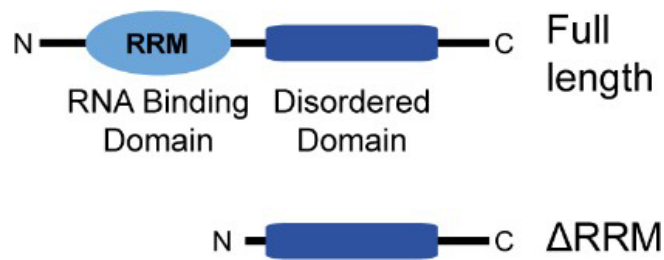


Figure 5. 12: Comparison of the full-length RBMXL2 protein and RBMXL2 Δ RRM that was used for the stable cell line.

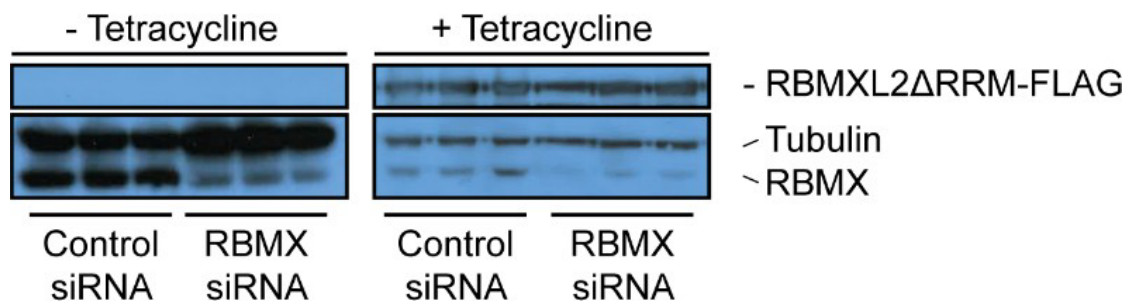


Figure 5. 13: Western analysis of the RBMXL2 Δ RRM rescue experiment. The western blot shows the successful depletion of *RBMX* compared the siRNA control and tubulin as a control (even across the membrane) confirms the knockdown of *RBMX*. Finally, the probing with the flag antibody indicates that HEK293 cells express RBMXL2 Δ RRM. However, the membrane exhibits some degree of over-exposure in the Tubulin control.

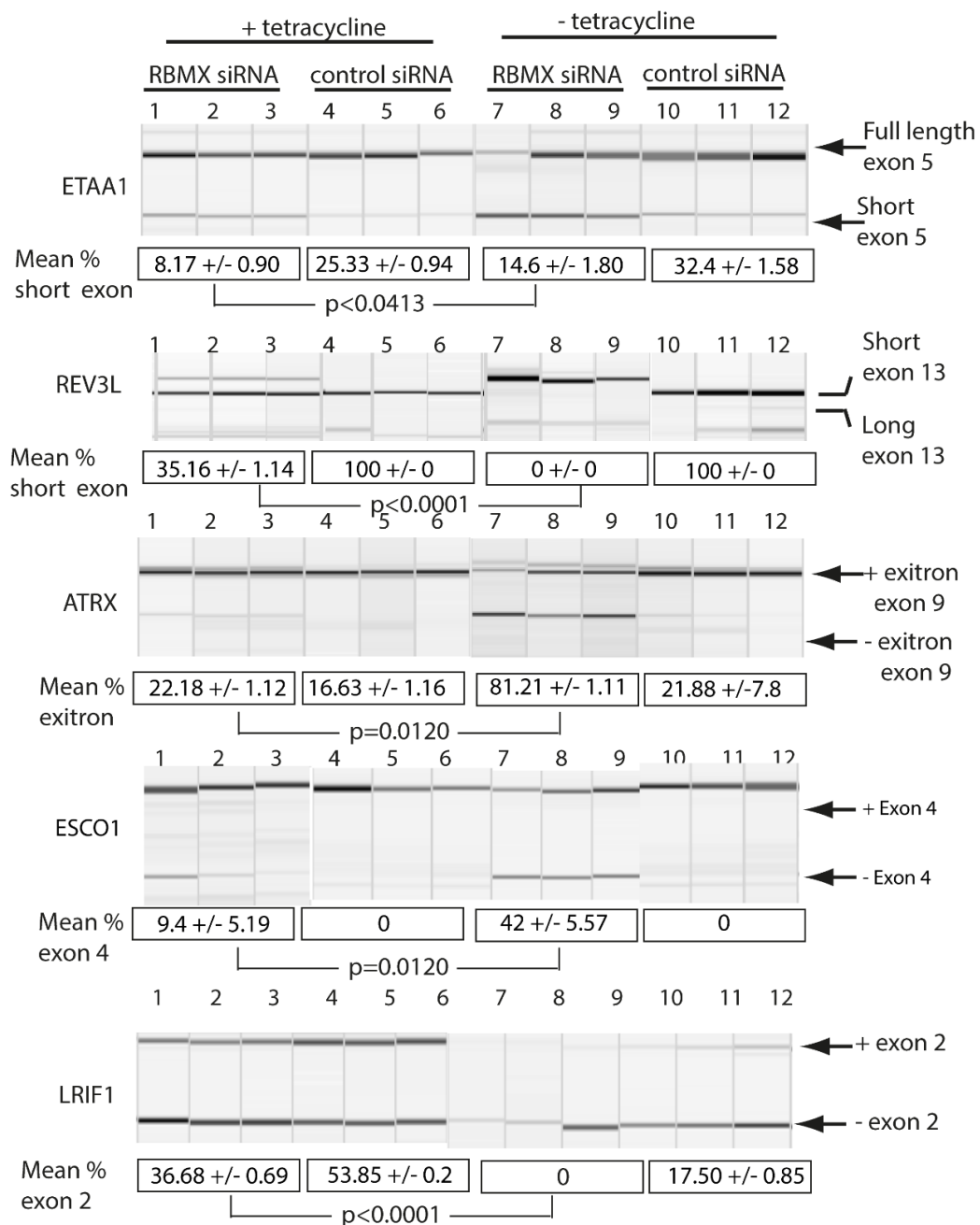


Figure 5. 14: Capillary gel electrophoretograms show RBMXL2ΔRRM's ability to substitute for RBMX in somatic cells and ensure the proper splicing inclusion of ultra-long exons on different genes (Siachisumo et al., 2023). Samples from each gene were run in triplicate, comparing the *RBMX* siRNA to the siRNA control, and when tetracycline was added to induce RBMXL2ΔRRM, to check if RBMXL2ΔRRM is rescuing the use of the cryptic splice site, thus replacing *RBMX* function. Boxes below each triplicate gel electrophoresis show the mean after calculating the PSI percentage. The P value was calculated using a t test (graphpad prism) to show the significance between *RBMX* knockdown and *RBMX* knockdown with RBMXL2ΔRRM overexpression. These experiments were performed by me and are part of the Siachisumo et al 2023 manuscript of which I am joint first author.

I followed a similar kind of “rescue” experiment to those described above for *RBMXL2*. This experiment was analysed using RT-PCR for the same ultra-long exon targets of *RBMX*: *ETAA1*, *REV3L*, *ATRX*, *ESCO1*, and *LIRF1*. Rather than doing RNAseq, I only analysed the experiment using a one-step RT-PCR (Qiagen) and capillary gel electrophoresis figure 5.14.

I tested the effect of *RBMX* protein on *ETAA-1* expression using RT-PCR from an inducible cell line that overexpressed *RBMXL2ΔRRM* if Tetracycline was added. I did deplete *RBMX* using siRNA or treated with siRNA control in three biological replicates (without overexpression of *RBMXL2ΔRRM*). Lanes (10-12), containing control siRNA exhibited two bands representing the full length of exon 5 and the shorter isoform. However, lanes (7-9), where *RBMX* is depleted, increase of the shorter isoform was detected along with the full length of *ETAA-1* exon 5. On the other plate, that was induced with tetracycline to overexpress *RBMXL2ΔRRM*. Similar to lanes (10-12), which had the control siRNA, lanes (4-6), two bands were seen, one indicating the full length of exon 5 and another faint band represent the expression of the shorter isoform. Last lanes (1-3) where *RBMX* was knockdown using siRNA and *RBMXL2ΔRRM* was overexpressed, partial restoration of *RBMX* function were seen as more inclusion of the full length were detected. However, small amount of the short isoform was also observed. PSI was determined, and GraphPad prism was used to calculate the t test and obtain the P value to compare the *RBMX* knockdown in the control plate (without *RBMXL2RRM* overexpression) to the *RBMX* depletion with *RBMXL2RRM* overexpression. *RBMXL2RRM* partially restored the function of *RBMX*; the result was statistically significant.

The *REV3L* gene is also associated with cell cycle regulation and genome stability. Depletion of RBMX caused the use of the 3' cryptic splice sites of exon 13, produces a shorter defective mRNA that was not seen if RBMX was expressed. The previous experiment analysed the RT-PCR of biological replicates from the control plate. On lanes (10-12) treated with control siRNA, amplification of the long exon 13 was the only one detected. However, in lanes (7-9) where RBMX was depleted, only the short isoform was observed (although some of these bands did not align perfectly with each other on the capillary gel electrophoretogram), which confirmed the use of the 3' cryptic splice sites for the overexpressed experimental plate RBMXL2ΔRRM after the addition of tetracycline. Lanes (4-6) treated with control siRNA were similar to lanes (10-12), showing the amplification of the long exon 13 only. However, unlike lanes (1-3), showed both the long exon and the faint band that stands for the shorter isoform. Again, PSI was obtained, and GraphPad prism was used to calculate the t test and P value. The result was statistically significant, and RBMXL2ΔRRM replaced the function of RBMX.

The *ATRX* gene was the third example; depletion of *RBMX* causes an exitron on the ultra-long exon 9 of the *ATRX* gene that is usually not seen when RBMX is expressed. The exitron is formed due to the selection of cryptic 5' and 3' splice sites in exon 9. I tested the importance of the RRM in RBMXL2 using the same method as the last experiment. On the control plate no RBMXL2ΔRRM overexpression is introduced in the absence of tetracycline. On (lanes 10-12) treated with control siRNA, detection of the full-length exon 13 was observed along with a slight expression of the exitron. However, on (lanes 7-9), where siRNA reduced the amount of RBMX, the stronger lower band showed the increased use of the exitron. The samples from the other plate where RBMXL2ΔRRM was overexpressed after tetracycline induction gave the following results: on (lanes 4-6) where cells were treated with control siRNA, one band was observed, indicating the expression of the full length of exon 13. However, on (lanes 1-3) where *RBMX* was depleted, and RBMXL2ΔRRM was overexpressed after tetracycline induction, two bands were observed, displaying the expression of the full-length exon 13 and a faint band representing the slight expression of the exitron. The mean PSI values were determined, and Graphpad prism was used to get the P value using a t-test. The P value was significant, and indicates that RBMXL2ΔRRM substituted the function of RBMX.

