An ancient germ cell-specific RNA-binding protein protects the germline from cryptic splice site poisoning

Citation for published version:
Ehrmann, I, Crichton, JH, Gazzara, MR, James, K, Liu, Y, Grellscheid, SN, Curk, T, de Rooij, D, Steyn, JS, Cockell, S, Adams, IR, Barash, Y & Elliott, DJ 2019, 'An ancient germ cell-specific RNA-binding protein protects the germline from cryptic splice site poisoning' eLIFE, vol. 8. DOI: 10.7554/eLife.39304

Digital Object Identifier (DOI):
10.7554/eLife.39304

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
eLIFE

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
An ancient germ cell-specific RNA binding protein protects the germline from cryptic splice site poisoning

Ingrid Ehrmann¹, James H. Crichton², Matthew R. Gazzara³,⁴, Katherine James⁵, Yilei Liu¹,⁶, Sushma Grellscheid¹,⁷, Tomaz Curk⁸, Dirk G. de Rooij⁹,¹⁰, Jannetta S. Steyn¹¹, Simon J. Cockell¹¹, Ian R. Adams², Yoseph Barash³,¹² and David J. Elliott¹,¹³

¹ Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle NE1 3BZ
² MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU
³ Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA;
⁴ Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.
⁵ Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK.
⁶ Department of Plant and Microbial Biology, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
⁷ School of Biological and Biomedical Sciences, University of Durham, South Road, Durham DH1 3LE, UK
⁸ Laboratory of Bioinformatics, Faculty of Computer and Information Sciences, University of Ljubljana, Ljubljana, Slovenia.
⁹ Reproductive Biology Group, Division of Developmental Biology, Department of Biology, Faculty of Science, Utrecht University, 3584 CH, Utrecht, The Netherlands; and ¹⁰ Center for Reproductive Medicine, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands.
¹¹ Bioinformatics Support Unit, Faculty of Medical Sciences, Framlington Place, Newcastle University, Newcastle, NE2 4HH
¹² Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104, USA

Corresponding authors: Yoseph Barash, e-mail: yosephb@seas.upenn.edu
David Elliott, e-mail: David.Elliott@ncl.ac.uk

¹³ Lead contact.
Abstract
Male germ cells of all placental mammals express an ancient nuclear RNA binding protein of unknown function called RBMXL2. Here we find that deletion of the retrogene encoding RBMXL2 blocks spermatogenesis. Transcriptome analyses of age-matched deletion mice show that RBMXL2 controls splicing patterns during meiosis. In particular, RBMXL2 represses the selection of aberrant splice sites and the insertion of cryptic and premature terminal exons. Our data suggest an Rbmxl2 retrogene has been conserved across mammals as part of a splicing control mechanism that is fundamentally important to germ cell biology. We propose that this mechanism is essential to meiosis because it buffers the high ambient concentrations of splicing activators, thereby preventing poisoning of key transcripts and disruption to gene expression by aberrant splice site selection.

Keywords. RNA splicing, cryptic exons, male infertility, meiosis, gene expression.
Introduction

It has been a long recognised hallmark of mammalian gene expression patterns that there are extremely high levels of transcription and transcriptome complexity in testicular cells (de la Grange et al., 2010, Licatalosi, 2016, Soumillon et al., 2013, Pan et al., 2008, Wang et al., 2008, Clark et al., 2007, Grosso et al., 2008, Yeo et al., 2004). These high gene expression levels are thought to result from epigenetic changes that favour relaxed patterns of gene expression during meiosis – the unique form of division used to generate sperm and eggs (Soumillon et al., 2013). Many nuclear RNA binding proteins are differentially expressed during and immediately after meiosis (Grellscheid et al., 2011, Schmid et al., 2013). These include the nuclear RNA binding protein RBMXL2 (also known as hnRNP GT) that is expressed only during and immediately after male meiosis, but not in the preceding spermatogonial cells (Ehrmann et al., 2008) (Figure 1A). Consistent with an important function in germ cell development, genetic studies have identified point mutations within infertile men in the human chromosome 11 RBMXL2 gene (Westerveld et al., 2004). A further connection to male infertility is that RBMXL2 belongs to the same gene family as RBMY, which was historically the first human Y chromosome infertility gene identified in the search for the AZF (AZOOSPERMIA FACTOR) gene (Ma et al., 1993). RBMXL2 evolved ~65 million years ago via retro-transposition of an mRNA from the X chromosome-located RBMX gene (Figure 1A).

An RBMXL2 gene is found in all placental mammals, which is consistent with a fundamental role in germ cell biology. This role remains to be identified, but RBMXL2 protein has an N-terminal RNA Recognition Motif (abbreviated RRM, Figure 1B). RBMX and RBMY proteins are also nuclear RNA binding proteins that are very similar to RBMXL2 (73.2% and 36.8% overall identity to RBMX and RBMY, respectively; 93.7% and 77.2% identity within the RRMs, Figure 1 – figure supplement 1). The RNA binding specificity of RBMXL2 protein is unknown, but both RBMX and RBMY proteins bind to AA dinucleotide-containing RNA sequences (Clery et al., 2011, Moursy et al., 2014, Nasim et al., 2003). The RBMXL2, RBMX and RBMY proteins interact with and modulate the splicing activity of Tra2β and SR proteins in vitro (Figure 1B) (Clery et al., 2011, Moursy et al., 2014, Liu et al., 2009, Nasim et al., 2003, Elliott et al., 2000a), suggesting a role in splicing control. Maintaining proper ratios of mRNA splice isoforms can be critical in normal development (Kalsotra and Cooper, 2011), where changes in isoforms can have effects on encoded proteins ranging from major to subtle. Alternative splicing is known to be critical for germ cell development. For example, deletion of the splicing regulator protein PTBP2 within germ cells affects mRNA isoforms important for cell-cell communication with Sertoli cells (Hannigan et al., 2017). Some alternative splice events in the testis are conserved between humans and mice so may
control fundamental aspects of germ cell biology (Schmid et al., 2013). However many alternative splicing patterns are not conserved between humans and mice (Kan et al., 2005). RBMX protein is also reported to control transcription (Takemoto et al., 2007), affect DNA double strand repair and mitotic sister chromatid cohesion (Adamson et al., 2012, Matsunaga et al., 2012), and to bind to m6A methylated RNA (Liu et al., 2017).