The *ESCO1* gene is one of the genes containing an exon activated by RBMX. *RBMX* depletion causes a skipping of the ultra-long exon 4. To test whether the RRM is important for splicing control of the *ESCO1* gene using the stable cell line, I started by analysing the RT-PCR of the control plate. One band was amplified on (lanes 10-12) treated with siRNA control, representing splicing inclusion of the ultra-long exon 4. While on (lanes 7-9), where *RBMX* was knocked down by siRNA, two bands were observed. Corresponding to the splicing of the ultra-long exon 4 and also skipping of exon 4 that was not seen in the previous lanes. On the other plate, *RBMXL2ΔRRM* was overexpressed after tetracycline induction. (Lanes 4-6) were similar to (lanes 10-12) and showed the activation of the ultra-long exon 4. While (lanes 1-3) presented with two lanes, the activation of exon 4 and another faint band represent the slight level of exon 4 skipping. Mean PSI values were calculated for exon 4 inclusion and GraphPad prism was used to calculate the P value using a t-test. Despite differences of expression between replicates, the results were significant, showing that *RBMXL2ΔRRM* was partially able to replace the RBMX function.

The *LRIF1* gene is another example of the genes that RBMX control via exon activation. Depletion of RBMX leads to the skipping of the ultra-long exon 2. I tested if *RBMXL2ΔRRM* is able to replace the function of RBMX in the activated genes. To test splicing of this gene using the stable cell line, I started by analysing the RT-PCR of the control plate. In (lanes 7-9), where RBMX was knocked down by siRNA, the skipping of exon 2 was the only product seen. However, on (lanes 10-12) treated with siRNA control, two bands were observed: the skipping of exon 2 and the faint band representing the inclusion of exon 2. On the experimental plate, *RBMXL2ΔRRM* was overexpressed after tetracycline induction. (Lanes 4-6) were treated with siRNA control with overexpression of *RBMXL2ΔRRM*. This samples presented with two bands, the activation of exon 2 and the skipping of the same exon. Finally, (lanes 1-3) treated with RBMX siRNA and the overexpression of *RBMXL2ΔRRM*, also gave two bands, one for the inclusion of exon 2 and a shorter band showing the expression of skipping exon 2. Importantly, more exon 2 splicing was seen when *RBMXL2ΔRRM* was expressed. Exon 2 activation was seen when *RBMXL2ΔRRM* overexpressed in both conditions. The mean PSI values were determined, and GraphPad prism was used to obtain the P value by performing a t-test and compare results from the *RBMX* knockdown in the control plate (without *RBMXL2ΔRRM* overexpression) to the *RBMX* depletion and *RBMXL2ΔRRM* overexpression. These data showed that *RBMXL2ΔRRM* restored the function of RBMX; the result was statistically significant.

5.5 Discussion:

In this chapter I showed that a “rescue” experiment can be used to study the function of RBMXL2 protein much better than a minigene. This kind of rescue experiment involves the use of a cell line that expresses RBMXL2 after tetracycline induction. I then analysed RNA samples after *RBMX* depletion to test whether RBMXL2 expression is able to replace the loss of RBMX. This approach showed that RBMXL2 is able to replace RBMX. Even more surprising, RBMY is also able to replace RBMX function. I also did experiments that showed that the C-terminal disordered domain of RBMXL2 is able to restore function of RBMX as well as full length protein.

5.5.1 RBMXL2 is replacing RBMX in somatic cells:

The Elliott Lab established that *RBMX* is important for proper splicing inclusion of a group of long exons, called "ultra-long exons" that are more than 1 kb in length. Also, they found that RBMX, which is ubiquitously expressed, frequently functions as a splicing repressor. RBMX and RBMXL2 both act as splicing repressors. RBMX is known to bind to RNA from genes essential for genome stability, and the depletion of *RBMX* leads to increased DNA damage. Moreover, a study showed that *RBMX* is vital for genome stability during replication by activating the ATR pathway (Zheng et al., 2020). Professor Elliott's lab showed that a shorter version of the *ETAA1* mRNA was produced in the case of depletion of *RBMX*. However, a full length version of *ETAA1* mRNA and encoded protein is required for repairing stalled replication forks and enabling DNA repairs to sustain the genome's stability (Lee et al., 2016). A study found that *RBMX* also turns on the homologous recombination pathway to ensure *BRCA2* is expressed correctly (Adamson et al., 2012). My data demonstrated the ability of RBMXL2 to replace the function of RBMX in somatic cells despite their divergence 65 million years ago.

5.5.2 RBMY another member of the family is also replacing the function of RBMX:

RBMY is located on chromosome Y and has a testis-specific expression. Deletion of the *RBMY* gene leads to azoospermia. A previous study showed that *RBMXL2* and *RBMX* work as co-regulators for the splicing inclusion of *TLE4-T* (Liu et al., 2009) after confirming that *RBMXL2* is able to replace the function of *RBMX* in somatic cells. *RBMY* is less like *RBMX* than *RBMXL2*, but still rescue the function of *RBMX*.

5.5.3 The disordered domain of *RBMXL2* is responsible for RNA-protein binding:

RBMY and *RBMX* share 88% similarity in the RRM sequence. In contrast, the RRM in *RBMXL2* shared 84% similarity with *RBMY* (Skrisovska et al., 2007). In this chapter, I tested the importance of the RRM in *RBMXL2*. I hypothesized that RRM of *RBMXL2* directly binds to RNA and so would likely be important for splicing control. Therefore, Mrs Caroline Dalglish constructed a tetracycline- inducible cell line to test this theory that would overexpress the C-terminus disordered domain without the RRM. Then I tested the same genes that I analysed for the full-length *RBMXL2*. Remarkably, I found that the RRM is not essential in rescuing cryptic splicing pattern. Thus, the disordered domain in the *RBMXL2* is the key region for rescuing cryptic splicing. In the literature, Liu et al. (2017) showed that the disordered domain on *RBMX* is responsible for RNA binding as well as the RRM. My data showed that *RBMXL2* Δ RRM successfully restored the function of *RBMX*. The RRM is the most conserved region but seemingly not important for rescue. For example, a direct comparison of wild-type (WT) and *RBMX* knockdown on splicing of *LIRF1* was carried out using *RBMXL2* overexpression without the RRM domain. This showed increased activation in response to tetracycline and the control siRNA, as compared to tetracycline plus the siRNA specific to *RBMX*. This shows that endogenous *RBMX* can activate this exon as well as *RBMXL2*.

Over-expressing other regions of *RBMXL2* would be interesting as these would define the important sequences for splicing function. This comparison also aids in determining whether the RRM domain can compensate for the absence or dysfunction of other protein regions, which could be significant.

5.6 Chapter summary:

The RBM family of RNA-binding proteins consists of RBMX, RBMXL2, and RBMY. Here I found that despite the divergence between *RBMX* and *RBMY* around 200 million years ago and between *RBMX* and *RBMXL2* around 65 million years ago, both RBMY and RBMXL2 substitute for the function of RBMX in somatic cells. In addition, RBMXL2 is expected to directly replace the role of RBMX when XY is inactivated during meiosis. I also demonstrated the significance of *RBMXL2*'s disordered c-terminus domain for binding.

Chapter 6 : Discussion and future work:

Gene expression involves different RNA processing steps, including capping, polyadenylation, and Pre-mRNA splicing. Pre-mRNA splicing is an integral part of controlling gene expression. It involves taking out introns and joining exons in different ways to make proteins with diverse functions. This process, known as “alternative splicing,” is coordinated by various RNA-binding proteins. The testis-specific *RBMXL2* was identified as a splicing repressor, and the same study from Elliott’s lab revealed that RBMXL2 is repressing the splicing of weak splice sites, known as “cryptic splicing events.”

It is fundamental to understand the mechanisms of gene expression and protein synthesis, especially in the testis, where numerous splicing events occur. Therefore, by knowing the RNA targets of these proteins, we can gain insights into their role in splicing and gene expression. In this thesis I described experiments to identify the role of the RNA-binding protein RBMXL2.

6.1 *Rbmxl2* knockout has a severe phenotype on the Sv/129 background.

The first aim of my PhD work was to determine whether the phenotype caused by the *Rbmxl2* knockout gene was strain-dependent. During my PhD, I managed to compare the effect of deleting the *Rbmxl2* gene within two mouse strains (C57BL/6 and Sv/129). Lack of *Rbmxl2* in the C57BL/6 strain causes infertility and arrest at diplotene, according to a previous study from the Elliott lab (Ehrmann et al., 2019).

First, I confirmed Dr Ingrid Ehrmann's findings on the C57BL/6 mouse strain. The phenotype caused by *Rbmxl2* knockout was then investigated in the Sv/129 mouse strain. I discovered that mice lacking *Rbmxl2* were infertile, and their development was arrested at the pachytene stage. This demonstrates that deletion of *Rbmxl2* causes an earlier meiotic defect on the Sv/129 mouse background. Apoptosis was also observed at the pachytene stage, with many seminiferous tubules devoid of germ cells. These results demonstrate that RBMXL2 protein is essential during meiosis and not only during diplotene, as we had previously thought based on the original C57BL/6 mouse model. Despite the differences in phenotypes, both mouse strains lack sperm and are infertile, indicating the importance of *Rbmxl2* in meiosis and germ cell development.

These differences in phenotype were not caused by differences in RBMXL2 protein expression patterns. In both strains, RBMXL2 protein is still only expressed in the testis during and immediately after meiosis (from pachytene to the round spermatid stage). Histological staining demonstrated that deleting *Rbmxl2* resulted in strain-specific phenotypes in both backgrounds. These phenotypes indicate the presence of modifier genes that might lead to differences in germ cell development dependence on RBMXL2 between C57BL/6 and Sv/129 mice (Ramsbottom et al., 2020).

In my research, I chose not to employ a random sampling approach to investigate patterns of Hematoxylin and Eosin (H&E) staining because my main goal was to visually identify and distinguish different cell types in mouse testicular tubules for staging purposes. Since my study focused on qualitative observations, it was essential for me to be precise in choosing tubules that contained a variety of cell types. Unfortunately, the SV129 mouse strain I was studying had numerous tubules with only Sertoli cells, making the selection process more challenging. This unique characteristic of the mouse strain supported my decision to opt for a non-random staining approach, ensuring the accuracy and relevance of my qualitative investigation.

If I had more time, I would use apoptotic markers like caspase 3 or TUNEL antibodies to stain both mouse strains, to determine if cell death in both cases was associated with apoptosis.

6.2 Using iCLIP to identify direct targets of RBMXL2:

The second Aim of my thesis was to identify direct targets of RBMXL2 in mouse testis. To answer this, I used iCLIP to conduct a global transcriptome analysis of RBMXL2-RNA binding targets in mouse testis. A modified iCLIP protocol (unpublished) was utilised to identify RBMXL2 binding sites.

My results show that RBMXL2 protein RNA-binding was high in intergenic and intronic regions. Furthermore, the RBMXL2 iCLIP revealed an enrichment of binding sites in ultra- long exons. RBMXL2 was previously recognised as a splicing suppressor because it inhibited the cryptic splice sites of long exons (Ehrmann et al., 2019; Aldalaqan et al., 2022).

Furthermore, my iCLIP results shown an enrichment of binding sites in meiosis-essential genes like *Meioc* and *Esco1*, defects in splicing of which could contribute to the male infertility that is caused by *Rbmxl2* deletion. In addition, gene ontology analysis reveals that RBMXL2 protein binding is enriched in transcripts from genes involved in spermatogenesis, particularly meiosis and sperm formation and movement genes. My iCLIP results showed a

high density of RBMXL2 protein binding within ultra-long exons, which is consistent with a model that the RBMXL2 protein binds to RNAs to mask sequences needed for cryptic splice site selection. I successfully analysed the global transcriptome of RBMXL2-RNA binding targets in mouse testis. Furthermore, the minigene experiment did not replicate what I observed in the iCLIP and RNA-seq results. This could be due to GFP-fusion proteins that have an added GFP tag that could affect protein function. GFP does have several advantages in research, including low cellular toxicity, pH stability, and non-invasive monitoring of specific promoter activity. GFP is an effective environmental toxicity assessment reporter, helps with protein localization, tracks DNA and protein distribution, and aids in protein localization. It does, however, have some disadvantages. Despite these advantages, when GFP is fused to a target protein, it can significantly increase its size and molecular mass, potentially altering its natural function and intracellular transport. Care must be taken when employing GFP-fusion constructs. This could explain the ineffectiveness of the minigene experiment involving RBMXL2 (Remington, 2011; Ansari et al., 2016).

Following the differences between minigene and stable cell line experiments, we speculate that RBMX requires chromatin to function as a cryptic splice suppressor. Therefore, utilizing an alternative minigene experiment becomes of interest. In this approach, after transfecting cells, we would extend the incubation period beyond 24 hours, allowing for chromatin structure remodelling. This strategy draws inspiration from the work of Ast et al. (2014), where they tested a minigene with reduced nucleosome occupancy, leading to delayed RNA polymerase II progression, coupled with the addition of Histone deacetylase (HDAC) inhibitors. Furthermore, investigating RBMXL2 without its C-terminus in iCLIP analysis is an interesting avenue to explore its RNA-binding behaviour. Furthermore, performing a rescue experiment with a stable cell line expressing RBMXL2 minus the C-terminus allows for data comparison with RBMXL2 Δ RRM overexpression, assisting in the identification of critical functional domains. Another possible experiment could be a gel shift assay, in which purified RBMXL2 protein with and without the RRM domain are incubated with RNA, followed by gel electrophoresis to observe differences in RNA-binding based on gel shift.

If I have more time I would do detailed study of the RNA binding of processing control. I would learn more bioinformatics so I can correlate binding patterns with RNA processing. Alternately, I will use RNA-seq data from purified cell types (recently generated by Dr.

Ehrmann) also to correlate with sites of RNA-protein interaction that I have already mapped to determine the rules by which the RBMXL2 protein regulates RNA processing. Furthermore, preliminary RBMXL2 iCLIP experiments were conducted without the RRM domain, and binding was observed. Replicating this experiment and comparing the results with my RBMXL2 iCLIP data would be of interest.

6.3 RBMXL2 can substitute for RBMX and RBMY in cell lines to ensure accurate splicing of ultra-long exons.

Previously, Professor Elliott's lab demonstrated that the testis-specific protein RBMXL2 inhibits splicing of cryptic splice sites within some ultra-long exons during meiosis. Dr. Chile Siachisumo (a former PhD student in Elliott's lab) revealed that the paralog RBMX protein also suppresses some cryptic splice sites within ultra-long exons in somatic cells. To test if these proteins are functionally interchangeable, I replaced RBMX with RBMXL2 and RBMY in stable cell lines and demonstrated that RBMXL2 and RBMY directly replaced RBMX functions in somatic cells. RBMXL2 suppresses cryptic splice sites for important genes in genome stability, such as *Brca2* (Ehrmann et al., 2019). Similarly to RBMX, which has been found to regulate important genes in genome stability, such as *REV3L*, *ATRX*, and *ETAA1* (Chile Siachisumo, PhD thesis, 2022).

Consequently, RBMXL2 replaces RBMX during meiosis when the X chromosome is inactive. This fits into a broader picture, where numerous X-linked genes have testis-specific retrogenes that are only expressed during meiosis when both X and Y are inactive. RBMX has thus maintained its function as a splicing regulator for at least 200 million years. Finally, I demonstrated that the disordered domain at the RBMXL2 protein's C-terminus is required for cryptic splicing repression. This might be clinically important. A frameshift in the disordered C-terminus domain of human RBMX led to intellectual disability syndrome. RBMX was found to be more abundant in the isolate with the m6A-methylated hairpin. These findings identify RBMX as a protein that recognizes m6A modifications and has a specific relationship with an m6A-modified hairpin structure observed in MALAT1, a long non-coding RNA. In addition, a study demonstrated that the C-terminal disordered region of RBMX promotes exon inclusion by recognising and binding to m6A (N6-methyladenosine) (Liu et al., 2017; Shashi et al., 2015).

6.4 Conclusion:

In this thesis, I demonstrate that RBMXL2 regulates the mRNA processing of meiosis- and fertility-related genes. I also showed that RBMXL2, like RBMX, inhibits the cryptic splice sites in ultra-long exons. Furthermore, I have shown that RBMXL2 directly replaces the function of RBMX during meiosis when the X and Y are inactivated. In addition, I have demonstrated that the RBMXL2 knockout phenotype is more severe in the Sv/129 mouse strain than in the previously reported 57BL/6 background.

Bibliography:

- Abby, E., Tourpin, S., Ribeiro, J., Daniel, K., Messiaen, S., Moison, D., Guerquin, J., Gaillard, J.C., Armengaud, J., Langa, F., Toth, A., Martini, E. & Livera, G. (2016) 'Implementation of meiosis prophase I programme requires a conserved retinoid-independent stabilizer of meiotic transcripts', *Nature Communications*, 7.
- 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.
- Adham, I.M., Nayernia, K., Burkhardt-Göttges, E., Topaloglu, Ö., Dixkens, C., Holstein, A.F. & Engel, W. (2001) 'Teratozoospermia in mice lacking the transition protein 2 (Tnp2)', *Molecular Human Reproduction*, 7(6), pp. 513–520.
- Aldalqaan, S., Dalglish, C., Luzzi, S., Siachisumo, C., Reynard, L.N., Ehrmann, I. & Elliott, D.J. (2022) 'Cryptic splicing: common pathological mechanisms involved in male infertility and neuronal diseases', *Cell Cycle*, 21(3), pp. 219–227.
- Alomer, R.M., Da Silva, E.M.L., Chen, J., Piekarz, K.M., McDonald, K., Sansam, C.G., Sansam, C.L. & Rankin, S. (2017) 'Esco1 and Esco2 regulate distinct cohesin functions during cell cycle progression', *Proceedings of the National Academy of Sciences of the United States of America*, 114(37), pp. 9906–9911.
- Amann, R.P. & Howards, S.S. (1980) 'Daily spermatozoal production and epididymal spermatozoal reserves of the human male', *Journal of Urology*, 124(2), pp. 211–215.
- Anna, A. & Monika, G. (2018) 'Splicing mutations in human genetic disorders: examples, detection, and confirmation'. *Journal of Applied Genetics* 59 (3) p.pp. 253–268.
- Ansari, A.M., Ahmed, A.K., Matsangos, A.E., Lay, F., Born, L.J., Marti, G., Harmon, J.W. & Sun, Z. (2016) 'Cellular GFP Toxicity and Immunogenicity: Potential Confounders in in Vivo Cell Tracking Experiments', *Stem Cell Reviews*, 12(5), p. 553.
- Arsov, T., Silva, D.G., O'Bryan, M.K., Sainsbury, A., Lee, N.J., Kennedy, C., Manji, S.S.M., Nelms, K., Liu, C., Vinuesa, C.G., de Kretser, D.M., Goodnow, C.C. & Petrovsky, N. (2006) 'Fat Aussie—A New Alström Syndrome Mouse Showing a Critical Role for ALMS1 in Obesity, Diabetes, and Spermatogenesis', *Molecular Endocrinology*, 20(7), pp. 1610–1622.

- 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.
- Beeram, E., Suman, B. & Divya, B. (2019) 'Proteins as the molecular markers of male fertility', *Journal of Human Reproductive Sciences*, 12(1), pp. 19–23.
- Bellve, A.R., Cavicchia, J.C., Millette, C.F., O'Brien, D.A., Bhatnagar, Y.M. & Dym, M. (1977) 'Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization', *The Journal of cell biology*, 74(1), pp. 68–85.
- Best, A., Dalglish, C., Kheirollahi-Kouhestani, M., Danilenko, M., Ehrmann, I., Tyson-Capper, A. & Elliott, D. (2014) 'Tra2 protein biology and mechanisms of splicing control', *Biochemical Society transactions*, 42(4), pp. 1152–1158.
- Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R. & Blencowe, B.J. (2013) 'Dynamic Integration of Splicing within Gene Regulatory Pathways', *Cell*, 152(6), pp. 1252–1269.
- Campopiano, R., Ryskalin, L., Giardina, E., Zampatti, S., Busceti, C.L., Biagioni, F., Ferese, R., Storto, M., Gambardella, S. & Fornai, F. (2017) 'Next Generation sequencing and ALS: Known genes, different phenotypes', *Archives Italiennes de Biologie*, 155(4), pp. 159–166.
- Celeste, A., Petersen, S., Romanienko, P.J., Fernandez-Capetillo, O., Chen, H.T., Sedelnikova, O.A., Reina-San-Martin, B., Coppola, V., Meffre, E., Difilippantonio, M.J., Redon, C., Pilch, D.R., Orlu, A., Eckhaus, M., Camerini-Otero, R.D., Tessarollo, L., Livak, F., Manova, K., Bonner, W.M., et al. (2002) 'Genomic instability in mice lacking histone H2AX', *Science*, 296(5569), pp. 922–927.
- Chen, D., Zheng, W., Lin, A., Uyhazi, K., Zhao, H. & Lin, H. (2012) 'Pumilio 1 suppresses multiple activators of p53 to safeguard spermatogenesis', *Current Biology*, 22(5), pp. 420–425.
- Chi, M.N., Auriol, J., Jégou, B., Kontoyiannis, D.L., Turner, J.M.A., De Rooij, D.G. & Morello, D. (2011) 'The RNA-binding protein ELAVL1/HuR is essential for mouse spermatogenesis, acting both at meiotic and postmeiotic stages', *Molecular Biology of the Cell*, 22(16), pp. 2875–2885.
- Siachisumo, C., Luzzi, S., Aldalagan, S., Hysenaj, G., Dalglish, C., Cheung, K., Gazzara, M.R., Yonchev, I.D., James, K., Chadegani, M.K., Ehrmann, I., Smith, G.R., Cockell, S.J., Munkley, J., Wilson, S.A., Barash, Y. & Elliott, D.J. (2023) 'An anciently diverged family of RNA binding proteins maintain correct splicing of ultra-long exons through cryptic splice site repression', *bioRxiv*, p. 2023.08.15.553384.