The function of RBMXL2 and why this RNA binding protein has been conserved across placental mammals are not known. A major factor limiting understanding of endogenous RBMXL2 functions has been the absence of a reliable mouse model. Development of a mouse model is also critical to test the importance of this wider family of RNA binding proteins in germ cell development. Men carrying the AZFb deletion on their Y chromosomes are missing RBMY genes and undergo meiotic arrest. However, it is unclear if RBMY loss is causing this phenotype, because the deletion interval encompasses several other genes which could contribute to male infertility (Elliott, 2000, Vogt et al., 1996). Within the meiotic and immediately post-meiotic germ cells which express RBMXL2, the RBMX and RBMY genes are transcriptionally inactivated within a heterochromatic structure called the XY body (Wang, 2004). Meiosis thus provides a genetically tractable window to probe RBMXL2 function where there should be no redundancy effects possible with either RBMX or RBMY. Hence to discover what RBMXL2 does in the germline we have made a conditional Hrpmgt gene knockout mouse. Analysis of this knockout mouse reveals that RBMXL2 protein is essential for meiosis and has a major role in protecting the meiotic transcriptome from aberrant selection of cryptic splice sites that are normally ignored by the spliceosome. Our data suggest this fundamentally important process operates so efficiently in meiosis that it has been previously undetected, yet is critical to avoid male infertility caused by aberrant splicing of key meiotic transcripts.
Results

RBMXL2 protein is essential for male fertility

To test how important RBMXL2 protein is for male germline development we made a conditional mouse model in which we flanked the Rbmxl2 open reading frame with LoxP sites. Since Rbmxl2 is only expressed in the testis (Elliott et al., 2000b), we chose to delete the entire Rbmxl2 open reading frame to create a null allele. We achieved this by crossing our conditional model with a mouse strain expressing Cre recombinase under control of the ubiquitous Pgk promoter (experimental details are provided in the Methods). We confirmed deletion of this genomic region in homozygous Rbmxl2 gene knockout (Rbmxl2 -/-) mice by Southern blotting (Figure 1 –figure supplement 2) and the specific absence of the 50KDa RBMXL2 protein from knockout testes by Western blotting (Figure 1C).

Rbmxl2 -/- mice developed comparably to their wild type littermates but their testes were much smaller (Figure 2A and 2B). This small testis phenotype correlated with a severe disruption of testicular histology. Adult Rbmxl2 -/- mice contained cells undergoing meiosis but almost no post-meiotic cells (Figure 2C, meiotic spermatocytes are abbreviated Spc, and post-meiotic round spermatids are abbreviated Rtd). The epididymis dissected from Rbmxl2 -/- mice were completely devoid of sperm. Four wild type mice tested had an average of 7.07+/−1.39X10^6 epididymal sperm ml^-1, compared with zero in four Rbmxl2 -/- mice (Figure 2D and 2E). No effect on female fertility was observed in Rbmxl2 -/- mice (not shown).

Very rarely a few round spermatids were observed in adult Rbmxl2 -/- testis sections (Figure 2 –figure supplement 1A), although no elongated spermatids were detected. The presence of these round spermatids indicates that meiosis can occasionally complete in the absence of RBMXL2 protein. This is consistent with loss of RBMXL2 causing either (1) a developmental block in meiosis from which a few cells can escape; or alternatively (2) a slow attrition effect, in which the Rbmxl2 -/- testis phenotype is caused by post-meiotic stages dying in the adult testis. To differentiate between these two possibilities we histologically analysed testes at 21 days postpartum (21dpp), at which point germ cells in the wild type (Rbmxl2 +/-) mice have just started to enter the post-meiotic round spermatid stage. At 21dpp there were still significantly fewer round spermatids in the Rbmxl2 +/- sections compared to wild type (Figure 2 Supplement 1B), even though there would have been less time for round spermatid cell death compared with the adult. This result thus supports a strong meiotic block in Rbmxl2 +/- mice, with occasional completion of meiosis rather than a gradual attrition of round spermatids.
Most Rbmxl2 −/− germ cells do not progress past meiotic diplotene

The above data demonstrate that an Rbmxl2 gene, which is conserved in all placental mammals and specifically expressed in male meiosis, is essential for mouse spermatogenesis somewhere within meiotic prophase. Staining of nuclear spreads with antibodies specific to the meiotic chromosome proteins SYCP1 and SYCP3 more precisely showed that Rbmxl2 −/− mice arrest germ cell development during the diplotene substage of meiotic prophase (Figure 2F and Figure 2 –figure supplement 2). Metaphase I nuclei were only detected in wild type mice and not in Rbmxl2 −/− mice (N=3 wild type and N=3 Rbmxl2 −/− testes, 180 and 138 spermatocyte nuclei scored for wild type and Rbmxl2 −/− respectively). No significant differences in the frequency of earlier stages of meiotic prophase were detected between Rbmxl2 −/− and wild type testes. The presence of germ cell populations up to diplotene in Rbmxl2 −/− testes was further confirmed by analysis of histological sections stained with Periodic Acid Schiff (PAS) (Figure 2 –figure supplement 3A-D). Staining of adult testis sections with PAS or antibodies specific to the apoptotic marker activated Caspase 3 further showed that adult Rbmxl2 −/− germ cells die via apoptosis (Figure 2 –figure supplement 3E-F).

Analysis of diplotene nuclear spreads revealed further abnormalities in the Rbmxl2 −/− testis. There was a decreased number of H3K9me3-marked centromere clusters (Takada et al., 2011), with each individual cluster also containing more centromeres than in wild type testis (Figure 2 –figure supplement 4A-C). H3K9me3 staining was also present at the sex body in a proportion (82%) of wild type diplotene spermatocytes, but essentially absent in mutant diplotene spermatocytes (1.5% of diplotene spermatocytes stained, 66 Rbmxl2 −/− and 66 wild type nuclei scored, n=3 for both) (Figure 2 –figure supplement 4D). These defects in centromere clustering and H3K9me3 modification of the sex body were still detectable but less severe at pachytene (Figure 2 –figure supplement 4D). Asynapsis was rarely observed in either control or mutant pachytene stage spermatocytes, indicating that the defect causing diplotene arrest does not significantly impact either synaptonemal complex formation or meiotic homology searching (Figure 2 –figure supplement 4E) (N=3, 109 and 94 nuclei scored for wild type and Rbmxl2 −/− respectively).

In summary, the above mouse phenotype showed that deletion of the ancient RNA binding protein RBMXL2 induces progressive defects in male mouse meiotic prophase. These defects culminate with a major block during diplotene that prevents entry into metaphase I, but with defects becoming already apparent during pachytene.

Gene expression analysis of age matched wild type and Rbmxl2 −/− testes
Since RBMXL2 is a nuclear RNA binding protein we predicted that the above phenotype could be associated with a primary molecular defect in generating or processing RNAs in the Rbmxl2-/- testis. To test this we analysed wild type and Rbmxl2-/- testes using RNAseq. Based on the knockout phenotype above, and because the adult wild type testis contains additional more advanced germ cells that are missing from the adult Rbmxl2-/- testis, we analysed testes at 18 days post partum (18dpp) during the first synchronised wave of mouse spermatogenesis. Such 18dpp wild type mouse testes contain germ cells between spermatogonia all the way through to diplotene, with 60% of cells engaged in meiotic prophase, but no post-meiotic cells (Bellve et al., 1977).