- Chow, L.T., Gelinias, R.E., Broker, T.R. & Roberts, R.J. (1977) 'An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA', *Cell*, 12(1), pp. 1–8.
- Colgan, D.F. & Manley, J.L. (1997) 'Mechanism and regulation of mRNA polyadenylation', *Genes & development*, 11(21), pp. 2755–2766.
- Collier, B., Gorgoni, B., Loveridge, C., Cooke, H.J. & Gray, N.K. (2005) 'The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells', *EMBO Journal*, 24(14), pp. 2656–2666.
- De Conti, L., Baralle, M. & Buratti, E. (2017) 'Neurodegeneration and RNA-binding proteins', *Wiley interdisciplinary reviews. RNA*, 8(2), .
- Dass, B., Tardif, S., Ji, Y.P., Tian, B., Weitlauf, H.M., Hess, R.A., Carnes, K., Griswold, M.D., Small, C.L. & MacDonald, C.C. (2007) 'Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility', *Proceedings of the National Academy of Sciences of the United States of America*, 104(51), pp. 20374–20379.
- Delbridge, M.L., Lingenfelter, P.A., Disteché, C.M. & Graves, J.A.M. (1999) 'The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome', *Nature Genetics*, 22(3), pp. 223–224.
- Edelmann, W., Cohen, P.E., Kane, M., Lau, K., Morrow, B., Bennett, S., Umar, A., Kunkel, T., Cattoretti, G., Chaganti, R., Pollard, J.W., Kolodner, R.D. & Kucherlapati, R. (1996) 'Meiotic pachytene arrest in MLH1-deficient mice', *Cell*, 85(7), pp. 1125–1134.
- 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., Liu, Y., Danilenko, M., Crosier, M., Overman, L., Arthur, H.M., Lindsay, S., Clowry, G.J., Venables, J.P., Fort, P. & Elliott, D.J. (2013) 'The Tissue-Specific RNA Binding Protein T-STAR Controls Regional Splicing Patterns of Neurexin Pre-mRNAs in the Brain', *PLoS Genetics*, 9(4).
- 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), .

- Elliott, D.J. (2000) 'RBMV genes and AZFb deletions'. *Journal of Endocrinological Investigation* 23 (10) p.pp. 652–658.
- Elliott, D.J. (2004) 'The role of potential splicing factors including RBMY, RBMX, hnRNP-G-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) 'RBMY 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., Millar, M.R., Oghene, K., Ross, A., Kiesewetter, F., Pryor, J., McIntyre, M., Hargreave, T.B., Saunders, P.T.K., Vogt, P.H., Chandley, A.C. & Cooke, H. (1997) 'Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm', *Proceedings of the National Academy of Sciences of the United States of America*, 94(8), pp. 3848–3853.
- 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), pp. 2117–2124.
- Fernandes, S., Huellen, K., Goncalves, J., Dukal, H., Zeisler, J., Rajpert De Meyts, E., Skakkebaek, N.E., Habermann, B., Krause, W., Sousa, M., Barros, A. & Vogt, P.H. (2002) 'High frequency of DAZ1/DAZ2 gene deletions in patients with severe oligozoospermia', *Molecular Human Reproduction*, 8(3), pp. 286–298.
- Fu, X.-F., Cheng, S.-F., Wang, L.-Q., Yin, S., De Felici, M. & Shen, W. (2015) 'DAZ Family Proteins, Key Players for Germ Cell Development.', *International journal of biological sciences*, 11(10), pp. 1226–1235.
- Furuichi, Y. & Shatkin, A.J. (2000) 'Viral and cellular mRNA capping: past and prospects', *Advances in virus research*, 55pp. 135–184.
- Gamble, J., Chick, J., Seltzer, K., Graber, J.H., Gygi, S., Braun, R.E. & Snyder, E.M. (2020) 'An expanded mouse testis transcriptome and mass spectrometry defines novel proteins', *Reproduction*, 159(1), pp. 15–26.
- Garcia-Moreno, M., Järvelin, A.I. & Castello, A. (2018) 'Unconventional RNA-binding proteins step into the virus–host battlefield', *Wiley Interdisciplinary Reviews: RNA*, 9(6), .
- Gashti, N.G., Gilani, M.A.S. & Abbasi, M. (2021) 'Sertoli cell-only syndrome: etiology and clinical management', *Journal of Assisted Reproduction and Genetics*, 38(3), p. 559.

- Gatta, V., Raicu, F., Ferlin, A., Antonucci, I., Scioletti, A.P., Garolla, A., Palka, G., Foresta, C. & Stuppia, L. (2010) 'Testis transcriptome analysis in male infertility: new insight on the pathogenesis of oligo-azoospermia in cases with and without AZFc microdeletion', *BMC Genomics*, 11(1), p. 401.
- Ghieh, F., Izard, V., Poulain, M., Fortemps, J., Kazdar, N., Mandon-Pepin, B., Ferlicot, S., Ayoubi, J.M. & Vialard, F. (2022) 'Cryptic splice site poisoning and meiotic arrest caused by a homozygous frameshift mutation in RBMXL2: A case report', *Andrologia*, 54(11), p. e14595.
- Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D.L. (2008) 'RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes', *Nature structural & molecular biology*, 15(1), p. 71.
- Goldstrohm, A.C., Hall, T.M.T. & McKenney, K.M. (2018) 'Post-transcriptional Regulatory Functions of Mammalian Pumilio Proteins'. *Trends in Genetics* 34 (12) p.pp. 972–990.
- Gonatopoulos-Pournatzis, T. & Cowling, V.H. (2014) 'Cap-binding complex (CBC)', *The Biochemical journal*, 457(2), pp. 231–242.
- Grellscheid, S., Dalgliesh, C., Storbeck, M., Best, A., Liu, Y., Jakubik, M., Mende, Y., Ehrmann, I., Curk, T., Rossbach, K., Bourgeois, C.F., Stévenin, J., Grellscheid, D., Jackson, M.S., Wirth, B. & Elliott, D.J. (2011) 'Identification of Evolutionarily Conserved Exons as Regulated Targets for the Splicing Activator Tra2 β in Development' Wendy A. Bickmore (ed.), *PLoS Genetics*, 7(12), p. e1002390.
- Griswold, M.D. (2016) 'Spermatogenesis: The Commitment to Meiosis', *Physiological Reviews*, 96(1), pp. 1–17.
- Grozdanov, P.N., Li, J., Yu, P., Yan, W. & MacDonald, C.C. (2018) 'Cstf2t Regulates expression of histones and histone-like proteins in male germ cells', *Andrology*, 6(4), pp. 605–615.
- Hammoud, S., Emery, B.R., Dunn, D., Weiss, R.B. & Carrell, D.T. (2009) 'Sequence alterations in the YBX2 gene are associated with male factor infertility.', *Fertility and sterility*, 91(4), pp. 1090–1095.
- Hannigan, M.M., Zagore, L.L. & Licatalosi, D.D. (2017) 'Ptpb2 Controls an Alternative Splicing Network Required for Cell Communication during Spermatogenesis', *Cell Reports*, 19(12), pp. 2598–2612.
- Harris, J.C., Martinez, J.M., Grozdanov, P.N., Bergeson, S.E., Grammas, P. & MacDonald, C.C. (2016) 'The Cstf2t polyadenylation gene plays a sex-specific role in learning behaviors in mice', *PLoS ONE*, 11(11), .
- Harrison, P.M., Kumar, A., Lang, N., Snyder, M. & Gerstein, M. (2002) 'A question of size: the eukaryotic

- proteome and the problems in defining it', *Nucleic acids research*, 30(5), pp. 1083–1090.
- Huang, Y.S., Mendez, R., Fernandez, M. & Richter, J.D. (2023) 'CPEB and translational control by cytoplasmic polyadenylation: impact on synaptic plasticity, learning, and memory', *Molecular Psychiatry* 2023, pp. 1–9.
- Huot, tienne, Vogel, G., Zabarauskas, A., Tuan-Anh Ngo, C., Coulombe-Huntington, J., Majewski, J. & phane Richard, S. (2012) 'Molecular Cell The Sam68 STAR RNA-Binding Protein Regulates mTOR Alternative Splicing during Adipogenesis', *Molecular Cell*, 46pp. 187–199.
- 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*, 65(3), pp. 274–287.
- Imai, A., Hagiwara, Y., Niimi, Y., Tokumoto, T., Saga, Y. & Suzuki, A. (2020) 'Mouse dead end1 acts with Nanos2 and Nanos3 to regulate testicular teratoma incidence', *PLOS ONE*, 15(4), p. e0232047.
- Jamsai, D., Watkins, D.N., O'Connor, A.E., Merriner, D.J., Gursoy, S., Bird, A.D., Kumar, B., Miller, A., Cole, T.J., Jenkins, B.J. & O'Bryan, M.K. (2017) 'In vivo evidence that RBM5 is a tumour suppressor in the lung', *Scientific Reports*, 7(1), .
- Jan, S.Z., Vormer, T.L., Jongejan, A., Röling, M.D., Silber, S.J., de Rooij, D.G., Hamer, G., Repping, S. & van Pelt, A.M.M. (2017) 'Unraveling transcriptome dynamics in human spermatogenesis.', *Development (Cambridge, England)*, 144(20), pp. 3659–3673.
- Kalsotra, A. & Cooper, T.A. (2011) 'Functional consequences of developmentally regulated alternative splicing'. *Nature Reviews Genetics* 12 (10) p.pp. 715–729.
- Kan, Z., Garrett-Engle, P.W., Johnson, J.M. & Castle, J.C. (2005) 'Evolutionarily conserved and diverged alternative splicing events show different expression and functional profiles', *Nucleic Acids Res*, 33pp. 5659–5666.
- Kojima, M.L., De Rooij, D.G. & Page, D.C. (2019) 'Amplification of a broad transcriptional program by a common factor triggers the meiotic cell cycle in mice', *eLife*, 8.
- 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), .

- Kress, C., Gautier-Courteille, C., Osborne, H.B., Babinet, C. & Paillard, L. (2007) 'Inactivation of CUG-BP1/CELF1 Causes Growth, Viability, and Spermatogenesis Defects in Mice', *Molecular and Cellular Biology*, 27(3), pp. 1146–1157.
- Kübler, M., Götz, P., Braumandl, A., Beck, S., Ishikawa-Ankerhold, H. & Deindl, E. (2021) 'Impact of c57bl/6j and sv-129 mouse strain differences on ischemia-induced postnatal angiogenesis and the associated leukocyte infiltration in a murine hindlimb model of ischemia', *International Journal of Molecular Sciences*, 22(21), .
- Kuwano, Y., Nishida, K., Kajita, K., Satake, Y., Akaike, Y., Fujita, K., Kano, S., Masuda, K. & Rokutan, K. (2015) 'Transformer 2 β and miR-204 regulate apoptosis through competitive binding to 3' UTR of BCL2 mRNA', *Cell Death and Differentiation*, 22(5), pp. 815–825.
- Lahn, B.T., Page, D.C., Lahn & Page (1997) 'Functional coherence of the human Y chromosome', *Science*, 278(5338), pp. 675–680.
- Lander, E.S. (2011) 'Initial impact of the sequencing of the human genome', *Nature* 2011 470:7333, 470(7333), pp. 187–197.
- 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.
- Legrand, J.M.D., Chan, A.L., La, H.M., Rossello, F.J., Änkö, M.L., Fuller-Pace, F. V. & Hobbs, R.M. (2019) 'DDX5 plays essential transcriptional and post-transcriptional roles in the maintenance and function of spermatogonia', *Nature Communications*, 10(1), pp. 1–21.
- Legrand, J.M.D. & Hobbs, R.M. (2018) 'RNA processing in the male germline: Mechanisms and implications for fertility'. *Seminars in Cell and Developmental Biology* 79 p.pp. 80–91.
- Li, H., Liang, Z., Yang, J., Wang, D., Wang, H., Zhu, M., Geng, B. & Xu, E.Y. (2019) 'DAZL is a master translational regulator of murine spermatogenesis.', *National science review*, 6(3), pp. 455–468.
- Licatalosi, D.D. (2016) '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*. [Online]. Springer New York LLC. pp. 123–151.
- Lin, Y. & Page, D.C. (2005) 'Dazl deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice', *Developmental Biology*, 288(2), pp. 309–316.

- Ling, J.P., Chhabra, R., Merran, J.D., Schaughency, P.M., Wheelan, S.J., Corden, J.L. & Wong, P.C. (2016) 'PTBP1 and PTBP2 Repress Nonconserved Cryptic Exons.', *Cell reports*, 17(1), pp. 104–113.
- Liu, D., Brockman, J.M., Dass, B., Hutchins, L.N., Singh, P., McCarrey, J.R., MacDonald, C.C. & Graber, J.H. (2007) 'Systematic variation in mRNA 3'-processing signals during mouse spermatogenesis', *Nucleic acids research*, 35(1), pp. 234–246.
- 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', *PLOS Genetics*, 5(11), p. e1000707.
- Lu, X., Li, N., Shushakova, N., Schmitt, R., Menne, J., Susnik, N., Meier, M., Leitges, M., Haller, H., Gueler, F. & Rong, S. (2012) 'C57BL/6 and 129/SV mice: Genetic difference to renal ischemia-reperfusion', *Journal of Nephrology*, 25(5), pp. 738–743.
- Lunde, B.M., Moore, C. & Varani, G. (2007) 'RNA-binding proteins: modular design for efficient function', *Nature reviews. Molecular cell biology*, 8(6), pp. 479–490.
- Ma, K., Inglis, J.D., Sharkey, A., Bickmore, W.A., Hill, R.E., Prosser, E.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.
- MacDonald, C.C. (2019) 'Tissue-specific mechanisms of alternative polyadenylation: Testis, brain, and beyond (2018 update)'. Wiley Interdisciplinary Reviews: RNA 10 (4).
- Mardon, H.J., Sebastio, G. & Baralle, F.E. (1987) 'A role for exon sequences in alternative splicing of the human fibronectin gene', *Nucleic acids research*, 15(19), pp. 7725–7733.
- Matsunaga, S., Takata, H., Morimoto, A., Hayashihara, K., Higashi, T., Akatsuchi, K., Mizusawa, E., Yamakawa, M., Ashida, M., Matsunaga, T.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.

- Mende, Y., Jakubik, M., Riessland, M., Schoenen, F., Roßbach, K., Kleinridders, A., Köhler, C., Buch, T. & Wirth, B. (2010) 'Deficiency of the splicing factor Sfrs10 results in early embryonic lethality in mice and has no impact on full-length SMN/Smn splicing', *Human Molecular Genetics*, 19(11), pp. 2154–2167.
- Mikedis, M.M., Fan, Y., Nicholls, P.K., Endo, T., Jackson, E.K., Cobb, S.A., de Rooij, D.G. & Page, D.C. (2020) 'Dazl mediates a broad translational program regulating expansion and differentiation of spermatogonial progenitors', *eLife*, 9pp. 1–96.
- Mischo, H.E. & Proudfoot, N.J. (2013) 'Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast', *Biochimica et biophysica acta*, 1829(1), pp. 174–185.
- Mittleman, B.E., Pott, S., Warland, S., Zeng, T., Mu, Z., Kaur, M., Gilad, Y. & Li, Y. (2020) 'Alternative polyadenylation mediates genetic regulation of gene expression', *eLife*, 9pp. 1–21.
- Miyata, H., Morohoshi, A. & Ikawa, M. (2020) 'Analysis of the sperm flagellar axoneme using gene-modified mice', *Experimental animals*, 69(4), pp. 374–381.
- Monzón-Casanova, E., Matheson, L.S., Tabbada, K., Zarnack, K., Smith, C.W. & Turner, M. (2020) 'Polypyrimidine tract-binding proteins are essential for B cell development', *eLife*, 9.
- Murakami, H. & Keeney, S. (2008) 'Regulating the formation of DNA double-strand breaks in meiosis', *Genes & Development*, 22(3), p. 286.
- Nandakumar, P., Mansouri, A. & Das, S. (2017) 'The role of ATRX in glioma biology', *Frontiers in Oncology*, 7(SEP), p. 236.
- Niwa, M., Rose, S.D. & Berget, S.M. (1990) 'In vitro polyadenylation is stimulated by the presence of an upstream intron', *Genes & development*, 4(9), pp. 1552–1559.
- 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.
- Ogonuki, N., Mori, M., Shinmen, A., Inoue, K., Mochida, K., Ohta, A. & Ogura, A. (2010) 'The Effect on Intracytoplasmic Sperm Injection Outcome of Genotype, Male Germ Cell Stage and Freeze-Thawing in Mice', *PLOS ONE*, 5(6), p. e11062.

- Ottoz, D.S.M. & Berchowitz, L.E. (2020) 'The role of disorder in RNA binding affinity and specificity', *Open Biology*, 10(12), .
- Oud, M.S., Smits, R.M., Smith, H.E., Mastrososa, F.K., Holt, G.S., Houston, B.J., de Vries, P.F., Alobaidi, B.K.S., Batty, L.E., Ismail, H., Greenwood, J., Sheth, H., Mikulasova, A., Astuti, G.D.N., Gilissen, C., McEleny, K., Turner, H., Coxhead, J., Cockell, S., et al. (2022) 'A de novo paradigm for male infertility', *Nature Communications* 2022 13:1, 13(1), pp. 1–10.
- Pan, J., Eckardt, S., Leu, N.A., Buffone, M.G., Zhou, J., Gerton, G.L., McLaughlin, K.J. & Wang, P.J. (2009) 'Inactivation of Nxf2 causes defects in male meiosis and age-dependent depletion of spermatogonia', *Developmental Biology*, 330(1), pp. 167–174.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J. (2008) 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing', *Nature Genetics*, 40(12), pp. 1413–1415.
- Papasaikas, P., Valcárcel, J., Eu, J.V. & Valcárcel, J. (2016) 'Special Issue: 40 Years of TiBS The Spliceosome: The Ultimate RNA Chaperone and Sculptor', *Trends in Biochemical Sciences*, 41(1), pp. 33–45.
- Paronetto, M.P., Messina, V., Barchi, M., Geremia, R., Richard, S. & Sette, C. (2011) 'Sam68 marks the transcriptionally active stages of spermatogenesis and modulates alternative splicing in male germ cells', *Nucleic Acids Research*, 39(12), p. 4961.
- Paronetto, M.P. & Sette, C. (2010) 'Role of RNA-binding proteins in mammalian spermatogenesis.', *International journal of andrology*, 33(1), pp. 2–12.
- Patel, A.A. & Steitz, J.A. (2003) 'Splicing double: insights from the second spliceosome', *Nature Reviews Molecular Cell Biology* 2003 4:12, 4(12), pp. 960–970.
- Pertea, M., Shumate, A., Pertea, G., Varabyou, A., Breitwieser, F.P., Chang, Y.C., Madugundu, A.K., Pandey, A. & Salzberg, S.L. (2018) 'CHES: A new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise', *Genome Biology*, 19(1), pp. 1–14.
- Prelich, G. (2012) 'Gene Overexpression: Uses, Mechanisms, and Interpretation', *Genetics*, 190(3), p. 841.
- Proudfoot, N. (1991) 'Poly(A) signals', *Cell*, 64(4), pp. 671–674.

- Qin, H., Ni, H., Liu, Y., Yuan, Y., Xi, T., Li, X. & Zheng, L. (2020) 'RNA-binding proteins in tumor progression', *Journal of Hematology and Oncology*, 13(1), pp. 1–23.
- Quinlan, A.R. & Hall, I.M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics (Oxford, England)*, 26(6), pp. 841–842.
- Ramsbottom, S.A., Thelwall, P.E., Wood, K.M., Clowry, G.J., Devlin, L.A., Silbermann, F., Spiewak, H.L., Shril, S., Molinari, E., Hildebrandt, F., Gunay-Aygun, M., Saunier, S., Cordell, H.J., Sayer, J.A. & Miles, C.G. (2020) 'Mouse genetics reveals Barttin as a genetic modifier of Joubert syndrome', *Proceedings of the National Academy of Sciences of the United States of America*, 117(2), pp. 1113–1118.
- Remington, S.J. (2011) 'Green fluorescent protein: A perspective', *Protein Science : A Publication of the Protein Society*, 20(9), p. 1509.
- 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) p.pp. 24–26.
- Rogalska, M.E., Vivori, C. & Valcárcel, J. (2022) 'Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects', *Nature Reviews Genetics* 2022, pp. 1–19.
- Ruggiu, M., Speed, R., Taggart, M., McKay, S.J., Kilanowski, F., Saunders, P., Dorin, J. & Cooke, H.J. (1997) 'The mouse Dazl gene encodes a cytoplasmic protein essential for gametogenesis', *Nature*, 389(6646), pp. 73–77.
- Sakai, K., Ito, C., Wakabayashi, M., Kanzaki, S., Ito, T., Takada, S., Toshimori, K., Sekita, Y. & Kimura, T. (2019) 'Usp26 mutation in mice leads to defective spermatogenesis depending on genetic background', *Scientific Reports* 2019 9:1, 9(1), pp. 1–12.
- Saunders, P.T.K., Turner, J.M.A., Ruggiu, M., Taggart, M., Burgoyne, P.S., Elliott, D. & Cooke, H.J. (2003) 'Absence of mDazl produces a final block on germ cell development at meiosis', *Reproduction*, 126(5), pp. 589–597.
- Schmid, R., Grellscheid, S.N., Ehrmann, I., Dalglish, C., Danilenko, M., Paronetto, M.P., Pedrotti, S., Grellscheid, D., Dixon, R.J., Sette, C., Eperon, I.C. & Elliott, D.J. (2013) 'The splicing landscape is globally reprogrammed during male meiosis.', *Nucleic acids research*, 41(22), pp. 10170–10184.
- Sciabica, K.S. & Hertel, K.J. (2006) 'The splicing regulators Tra and Tra2 are unusually potent activators of pre-mRNA splicing', *Nucleic Acids Research*, 34(22), pp. 6612–6620.