Gene expression analysis of this RNAseq data (Anders and Huber, 2010) showed overall patterns of transcription were similar between wild type and Rbmxl2-/- 18dpp testes (Figure 3A, and Figure 3 Source Data 1). Only 45 genes showed a fold change greater or equal to two between the wild type and Rbmxl2-/- testes, with an adjusted p value of less than 0.05, and only 23 of these changes were for known protein coding genes (Figure 3 Source Data 1). The strongest difference in overall gene expression between the wild type and Rbmxl2-/- backgrounds was for the Rbmxl2 gene itself, as expected since the Rbmxl2 gene is deleted from the knockout mouse. We also detected strong expression changes within the Fsip2 gene, particularly for Fsip2 exon 16 and downstream exons that were expressed only in the wild type background (Figure 3 – figure supplement 1A and 1B). Mutations in Fsip2 correlate with defects in human sperm flagella – cellular structures which develop after meiosis (Martinez et al., 2018). RNAseq analysis detected more subtle expression changes in Cul4a, Slc9c1 and Tex15, each of which are required for mouse male fertility (Smith et al., 2018) (Figure 3 Source Data 1) (Kopanja et al., 2011). Genes encoding transcription factors (Myf6 and Nxn) and a signalling protein (Cyr61) that controls apoptosis (Jun and Lau, 2011) also changed expression. Mouse phenotype information at the Mouse Genome Database (MGD) (Smith et al., 2018) indicate that each of these latter 3 genes have important roles in normal development but not specifically of the testis (Figure 3 Source Data 2). Sixteen (36% of the total detected) gene expression changes in the Rbmxl2-/- testis were for predicted non-coding RNAs of unknown function (Figure 3 Source Data 1).

**RBMXL2 protein controls splicing patterns during meiosis**

We carried out further bioinformatic analysis to specifically search for mis-regulated splicing events in the Rbmxl2-/- testes (Vaquero-Garcia et al., 2016). A total of 237 high-confidence, mis-regulated local splicing variations were identified in 186 genes (Figure 3B, and Figure 3 Source Data 3). Using RT-PCR to distinguish splice isoforms 27 of these splicing changes were experimentally tested, validating 23/27 splice isoform switches (Figure 3 Source Data 4). Some genes had more than one splicing event controlled by RBMXL2 (e.g. the Catsperb
gene had 3 events). Gene Ontology (GO) analysis showed that a number of the genes regulated at the splicing level by RBMXL2 have established roles in spermatogenesis, meiosis and germ cell development (Figure 3 Source Data 5). However, amongst the complete set of regulated genes there was no significant enrichment of particular GO terms. This is consistent with RBMXL2 regulating splicing of a functionally diverse group of genes. Analysis of knockout phenotypes provided by the MGD (Smith et al., 2018) indicated that whole gene deletion of 25/186 RBMXL2-regulated genes cause male infertility. These latter target genes must thus have a key role in germ cell development (Figure 3C, Figure 3 – figure supplement 2, Figure 3 Source Data 6). Genetic deletion of some other RBMXL2 target genes cause either embryonic or neonatal lethality (33/186 genes), developmental (19/36 genes) or pleiotropic defects (43/186 genes). Cell type mis-splicing of these latter genes during meiosis could thus cause severe phenotypic effects on germ cell biology.

Fourteen (33%) of the genes originally identified to have changed overall expression levels between wild type and *Rbmxl2*+/− testes also changed splicing patterns (Figure 3C, and Figure 3 Source Data 1 and 3). These included splice variants in *Esco1* (encoding a protein involved in chromatid cohesion), *Slc39a8* gene (that encodes the transporter protein responsible for cadmium toxicity in the testis (Dalton et al., 2005); and the *Slc9c1* gene (annotated on the MGD as essential for male fertility (Smith et al., 2018)).

The RNA binding specificity of RBMXL2 protein was unknown. Thus we used high throughput sequencing cross linking immunoprecipitation (HITS-CLIP) to enable us to correlate splicing changes detected within the *Rbmxl2*+/− 18dpp mouse testis with global RBMXL2 protein-RNA interactions (Grellscheid et al., 2011). Antibodies specific to mouse RBMXL2 immuno-precipitated a radiolabelled RNA protein adduct of the known size of RBMXL2 protein (50KDa) after treatment with high concentrations of RNase (Figure 3 Supplement 3A) (Elliott et al., 2000b). Lower concentrations of RNase were used to retrieve an average tag length of 40 nucleotides. Enriched motif analysis of the sequenced RBMXL2 CLIP tags showed that each of the top 10 5-mers contained the dinucleotide AA (Figure 3D). Interestingly, AA is also the dinucleotide bound by RBMX (Moursy et al., 2014), which is consistent with over 90% shared sequence identity within the RRMs of these proteins (Figure 1 – figure supplement 1). These intragenic cross-linked sites mapped to the mouse genome most frequently within introns, consistent with RBMXL2 being involved in nuclear RNA processing events (Figure 3 Supplement 3B). Next we searched for CLIP tags mapping to regions near the splicing events controlled by RBMXL2, which would suggest direct regulation. We observed enrichment of RBMXL2 binding both within alternative exon sequences and in regions proximal to regulated splice sites compared to non-regulated
events (Figure 3 – figure supplement 4A). Consistent with these enriched binding occurrences, motif maps of the top pentamers identified by CLIP (Figure 3D) showed regions of enrichment proximal to regulated versus non-regulated splice junctions (Figure 3 – figure supplement 4B).

**RBMXL2 protein represses splicing of exons that would compromise meiotic gene expression**

Overall, the above bioinformatics analysis indicated an accumulation of defective splice isoform patterns in the *Rbmxl2^-/-* testis. MAJIQ captures local splicing variations (LSVs) that involve both known and un-annotated (*de-novo*) splice sites, junctions and exons. These LSVs can correspond to both classical binary splicing events and more complex events involving three or more junctions. Since classical binary events (e.g. skipped exons) are easier to inspect and visualise with the RNAseq reads on the UCSC genome browser, we focused on the binary events for further investigation. This showed that 60% of the 87 most easily visualised classical splicing events controlled by RBMXL2 involve the altered selection of *de-novo* splice sites (defined as not currently annotated on the most recent build of the *Mus musculus* genome GRCm38/mm10, Figure 3 Source Data 3). These *de-novo* splicing variations controlled by RBMXL2 are putatively cryptic events as they insert novel internal exons that were flanked by consensus GT-AG 5' and 3' splice sites. Moreover, these *de-novo* splice sites showed low expected percent spliced in (E(PSI)) values in wild type 18 dpp testes and across a panel of twelve mouse tissues, further suggesting they are cryptic events only included in the absence of RBMXL2 and not in wild type mice (Figure 4 – figure supplement 1). Detailed investigation indicated 84% of such cryptic exons were either not multiples of three or introduced stop codons into their mRNAs, meaning their insertion into mRNAs would disrupt protein reading frames and interfere with meiotic protein expression.