- Senoo, M., Takijiri, T., Yoshida, N., Ozawa, M. & Ikawa, M. (2019) 'PTBP1 contributes to spermatogenesis through regulation of proliferation in spermatogonia', *Journal of Reproduction and Development*, 65(1), pp. 37–46.
- Sharan, S.K., Pyle, A., Coppola, V., Babus, J., Swaminathan, S., Benedict, J., Swing, D., Martin, B.K., Tessarollo, L., Evans, J.P., Flaws, J.A. & Handel, M.A. (2004) 'BRCA2 deficiency in mice leads to meiotic impairment and infertility', *Development*, 131(1), pp. 131–142.
- Sharp, P., Roberts, R. & Shi, Y. (2017) 'Mechanistic insights into precursor messenger RNA splicing by the spliceosome', *Nature Reviews Molecular Cell Biology* 2017 18:11, 18(11), pp. 655–670.
- 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.
- Shuman, S. (1997) 'Origins of mRNA identity: Capping enzymes bind to the phosphorylated C- terminal domain of RNA polymerase II', *Proceedings of the National Academy of Sciences of the United States of America*, 94(24), pp. 12758–12760.
- Siachisumo, C., Luzzi, S., Aldalaqan, S., Hysenaj, G., Dalglish, C., Cheung, K., Gazzara, M.R., Yonchev, I.D., James, K., Chadegani, M.K., Ehrmann, I., Smith, G.R., Cockell, S.J., Munkley, J., Wilson, S.A., Barash, Y. & Elliott, D.J. (2023) 'An anciently diverged family of RNA binding proteins maintain correct splicing of ultra-long exons through cryptic splice site repression', *bioRxiv*, p. 2023.08.15.553384.
- Sibley, C.R., Blazquez, L. & Ule, J. (2016) 'Lessons from non-canonical splicing', *Nature Reviews Genetics* 2016 17:7, 17(7), pp. 407–421.
- Singh, K.D., Zheng, X., Milstein, S., Keller, M., Roschitzki, B., Grossmann, J. & Hengartner, M.O. (2017) 'Differential regulation of germ line apoptosis and germ cell differentiation by CPEB family members in *C. elegans*', *PLoS ONE*, 12(7), .
- Skrisovska, L., Bourgeois, C.F., Stefl, R., Grellscheid, S.N., Kister, L., Wenter, P., Elliott, D.J., Stevenin, J. & Allain, F.H.T. (2007) 'The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction', *EMBO Reports*, 8(4), p. 372.
- Smith, R.W.P., Anderson, R.C., Smith, J.W.S., Brook, M., Richardson, W.A. & Gray, N.K. (2011) 'DAZAP1, an RNA-binding protein required for development and spermatogenesis, can regulate mRNA translation', *RNA*, 17(7), p. 1282.

- Snyder, E., Chukrallah, L., Seltzer, K., Goodwin, L. & Braun, R.E. (2020) 'ADAD1 and ADAD2, testis-specific adenosine deaminase domain-containing proteins, are required for male fertility', *Scientific Reports*, 10(1), .
- Snyder, E., Soundararajan, R., Sharma, M., Dearth, A., Smith, B. & Braun, R.E. (2015) 'Compound Heterozygosity for Y Box Proteins Causes Sterility Due to Loss of Translational Repression', *PLoS Genetics*, 11(12), p. 1005690.
- Soh, Y.Q.S., Mikedis, M.M., Kojima, M., Godfrey, A.K., de Rooij, D.G. & Page, D.C. (2017) 'Meioc maintains an extended meiotic prophase I in mice', *PLoS Genetics*, 13(4), .
- Sone, R., Taimatsu, K., Ohga, R., Nishimura, T., Tanaka, M. & Kawahara, A. (2020) 'Critical roles of the ddx5 gene in zebrafish sex differentiation and oocyte maturation', *Scientific Reports*, 10(1), p. 14157.
- Song, H., Wang, L., Chen, D. & Li, F. (2020) 'The function of pre-mRNA alternative splicing in mammal spermatogenesis'. *International Journal of Biological Sciences* 16 (1) p.pp. 38–48.
- Stoilov, P., Daoud, R., Nayler, O. & Stamm, S. (2004) 'Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA.', *Human molecular genetics*, 13(5), pp. 509–24.
- Sutherland, J.M., Siddall, N.A., Hime, G.R. & McLaughlin, E.A. (2015) 'RNA binding proteins in spermatogenesis: an in depth focus on the Musashi family.', *Asian journal of andrology*, 17(4), pp. 529–536.
- Svitkin, Y. V., Pause, A., Haghighat, A., Pyronnet, S., Witherell, G., Belsham, G.J. & Sonenberg, N. (2001) 'The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure', *RNA (New York, N.Y.)*, 7(3), pp. 382–394.
- Tacke, R., Tohyama, M., Ogawa, S. & Manley, J.L. (1998) 'Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing', *Cell*, 93(1), pp. 139–148.
- Tavleeva, M.M., Belykh, E.S., Rybak, A. V., Rasova, E.E., Chernykh, A.A., Ismailov, Z.B. & Velegzhaninov, I.O. (2022) 'Effects of Antioxidant Gene Overexpression on Stress Resistance and Malignization In Vitro and In Vivo: A Review', *Antioxidants*, 11(12), .
- Tay, J. & Richter, J.D. (2001) 'Germ Cell Differentiation and Synaptonemal Complex Formation Are Disrupted in CPEB Knockout Mice', *Developmental Cell*, 1(2), pp. 201–213.

- Urlaub, H., Hartmuth, K. & Lührmann, R. (2002) 'A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles', *Methods*, 26(2), pp. 170–181.
- Venables, J.P., Elliott, D.J., Makarova, O. V, Makarov, E.M., Cooke, H.J. & Eperon, I.C. (2000) 'RBMV, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing.', *Human molecular genetics*, 9(5), pp. 685–694.
- Venables, J.P. & Eperon, I.C. (1999) 'The roles of RNA-binding proteins in spermatogenesis and male infertility', *Current Opinion in Genetics and Development*, 9(3), pp. 346–354.
- Wahl, M.C., Will, C.L. & Lührmann, R. (2009) 'The Spliceosome: Design Principles of a Dynamic RNP Machine'. *Cell* 136 (4) p.pp. 701–718.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, 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, 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., Gerstein, M. & Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nature Reviews Genetics* 2008 10:1, 10(1), pp. 57–63.
- Wassarman, K.M. & Steitz, J.A. (1993) 'Association with terminal exons in pre-mRNAs: a new role for the U1 snRNP?', *Genes & development*, 7(4), pp. 647–659.
- Westerveld, G.H., Gianotten, J., Leschot, N.J., van der Veen, F., Repping, S. & Lombardi, M.P. (2004) 'Heterogeneous nuclear ribonucleoprotein G-T (HNRNP G-T) mutations in men with impaired spermatogenesis', *Molecular Human Reproduction*, 10(4), pp. 265–269.
- 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), pp. 1–12.
- Wojtas, M.N., Pandey, R.R., Mendel, M., Homolka, D., Sachidanandam, R. & Pillai, R.S. (2017) 'Regulation of m6A Transcripts by the 3'→5' RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline', *Molecular Cell*, 68(2), pp. 374-387.e12.

- Xu, H., Guo, J., Wu, W., Han, Q., Huang, Y., Wang, Y., Li, C., Cheng, X., Zhang, P. & Xu, Y. (2022) 'Deletion of Hnrnpk Gene Causes Infertility in Male Mice by Disrupting Spermatogenesis', *Cells*, 11(8), p. 1277.
- Xu, X., Aprelikova, O., Moens, P., Deng, C.X. & Furth, P.A. (2003) 'Impaired meiotic DNA-damage repair and lack of crossing-over during spermatogenesis in BRCA1 full-length isoform deficient mice', *Development (Cambridge, England)*, 130(9), pp. 2001–2012.
- Ben Yamin, B., Ahmed-Seghir, S., Tomida, J., Despras, E., Pouvelle, C., Yurchenko, A., Goulas, J., Corre, R., Delacour, Q., Droin, N., Dessen, P., Goidin, D., Lange, S.S., Bhetawal, S., Mitjavila-Garcia, M.T., Baldacci, G., Nikolaev, S., Cadoret, J.C., Wood, R.D., et al. (2021) 'DNA polymerase zeta contributes to heterochromatin replication to prevent genome instability', *The EMBO journal*, 40(21), .
- Zagore, L.L., Grabinski, S.E., Sweet, T.J., Hannigan, M.M., Sramkoski, R.M., Li, Q. & Licatalosi, D.D. (2015) 'RNA Binding Protein Ptp2 Is Essential for Male Germ Cell Development.', *Molecular and cellular biology*, 35(23), pp. 4030–4042.
- Zagore, L.L., Sweet, T.J., Hannigan, M.M., Weyn-Vanhentenryck, S.M., Jobava, R., Hatzoglou, M., Zhang, C. & Licatalosi, D.D. (2018) 'DAZL Regulates Germ Cell Survival through a Network of PolyA-Proximal mRNA Interactions', *Cell Reports*, 25(5), pp. 1225-1240.e6.
- Zarnegar, B.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.
- Zhang, J., Zheng, N. & Zhou, P. (2003) 'Exploring the functional complexity of cellular proteins by protein knockout', *Proceedings of the National Academy of Sciences of the United States of America*, 100(SUPPL. 2), pp. 14127–14132.
- Zheng, J.T., Lin, C.X., Fang, Z.Y. & Li, H.D. (2020) 'Intron Retention as a Mode for RNA-Seq Data Analysis'. *Frontiers in Genetics* 11 p.p. 586.
- Zoabi, M., Nadar-Ponniah, P.T., Khoury-Haddad, H., Usaj, M., Budowski-Tal, I., Haran, T., Henn, A., Mandel-Gutfreund, Y. & Ayoub, N. (2014) 'RNA-dependent chromatin localization of KDM4D lysine demethylase promotes H3K9me3 demethylation', *Nucleic Acids Research*, 42(21), p. 13026.

Appendix 1: Thesis-related publications and conferences attended.

Publications:

Saad Aldalagan, Caroline Dalglish, 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.

An anciently diverged family of RNA binding proteins maintain correct splicing of ultra-long exons through cryptic splice site repression.

Chileleko Siachisumo^{1*}, Sara Luzzi^{1*}, **Saad Aldalagan**^{1*}, Gerald Hysenaj¹, Caroline Dalglish¹, Kathleen Cheung², Matthew R Gazzara³, Ivaylo D Yonchev⁴, Katherine James⁵, Mahsa Kheirollahi Chadegani¹, Ingrid Ehrmann¹, Graham R Smith², Simon J Cockell², Jennifer Munkley¹, Stuart A Wilson⁴, Yoseph Barash³ and David J Elliott (**Submitted**).

Poster presentation:

RBMX-family proteins bind to and regulate splicing of ultra-long exons

Saad Aldalagan¹, Chileleko Siachisumo¹, Sara Luzzi¹, Gerald Hysenaj¹, Emma Scott¹, Matthew Gazzara², Yoseph Barash² and David J. Elliott¹.

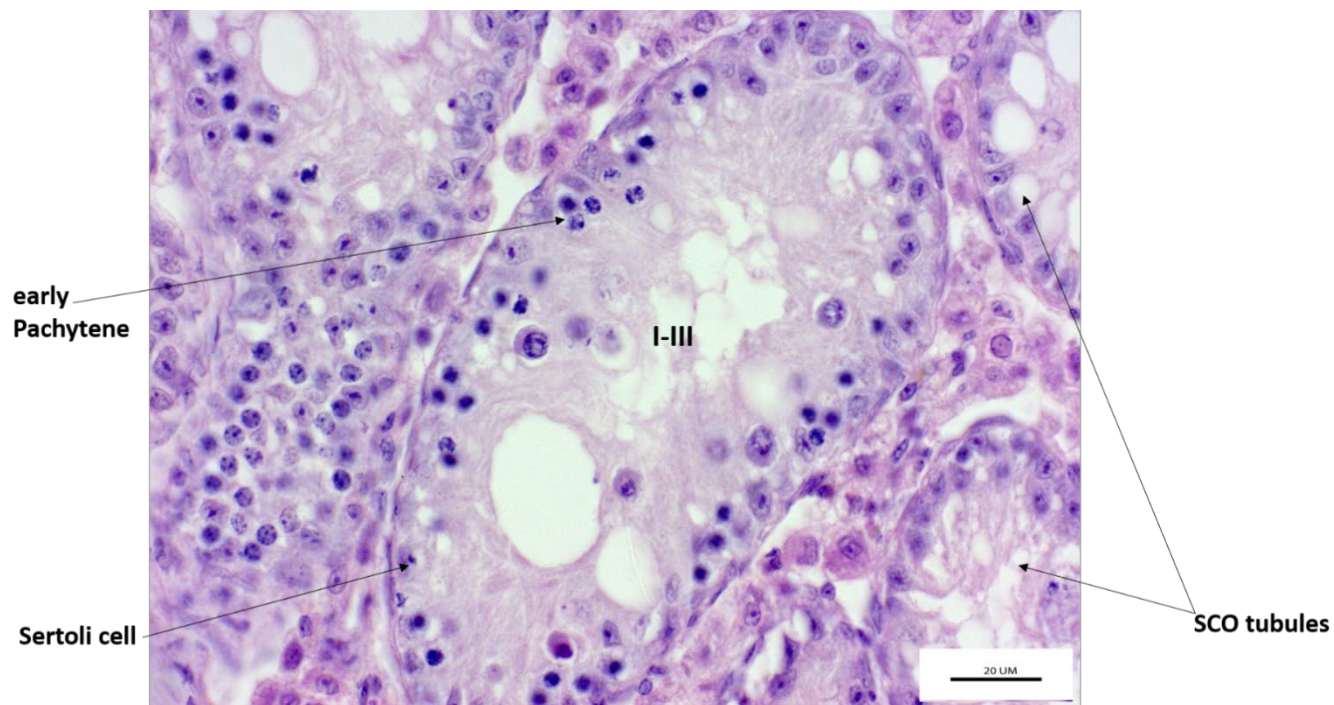
Presented on: 8th UK RNA Splicing Workshop 27th to 29th January, 2023 Rydal Hall, Cumbria, UK

Appendix 2: Table 1.1

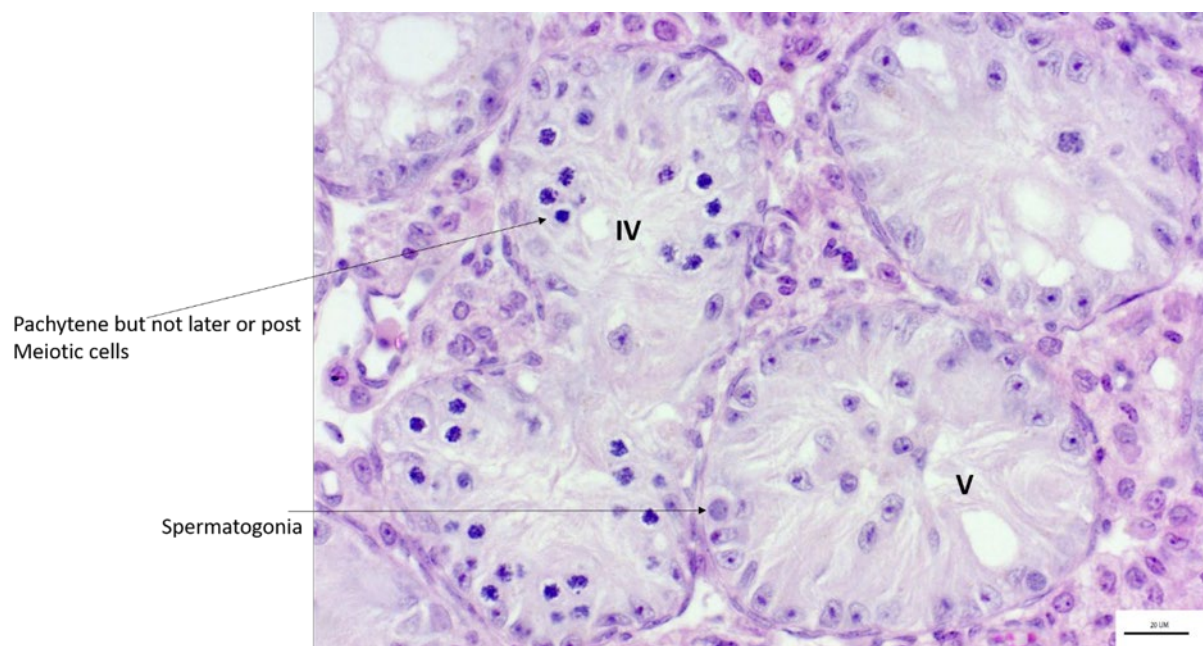
- 1-(Best et al., 2014)
- 2-(Campopiano et al., 2017)
- 3-(Snyder et al., 2020)
- 4-(Dass et al., 2007)
- 5-(Ehrmann et al., 2019)
- 6-(Licatalosi, 2016)
- 7-(Hammoud et al., 2009)
- 8-(Beeram et al., 2019)
- 9-(Chen et al., 2012)
- 10-(Goldstrohm et al., 2018)
- 11-(Huang et al., 2023)
- 12-(Licatalosi, 2016)
- 13-(Paronetto et al., 2011)
- 14-(Ehrmann et al., 2013)
- 15-(Kress et al., 2007)
- 16-(Chi et al., 2011)
- 17-(Ling et al., 2016)
- 18-(Monzón-Casanova et al., 2020)
- 19-(Oud et al., 2022)
- 20-(Legrand et al., 2019)