Splicing inclusion of cryptic terminal exons could also severely impact patterns of meiotic gene expression. We detected high amplitude splicing inclusion of a cryptic terminal exon within the 5' UTR of the *Kdm4d* gene (Figure 4A and 4B, note also increased splicing of an already annotated upstream *Kdm4d* alternative exon within the *Rbmxl2^-/-* testis). We confirmed splicing inclusion of this *Kdm4d* cryptic exon in the *Rbmxl2^-/-* testes using RT-PCR (primer positions are shown in Figure 4A). *Kdm4d* encodes a histone demethylase protein that is important for normal patterns of germ cell apoptosis in the testis (Iwamori et al., 2011). Exon 2 contains the entire CDS (Coding DNA Sequence, Figure 4A) of the *Kdm4d* gene. Analysis of exon junction read numbers using Sashimi plots confirmed different nuclear processing pathways are used for *Kdm4d* in wild type and knockout testes (Figure 4 – figure supplement 2A). There were 14-fold more exon junction reads connecting
Kdm4d exon 1 to the cryptic exon (140 reads), when compared to exon junction reads joining the cryptic exon to Kdm4d coding exon 2. This pattern is consistent with splicing inclusion of the cryptic exon being connected with use of an associated polyA site. Consistent with this, some individual RNAseq reads extended past the 5’ splice site of the Kdm4d cryptic exon and then terminated downstream following a consensus polyadenylation site (AATAAAA) sequence (Figure 4C). Both bioinformatics analysis (Figure 3 Source Data 1) and qPCR (Figure 4 –figure supplement 2B) detected reduced Kdm4d gene expression levels in each of 3 replicate Rbmxl2 −/− testes compared to their wild type equivalents, but these trends were not statistically significant due to biological variability between individual mice. We also detected increased selection in the Rbmxl2 −/− testes of another 17 splicing events that mapped to terminal exons (Figure 3 Source Data 3). These included an already annotated upstream terminal exon in the Lrcc1 gene (annotated on the MGD as essential for mouse fertility (Smith et al., 2018); and a terminal exon in the Slc39a8 gene (Figure 4 – figure Supplement 3A and B).

In addition to splicing of entire cryptic exons, loss of RBMXL2 protein also activated splicing selection of individual cryptic 5’ and 3’ splice sites within introns that increased the length of already annotated exons (Figure 3 Source Data 3). This kind of changed splicing pattern was observed for the Gm5134 mRNA that encodes a solute transporter important for normal metabolism (Figure 4D and 4E). The aberrant splice isoform of the Gm5134 transcript made in the Rbmxl2 −/− testis uses a cryptic 3’ splice site within intron 10, increasing the size of exon 11 by 1353nt.

**RBMXL2 protein is important to prevent mis-splicing of long exons in key genes during meiosis**

As well as cryptic splice sites within introns, some exon-located cryptic splice sites were activated within the Rbmxl2 −/− testes, including within genes important for testis development (Figure 3- source Data 6). Amongst the genes known to be important for meiosis, we observed a dip in RNAseq density for the interior of Meioc (meiosis specific gene with coiled coil domain) exon 5 in the Rbmxl2 −/− testis transcriptome (Figure 5A) (Abby et al., 2016, Soh et al., 2017). This dip in RNAseq read density was caused by high levels of cryptic splice site activation within Meioc exon 5. Sashimi plots showed that in the Rbmxl2 −/− testes ~28% of the splice junctions joined Meioc exon 4 to an exon-internal cryptic 3’ splice site within Meioc exon 5 (Figure 5 –figure supplement 1A). A further cryptic 5’ splice site within Meioc exon 5 was also used in 22% of the Rbmxl2 −/− testis transcripts. Use of both cryptic splice sites converted Meioc exon 5 into an extron (an exon with internal splice sites (Marquez et al.,
2015, Sibley et al., 2016). RT-PCR analysis confirmed both these defective Meioc exon 5 splicing patterns within the Rbmxl2 /− testes (Figure 5B-C).

The cryptic splice sites within Meioc exon 5 are hardly ever used in the wild type 18dpp testis transcriptome. *In silico* analysis showed that the internal cryptic 5′ splice site within Meioc exon 5 is weak compared to 5′ splice sites for other known alternative exons (7.5th percentile by SROOGLE, compared to 57th percentile for the downstream annotated 3′ splice site (Figure 5 –figure supplement 1B) (Schwartz et al., 2009). The first two nucleotide positions downstream of the Meioc exon 5- cryptic internal 5′ splice site are GC, so diverge from the GT consensus which is found flanking almost all eukaryotic introns. Similarly, the cryptic internal 3′ splice site within Meioc exon 5 is relatively weak (8th percentile of the average 3′ splice site strength for an alternative exon, compared to 41st percentile for the upstream annotated 3′ splice site). Meioc exon 5 is also directly bound by RBMXL2 protein, evidenced by three internal HITS-CLIP tags (Figure 5A). These include a CLIP tag directly overlapping the Meioc exon 5 internal cryptic 5′ splice site, as well as CLIP tags upstream of the internal 3′ splice site.

Meioc exon 5 is highly conserved across species suggesting it encodes a functionally important part of Meioc protein (Figure 5A). Exontrons frequently modify protein coding capacity of mRNAs by removing peptide coding segments from mRNAs (Marquez et al., 2015, Sibley et al., 2016). Both the cryptic splicing events in Meioc exon 5 remove RNA segments within the Meioc CDS that are multiples of 3. These events shorten the CDS just upstream of but largely not including the coding information for the MEIOC domain itself ([http://pfam.xfam.org/family/PF15189](http://pfam.xfam.org/family/PF15189)). We thus predicted that a shorter Meioc protein isoform may be produced in the Rbmxl2 /− testis. Consistent with this, Western blotting with a Meioc specific antiserum (Abby et al., 2016) detected a 40-50 KDa Meioc protein within the Rbmxl2 /− testis, but not in wild type or heterozygote Rbmxl2 /+ testes (Figure 5D). This 40-50 KDa molecular weight corresponds to an expected Meioc protein size following complete exitron removal. Although shorter Meioc protein isoforms were only detected in the Rbmxl2 /− testes, in each genotype we detected a protein doublet just above 100KDa corresponding to the reported size of full length Meioc protein (Abby et al., 2016) (Figure 5D).

We further analysed splicing patterns during the first wave of mouse spermatogenesis to test whether appearance of the Meioc splicing defects corresponded with the known expression window of RBMXL2 protein in meiotic prophase (Ehmann et al., 2008). Aberrant splice isoforms of Meioc exon 5 only appeared within Rbmxl2 /− testes from 14dpp, and were barely visible at any time point within similar age wild type testes. We performed similar experiments to analyse the dynamics of Brca2 splicing in the first wave of mouse spermatogenesis.
spermatogenesis. The RNAseq data predicted utilisation of an upstream cryptic 5′splice site within the 4809 nucleotide Brca2 exon 11 (Figure 3 Source Data 3) that removes 1263 nucleotides from Brca2 exon 11 in the Rbmxl2⁻/- testes (Figure 5 – figure supplement 2A). RT-PCR analysis during the first wave of spermatogenesis showed Brca2 exon 11 splicing was normal at 13dpp in the Rbmxl2⁻/- testes, with defects appearing only during and after 14dpp (Figure 5 – figure supplement 2B). These time-course data indicate that splicing defects in both Meioc and Brca2 appear and become progressively worse during meiotic prophase in the Rbmxl2⁻/- testes within genes that are expressed and processed normally in earlier germ cell types. An implication of this is that the aberrant splice isoforms detected in 18dpp whole testis RNA must be significantly diluted by signals from germ cells from earlier developmental stages.