Appendix 3: RBMXL2 KO Sv/129 testis sections. From figure 3.6

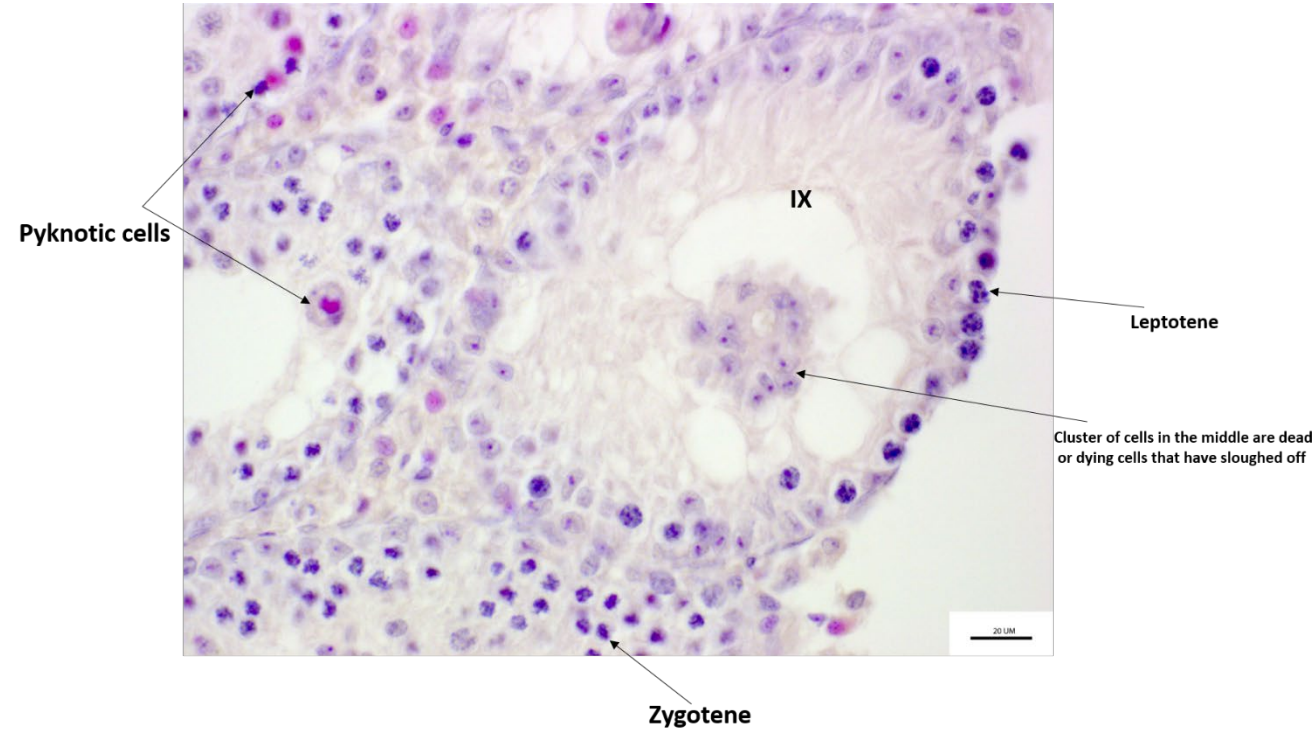
Stage I-III



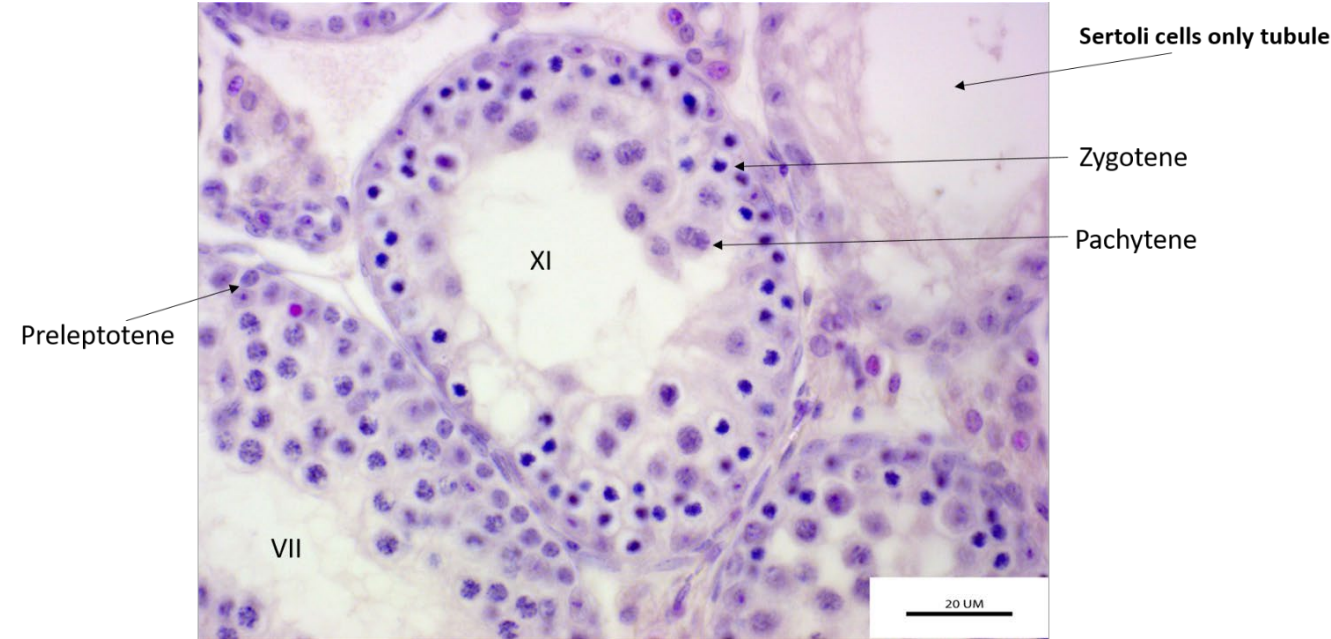
Stage IV-V



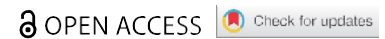
Stage IX



Stage XI



REVIEW



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

Saad Aldalaqan, Caroline Dalglish, 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.

Received 14 October 2021
Revised 16 November 2021
Accepted 30 November 2021

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.

CONTACT David J. Elliott  David.Elliott@ncl.ac.uk  Newcastle University Bioscience Institute, Newcastle University, Central Parkway Newcastle NE1 3BZ, UK

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

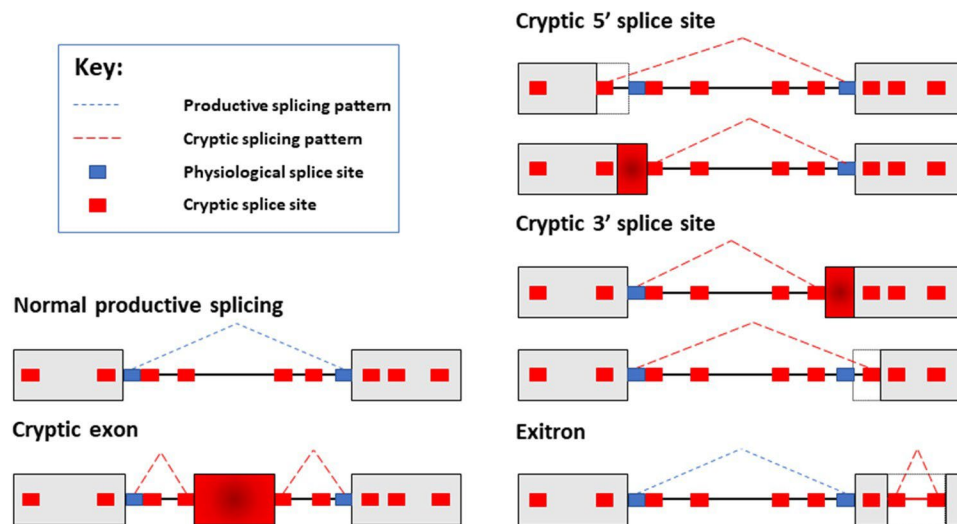


Figure 1. Schematic diagram of cryptic splicing patterns. Most genes are split between exons (shown as gray boxes here) and introns (shown as connecting lines between the boxes). Normal patterns of splice site selection will involve the spliceosome recognizing bona fide splice sites, and joining exons 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 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.

acterized 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

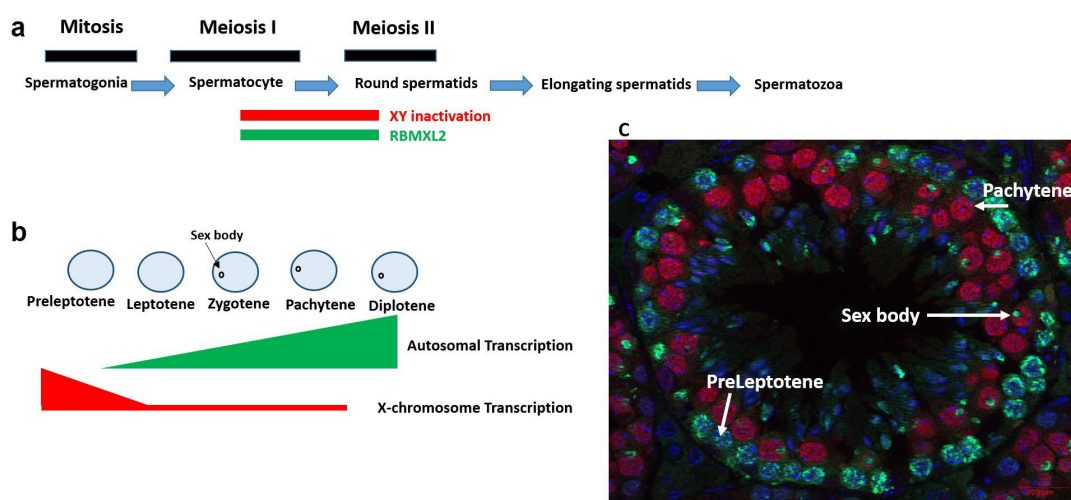


Figure 2. RBMXL2 is expressed during diplotene and pachytene of male meiotic prophase. a. Mouse germ cell development, showing the expression window of RBMXL2 and time period of XY inactivation. b. Transcription patterns of the X chromosome and autosomes during meiotic prophase. c. Seminiferous tubule counterstained with antibodies specific to RBMXL2 protein (pseudocolored red, detected within nuclei of cells in pachytene in this tubule) and γ H2AX (green color, detected within nuclei of preleptotene cells and within the sex body 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 retrotransposition 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 remainder 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 MCS1. 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

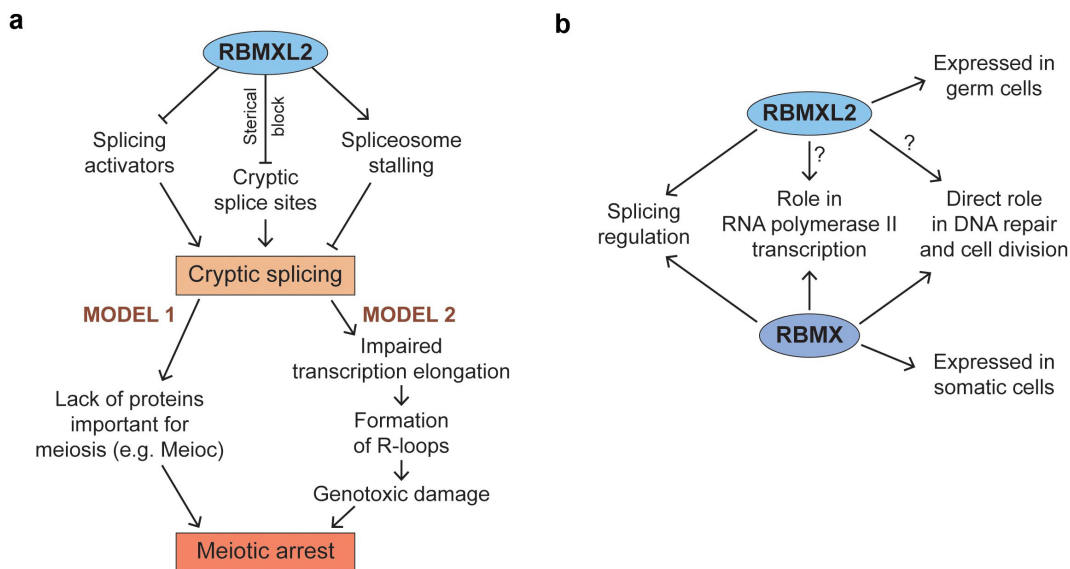


Figure 3. Mechanistic models to explain the impact of cryptic splicing on meiosis, and possible additional roles of RBMXL2.

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

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by the BBSRC (grants Biotechnology and Biological Sciences Research Council BB/S008039/1 and BB/P006612/1) and the King Fahad Medical City, Ministry of Health, Kingdom of Saudi Arabia. No potential competing interest was reported by the authors.

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-Engle 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, Necseulea 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, Vallotton 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, Nij 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. Meiosis 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, Tavares UL, et al. A targeted antisense therapeutic approach for Hutchinson–Gilford progeria syndrome. *Nat Med.* **2021**;27(3):536–545.
- [44] Elliott DJ, Dalglish 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 Paralogues 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.

- [1] Legrand JMD, Hobbs RM. RNA processing in the male germline: mechanisms and implications for fertility. *Semin Cell Dev Biol.* 2018;79:80–91.
- [2] 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.
- [3] Hannigan MM, Zagore LL, Licatalosi DD. Ptpb2 Controls an Alternative Splicing Network Required for Cell Communication during Spermatogenesis. *Cell Rep.* 2017;19(12):2598–2612.
- [4] Jp L, Chhabra R, Jd M, et al. PTBP1 and PTBP2 Repress Nonconserved Cryptic Exons. *Cell Rep.* 2016;17(1):104–113.
- [5] Naro C, Sette C. Timely-regulated intron retention as device to fine-tune protein expression. *Cell Cycle.* 2017;16(14):1321–1322.
- [6] Nasim MT, Chernova TK, Chowdhury HM, et al. HnRNP G and Tra2 β : opposite effects on splicing

matched by antagonism in RNA binding. *Hum Mol Genet.* 2003;12(11):1337–1348.

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.

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 HP1 α . *Blood.* 2018;132(Supplement 1):883