Other unusually long exons in genes important for germ cell development were also mis-spliced in the Rbmxl2⁻/- testes including exons in the Alms1 and Esco1 genes (Figure 3 Source Data 3). The Alms1 gene encodes a protein involved in microtubule organisation during cell division, and its knockout disrupts spermatogenesis resulting in a small testis phenotype (Smith et al., 2018). In the Rbmxl2⁻/- testis an upstream 5′splice site is selected within Alms1 exon 5, which is is unusually long at 1546nt (Figure 5F). This splice switch removes 1407 nucleotides from the Alms1 mRNA, and coding information for 469 amino acids from the predicted Alms1 reading frame.

**Ectopic expression of RBMXL2 protein represses cryptic exon splicing in vitro**

The above data indicate that loss of RBMXL2 protein within spermatocytes activates utilisation of cryptic splice sites. We next tested if cryptic splice site selection would be reciprocally repressed by ectopic expression of RBMXL2 protein. The genomic region spanning the Kdm4d cryptic exon and its flanking intron sequences were cloned into a minigene (Figure 6A). We detected ~40% splicing inclusion of the cryptic Kdm4d exon after co-transfection of this minigene into HEK293 cells (that do not normally express RBMXL2) with an expression vector encoding GFP (Figure 6B lane 1). Levels of Kdm4d cryptic exon splicing inclusion were repressed 6-fold upon co-transfection of a construct encoding an RBMXL2-GFP fusion protein. We also tested the activity of an RBMXL2 fusion protein without the RNP1 motif of the RRM (RBMXL2 ΔRNP1). Consistent with RNA-protein contacts not being critical for splicing repression of this cryptic exon, RBMXL2 ΔRNP1 could still repress Kdm4d cryptic exon splicing, although slightly less efficiently than the full length RBMXL2 fusion protein. Interestingly, splicing of the Kdm4d cryptic exon was also efficiently silenced by ectopic co-expression of an RBMX-GFP fusion protein (Figure 6B lane 7).
The above results show that RBMXL2 represses \textit{Kdm4d} cryptic exon splicing, and is consistent with RBMXL2 providing a direct functional replacement for RBMX during meiosis. iCLIP experiments in adult mouse testis further identified the \textit{Kdm4d} cryptic exon as directly binding the splicing activator protein Tra2β (Dalgliesh and Elliott, unpublished data). Consistent with this, Tra2β over-expression in HEK293 cells strongly increased \textit{Kdm4d} cryptic exon splicing inclusion (Figure 6B lane 4). Ectopic activation by Tra2β indicated that although this \textit{Kdm4d} exon cryptic exon is effectively ignored by the spliceosome in wild type testis, cryptic exon splicing is activated in response to increased expression of a splicing activator protein. Importantly, splicing activation by Tra2β was efficiently suppressed when the minigene was co-transfected at the same time with both RBMXL2 and Tra2β (Figure 6B, lanes 5 and 6).

We addressed if other cryptic exons repressed by RBMXL2 might also possess the capacity to be activated by splicing regulators that bind to their associated splicing enhancer sequences. \textit{In silico} analysis (Schwartz et al., 2009) of a panel of cryptic exons that were activated in the \textit{Rbmxl2\textsuperscript{-/-}} testes showed that these exons had significantly weaker 3′ splice sites, but similar Exonic Splicing Enhancer (ESE) sequence content compared to their immediately adjacent downstream exons (Figure 6C). These other cryptic exons that are repressed by RBMXL2 may be likewise poised for splicing during meiosis, because they bind to splicing activator proteins such as Tra2β.

\textbf{Discussion}

This study finds that loss of the ancient RBMXL2 protein completely prevents sperm production in the mouse, largely because of a developmental block during meiosis. \textit{Rbmxl2\textsuperscript{-/-}} mouse germ cells develop as far as meiosis, thus escaping the major developmental checkpoints that operate during meiotic pachytene (de Rooij and de Boer, 2003), but then undergo cell death by apoptosis after reaching diplotene. Our data further show that RBMXL2 protein is important to protect the meiotic transcriptome from a spectrum of splicing defects (Figure 7A). These include classic manifestations of cryptic splicing that would be strongly deleterious to gene expression (Marquez et al., 2015, Sibley et al., 2016), including insertion of novel cryptic exons and premature polyadenylation sites that would disrupt protein coding sequences, and the selection of cryptic splice sites that would shorten or extend exon lengths. Previous work reporting poor conservation of splicing patterns between species suggested that much alternative splicing might not control fundamental aspects of germ cell biology (Kan et al., 2005). In contrast, the work presented here suggest a generic requirement for cryptic splice site suppression during meiosis that could explain the
conservation of an Rbmxl2 gene in all placental mammals, even if individual cryptic splice sites are not conserved between species’ genomes (Elliott et al., 2000b).

Cells in meiotic prophase might be particularly susceptible to cryptic splice site poisoning because of their high levels of transcription, altered splicing regulator expression and relaxed chromatin environments (de la Grange et al., 2010, Licatalosi, 2016, Soumillon et al., 2013, Pan et al., 2008, Wang et al., 2008, Clark et al., 2007, Grosso et al., 2008, Yeo et al., 2004). In silico data suggest that the cryptic exons repressed by RBMXL2 would be poised for splicing selection in conditions of high concentrations of splicing activators because of their strong ESE contents. We suggest a model where RBMXL2 protein may buffer the activity of splicing activator proteins to prevent this cryptic splice site selection from happening in meiosis (Figure 7B). This model is also consistent with previously reported data showing that RBMXL2, RBMX and RBMY proteins physically interact with and antagonise the splicing activity of Tra2β and SR proteins (Liu et al., 2009, Nasim et al., 2003, Venables et al., 2000, Elliott et al., 2000a) (Figure 7B). Data presented in this study also show that Kdm4d cryptic exon splicing activation by Tra2β is blocked in vitro by co-expression of RBMXL2. As an alternative and not mutually exclusive model, RBMXL2 protein-RNA binding may sterically block recognition of some cryptic splice sites during meiosis, including the cryptic 5′ splice site within Meioc exon 5 which was utilised in the Rbmxl2−/− testis (Figures 5A and 7B). However, no HITS-CLIP tags for RBMXL2 were detected near the Kdm4d cryptic exon. Furthermore, a version of RBMXL2 deleted for the key RNP1 motif involved in RNA-protein interactions could still repress Kdm4d cryptic exon splicing (albeit slightly less efficiently than the wild type RBMXL2 protein). Thus buffering protein interactions between RBMXL2 and proteins like Tra2β may be potentially more important in repressing cryptic splicing in meiosis than direct RBMXL2 RNA-protein interactions.

Loss of RBMXL2 disrupts expression of many downstream target genes in parallel, both at splicing and transcriptional levels. This makes it difficult to resolve whether the phenotype detected in the Rbmxl2−/− mouse testes is caused by aberrant expression of a single target RNA, or is a compound phenotype involving multiple genes. Some genes affected by RBMXL2 protein deletion are annotated as important for spermatogenesis on the MGD (Smith et al., 2018). These include Alms1 (Arsov et al., 2006, Collin et al., 2005); Brca2 which is required for meiotic prophase (Sharan et al., 2004); Kdm4d, which helps control apoptosis during male germ cell development (Iwamori et al., 2011); and Cul4a that encodes an important ubiquitin ligase needed for DNA damage response. Two factors complicate direct comparison of MGD phenotypes with those after RBMXL2 knockout. Firstly, conventional knockouts will report a phenotype the first time a gene is needed in a developmental pathway. In contrast RBMXL2 is only expressed from the onset of meiosis,
so any defects in target gene expression that occur without RBMXL2 could result in a later phenotype. For example, Meioc knockout phenotype analysis has concentrated on entry into meiosis (Abby et al., 2016, Soh et al., 2017). Our data show that Meioc exon 5 is still spliced normally on meiotic entry in the Rbmxl2-/- mouse (at 12dpp), but becomes compromised later on during meiotic prophase (when RBMXL2 is expressed). It is possible that Rbmx and Rbmy will provide an equivalent function to Rbmxl2 in spermatogonia and early spermatocytes, and that it is only when meiotic sex chromosome inactivation initiates during pachytene that germ cells will depend on Rbmxl2. Consistent with this, in transfected cells RBMX was able to suppress a cryptic splice exon in the Kdm4d with similar efficiency to the RBMXL2 protein. Secondly, the MGD phenotype data (Smith et al., 2018) is not as detailed as that we have performed for RBMXL2.

High levels of gene expression and alternative splicing have also been observed in the nervous system, and this is another anatomic site where pathological cryptic exon inclusion has been reported. Neuronal cryptic exon inclusion occurs following depletion of the splicing repressor protein TDP43, and might contribute to neuron death in neurological diseases like ALS and Alzheimer’s disease (Sun et al., 2017). Cryptic splicing has also been observed in cultured human cells after depletion of the splicing repressor proteins hnRNP C, hnRNPL and PTB (Ling et al., 2016, McClory et al., 2018, Zarnack et al., 2013), and in mouse oocytes depleted for the splicing factor SRSF3 (Do et al., 2018). Taking all these results into consideration we suggest that RBMXL2 has a key role controlling the meiotic transcriptome, and suggest that human male infertility caused by loss of RBMXL2 or its paralog RBMY may be associated with germ cell type-specific cryptic splicing.

Acknowledgements

This work was funded by the BBSRC (grant BB/P006612/1, awarded to DJE). JC and IRA were funded by the Medical Research Council (http://www.mrc.ac.uk/) through an intramural programme grant award to IRA (MC_PC_U127580973). Work by MRG and YB was supported by R01 AG AG046544 to YB. We thank Laura Bellutti and Gabriel Livera for kindly supplying the Meioc antiserum. We are very grateful to Donal O’Carroll (Edinburgh University), Nick Europe Finner, Julian Venables and Louise Reynard for comments on the manuscript.

Declaration of interests

The authors declare no competing interests

Figure Legends
Figure 1. Creation of a mouse model which does not express the testis-specific RNA binding protein RBMXL2.

(A) The RBMXL2 gene evolved via retrotransposition of RBMX early in mammalian evolution. The cladogram shows three members of the gene family that derived from an ancestral RBMX gene, summarising their global functions, expression, and RNA binding sites of their encoded proteins.

(B) Modular structure of the RBMXL2 protein, showing the known protein interaction with Tra2β.

(C) Western blot confirming no RBMXL2 protein is made within Rbmxl2 −/− adult testis, compared to wild type (Rbmxl2 +/+) and heterozygous (Rbmxl2 +/-) adult testes.

Figure 2. RBMXL2 protein is important for mouse germ cells to progress past diplotene into metaphase I of meiosis.

(A) Testis morphologies of adult wild type and homozygous knockout (Rbmxl2 −/−) mice (scale above testes shown in millimeters).

(B) Testis:body weight ratios of different genotype mice. P values were calculated using a t test.

(C) Micrographs of haematoxylin-stained testis sections from wild type and Rbmxl2 −/− mice. Abbreviations: Spg, spermatogonia; Spc, spermatocyte; Rtd, round spermatid; Spd, elongating spermatid. Scale bar=20um.

(D) Epididymis of wild type and Rbmxl2 −/− mice stained with haemotoxylin and eosin. Asterisk indicates lumen (note lumen is empty in Rbmxl2 −/− section indicative of azoospermia, but contains multiple cells in the wild type). Scale bar=20um.

(E) Epididymal sperm counts of wild type and Rbmxl2 −/− mice (n=4 of each genotype). The error bar represents SEM.

(F) Meiotic prophase I stages detected in wild type and Rbmxl2 −/− testis chromosome spreads stained for SYCP3 (pseudocoloured red) and SYCP1 (pseudocoloured green); and DAPI and SYCP3 (pseudocoloured blue and red respectively). No metaphase I nuclei were identified in the Rbmxl2 −/− testes. Quantitative data are provided in Figure 2 Supplement 2.

Figure 3. RBMXL2 protein expression controls splicing patterns of important genes during meiosis
(A) MAplot showing gene expression levels in 18dpp testis transcriptomes and how they change between wild type and \textit{Rbmxl2} \textsuperscript{−/−} mice. Genes with more than a 2-fold change in gene expression, and an adjusted p value of less than 0.05 are shown as red dots. All other genes are shown as black dots. This data given in full within Figure 3 Source Data 1.

(B) Scatterplot showing splicing changes detected between the wild-type and \textit{Rbmxl2} \textsuperscript{−/−} testis detected by RNA-seq. The scatterplot shows expected percentage in (E(PSI)) for wild-type and \textit{Rbmxl2} \textsuperscript{−/−} testis detected by RNA-seq using the bioinformatics programme MAJIQ (Vaquero-Garcia et al., 2016). Splicing changes in some genes are named and arrowed. High-confidence splicing changes (P(|ΔPSI| > V) >95%) are marked in dark red for a predicted change of V = 20%, and dark yellow for a predicted change of V = 10%, and are also listed in Figure 3 Source Data 2. All other quantified splice changes are in yellow (21,280 events examined). Dashed lines indicate ΔPSI of +10%.

(C) Left pie chart: Proportions of genes identified by RNAseq analysis to have defects at the splicing and transcriptional levels. Right pie chart: Proportions of phenotypes reported by whole gene knockout annotated for genes detected to have splicing defects in the absence of RBMXL2 protein (Smith et al., 2018) (see also Figure 3 Source Data 6).

(D) The top 10 most frequently recovered pentamers after HITS-CLIP for RBMXL2 in the adult mouse testis (AA dinucleotides are shown in red).

\textbf{Figure 4 RBMXL2 protein represses splicing inclusion of exons that would compromise gene expression patterns}

(A) Screenshot from the UCSC genome browser (Karolchik et al., 2014) showing high amplitude splicing inclusion levels for a novel cryptic exon in the \textit{Kdm4d} gene in the \textit{Rbmxl2} \textsuperscript{−/−} testis transcriptome.

(B) Experimental confirmation of \textit{Kdm4d} splicing changes using RT-PCR and capillary gel electrophoresis (lower panel), and data from 3 \textit{Rbmxl2} \textsuperscript{−/−} and 3 wild type 18dpp mouse testes. Individual samples shown in the capillary gel electrophoretogram are indicated as red dots. The mean PSI is shown as a grey bar with SEM as error bar. The p value was calculated using a t test.

(C) Higher resolution screenshot (Karolchik et al., 2014) of the \textit{Rbmxl2} \textsuperscript{−/−} testis transcriptome showing RNAseq read density in the genomic region of the \textit{Kdm4d} cryptic exon. The position of the cryptic 5’ splice site is indicated by a red arrow. Splice junction sequences are shown underneath with the consensus polyadenylation site AATAAA in red. Some RNAseq
reads terminate downstream of the Kdm4d cryptic exon 5′ splice site. The final RNAseq read coverage detected downstream of this exon is shown as a red asterix.

(D) The left panel is a UCSC genome browser screenshot (Karolchik et al., 2014) showing increased splicing selection of an cryptic 3′ splice site within the Gm5134 mRNA in the Rbmxl2 −/− testis transcriptome.

(E) Chart showing experimental confirmation of the splicing switch in the Gm5134 gene in 3 Rbmxl2 −/− and 3 wild type 18dpp mouse testes. The individual samples shown in the capillary gel electrophoretogram are shown in the chart as red dots.

Figure 5. RBMXL2 protein is important to prevent splicing of long exons becoming disrupted during meiosis

(A) Screenshot from the UCSC genome browser (Karolchik et al., 2014) showing splicing pattern of Meioc exon 5 in wild type and Rbmxl2 −/− testis. RNAseq read density from Rbmxl2 −/− testis are shown in blue; RNAseq read density from wild type testis in red; the genomic positions of HITS-CLIP tags are mapped underneath. Splice sites for exon 5 are arrowed, and the binding sites for RT-PCR primers are shown.

(B) Representative capillary gel electrophoretogram showing RT-PCR validation of the changing Meioc splicing pattern. This RT-PCR analysis used three primers in multiplex (MeiocF-MeiocR1 detects splicing of exon 4 to exon 5, and MeiocF-MeiocR2 detects splicing from exon 4 to the cryptic 3′ splice site within exon 5, or splicing of exon 4 to exon 5 after complete exitron removal. Detection of full length Meioc exon 5 inclusion by this latter primer combination is precluded by the large size of the amplicon).

(C) Chart showing percentage splicing inclusion (PSI) of different splice variants of Meioc exon 5. PSI data was generated by RT-PCR from 3 wild type (WT) and 3 Rbmxl2 −/− mice, with the individual samples shown in the capillary gel electrophoretogram in part (B) indicated as red dots. Error bars represent the SEM, and p values were calculated using t tests.

(D) Western blot of 18dpp mouse testes protein probed with an antibody specific to Meioc protein (Abby et al., 2016). A non-specific protein that was also detected in mouse spleen as well as 18dpp testis is indicated by an asterisk.
(E) Capillary gel electrophoretogram showing patterns of Meioc exon 5 splicing between 10dpp and 21dpp in the first wave of mouse spermatogenesis.

(F) UCSC screenshot (Karolchik et al., 2014) showing that splicing of Alms1 exon 5 is disrupted in Rbmxl2−/− testes via selection of an internal 5′ splice site.

**Figure 6. Ectopic expression of RBMXL2 protein represses cryptic exon splicing in vitro**

(A) Schematic of the Kdm4d minigene, in which the Kdm4d cryptic exon and flanking introns (shown in blue) are inserted between two beta globin exons in the vector pXJ41 (Bourgeois et al., 1999).

(B) Bar chart (upper panel) showing data from biological triplicate experiments after quantitation using capillary gel electrophoresis (n=3 for each transfection). Capillary gel electrophoretogram (lower panel) showing splicing patterns in HEK293 cells. The minigene containing the Kdm4d cryptic exon and flanking intron sequences were transfected with expression plasmids encoding either GFP, RBMXL2, RBMXL2 ΔRNP1, Tra2β, or combinations of these. Splicing patterns were analysed using RT-PCR.

(C) Bar chart showing average annotated and cryptic splice site strengths from a panel of internal cryptic exons that were activated in Rbmxl2−/− testes. Splice site strengths were quantitated as a percentile of the average value for an alternative exon using Sroogle (Schwartz et al., 2009). Open bars show percentile values for cryptic exons, shaded bars (black or red) show percentile values for downstream exons. Error bars represent the SEM, and p values were calculated using t tests.

**Figure 7. RBMXL2 protein is needed to enable production of proper splice isoform patterns during meiosis**

(A) Stacked bar chart summarising percentage types of splicing event mis-regulated in the Rbmxl2−/− testes. Details of the individual genes contributing to this list are given in Supplementary Data 3.

(B) Models for RBMXL2 function. In wild type mice, RBMXL2 protein may block cryptic exon silencing through: 1. Inhibition of splicing regulator proteins including Tra2β protein via protein-protein interactions (Venables et al., 2000); or 2. Direct RNA protein interactions of RBMXL2 with important splicing sequences in the pre-mRNA. In the Rbmxl2−/− background the absence of this buffering would result in an increase in splicing inclusion of some cryptic exons which contain Tra2β binding sites.
Supplementary Figure Legends

Figure 1 – figure supplement 1. Alignment of RBMXL2 with RBMX and RBMY

(A) Visualization using MView (Brown et al., 1998) of Clustal Omega (version 1.2.4, (Sievers et al., 2011)) multiple sequence alignment of amino acid sequences of RBMXL2 (sequence 1, Uniprot ID O75526), RBMX (sequence 2, Uniprot ID P38159), and RBMY (sequence 3, Uniprot ID P0DJD3) highlighting overall identity to RBMXL2. The region of the alignment corresponding to the highly similar RRM (93.7% identity to RBMX and 77.2% identity to RBMY) is labelled.

Figure 1 – figure supplement 2. Genomic organisation of different genotype mice

(A–C) Genomic organisations of the different mouse genotypes were confirmed using Southern blotting (D).

Figure 2 – figure supplement 1. Germ cells occasionally progress through meiosis into the round spermatid stage

(A) Occasional round spermatids were identified in the Rbmxl2–/– background. Rbmxl2–/– testis section stained with Periodic Acid Schiff (PAS). Scale bar=20um.

(B) Numbers of round spermatids detected per tubule in wild type and Rbmxl2–/– testes at 21 dpp (n=20 tubules counted at random from either genotype). Error bars represent the SEM, and p value was calculated using a t test.

Figure 2 – figure supplement 2. Graphical analysis of meiotic chromosome spreads.

Percentage of SYCP3-positive meiosis I sub-stages that were detected within wild type and Rbmxl2–/– chromosome spreads. Data counted from N=179 nuclei from 3 individual wild type adult testes; and N=138 nuclei from 3 individual homozygous Rbmxl2–/– testes.

Figure 2 – figure supplement 3. Detailed histological analysis of Rbmxl2–/– testes

(A) Section of Rbmxl2–/– testis including a stage IV Rbmxl2–/– seminiferous tubule. This stage IV tubule has no identifiable problems, and is identified by presence of dividing Ln spermatogonia and newly born B spermatogonia. (B) Stage VII/VIII tubule from Rbmxl2–/–
mouse. In this tubule the pachytene cells still look normal, recognised by the presence of preleptotene spermatocytes (abbreviated Spc). (C) Stage IX-XI tubule from Rbmxl2−/− mouse showing zygotene cells, but rather few late pachytenes/diplotenes. These latter cells are disappearing because of sloughing off through the lumen and/or apoptosis. (D) Higher magnification image of stage IX-XI tubule from Rbmxl2−/− mouse, showing many apoptotic pachytene/diplotene cells, some of which are indicated by arrows. Wild type and Rbmxl2−/− testis sections were stained with Periodic Acid Schiff (PAS). Abbreviations: Spg, spermatogonia; Spc, spermatocyte.

(E-F) Caspase 3 expression detected by immunohistochemistry on (E) wild type and (F) Rbmxl2−/− testes sections. Scale bar=20um.

**Figure 2 – figure supplement 4. Organisation of H3K9me3 staining and centromere clusters within wild type and Rbmxl2−/− mice show defects arise during pachytene but become progressively works by diplotene.**

(A) Top panel: images for SCYP3 (red), anti-centromeric antibody (ACA, green), and H3K9me3 (blue). Bottom panel: greyscale images of the H3K9me3 channel alone, enabling the domains to be seen more clearly, and the XY body staining recognised (this is fainter than the staining of pericentromeric heterochromatin). Scale bar=10um.

(B-C) Box and Whisker plots of (B) heterochromatic domain frequency and (C) centromere clustering from wild type and Rbmxl2−/− mice (pachytene analysis based on 73 nuclei scored from n=3 wild type mice; and 65 nuclei scored from n=3 Rbmxl2−/− mice. Diplotene analysis based on 66 nuclei scored from n=3 wild type mice, and 66 nuclei from n=3 Rbmxl2−/− mice).

(D) Percentage of sex bodies labelled with H3K9me3 (error bar is the standard error of the mean, and statistical significance was calculated using a t test).

(E) Percentage asynapsis measured in wild type (n=108) and Rbmxl2−/− chromosome spreads (n=92). Pachytene spermatocytes were judged to be those with at least one pair of fully synapsed autosomes, recognised by complete co-localisation of the axial element and transverse filament proteins SYCP3 and SYCP1. Asynapsed pachytene nuclei contain axial elements lacking transverse filament staining co-localisation in combination with fully synapsed autosomes.
Figure 3 – figure supplement 1. Differential expression of the *Fsip2* gene in the *Rbmxl2* \(-/-\) testis.

(A) RNAseq density mapping to the *Fsip2* gene, visualised on the UCSC genome browser (Karolchik et al., 2014).

(B) Quantitation of fold change from the 3' end of the *Fsip2* gene using qPCR (n=3).

Figure 3 – Figure Supplement 2. List of RBMXL2 target genes that are already implicated in mouse development.

Summary of GO terms and MGI phenotypes of genes that are regulated at the splicing level by RBMXL2 (compiled from Figure 3 Source Data 5 and 6).

Figure 3 – figure supplement 3. HITS-CLIP analysis of RBMXL2 in the adult mouse testis.

(A) Autoradiograph showing SDS-PAGE gel of 32P-labelled RNA crosslinked to endogenous RBMXL2 protein from mouse testis during a HITS-CLIP experiment. The area indicated by the red box was excised for further analysis.

(B) Genome wide distribution of RBMXL2 HITS-CLIP tags.

Figure 3 – figure supplement 4. RBMXL2 CLIP and motif enrichment around regulated splicing events

(A) Barchart showing the fraction of RBMXL2 repressed events (red, n=64); enhanced events (green, n = 22); or nonchanging events (grey, n = 192) that contain RBMXL2 HITS-CLIP tags proximal to alternatively splice regions (intronic regions within 300 nt up or downstream of a regulated junction, within the alternative exonic region, or the union of any of these regions). Significance was assessed by comparing regulated to non-changing events using a two-tailed Fisher’s exact test.

(B) Motif map showing the per nucleotide frequency the top RBMXL2 HITS-CLIP defined pentamers (Figure 3D) around regulated and nonchanging events (colours as in (A)). A sliding mean of 20 nts was applied for smoothing. Dashed lines indicate the 3'ss and 5'ss with up to 50 nucleotides of the exonic regions shown.

Figure 4 – figure supplement 1.

(A) Boxplots showing the distribution of E(PSI) values for the putative cryptic splice junctions identified in 18 dpp *Rbmxl2* \(-/-\) KO testis across a panel of mouse tissue samples. The first twelve mouse tissues contain eight samples each (Zhang et al., 2014) and the testes samples are from this study. Not all events were deemed quantifiable in these tissues, likely
due to tissue-restricted gene expression. The following number of junctions were assessed in non-testes tissues: muscle 12; heart 12; lung 21; white adipose 14; adrenal 12; aorta 11; brown adipose 18; liver 11; kidney 14; brainstem 8; hypothalamus 18; cerebellum 20.

**Figure 4 – figure supplement 2.**

(A) Sashimi plots of $Kdm4d$ gene in the $Rbmxl2^{-/-}$ and $Rbmxl2^{+/+}$ mice.

(B) qPCR analysis of $Kdm4d$ exon 2 expression levels in $Rbmxl2^{-/-}$ and $Rbmxl2^{+/+}$ testes.

**Figure 4 – figure supplement 3.**

(A) UCSC screenshot (Karolchik et al., 2014) showing an annotated terminal exon within the $Lrrcc1$ gene that is activated in the $Rbmxl2^{-/-}$ testis (notice decreased utilisation of downstream terminal exon in the wild type mouse).

(B) UCSC screenshot (Karolchik et al., 2014) showing selection of an annotated terminal exon in the $Slc39a8$ gene in representative wild type and $Rbmxl2^{-/-}$ testes, use of which is predicted to disrupt a known Pfam domain.

**Figure 5 Supplement 1**

(A) Sashimi plot showing exon junction reads across the $Meioc$ gene.

(B) Strengths of splice sites associated with $Meioc$ exon 5. The actual splice sites are indicated in the sequences by downward facing arrows. Average annotated and cryptic splice site strengths were calculated using Sroogle (Schwartz et al., 2009).

**Figure 5 Supplement 2**

(A) Sashimi plot showing exon junction reads across the $Brca2$ gene.

(B) Capillary gel electrophoretogram showing patterns of $Brca2$ splicing between 10dpp and 21dpp in the first wave of mouse spermatogenesis.
Figure 3 Source Data 1. Complete list of gene expression changes between the 18dpp testes of wild type and \textit{Rbmxl2} \textsuperscript{-/-} testes detected by DESeq.

Figure 3 Source Data 2. MGI mouse phenotypes reported by the Mouse Genome Database (Smith et al., 2018) after whole gene knockout for RBMXL2 transcription targets.

Figure 3 Source Data 3. Complete list of 237 high-confidence, mis-regulated local splicing variations in 186 genes between wild type and \textit{Rbmxl2} \textsuperscript{-/-} testes detected by Majiq (Vaquero-Garcia et al., 2016).

Figure 3 Source Data 4. PCR primers used to analyse splicing events by RT-PCR that change between wild type and \textit{Rbmxl2} \textsuperscript{-/-} testes. Some genes have more than one regulated event. The coloured cells represent events in a gene also occurring in duplicate on the list.

Figure 3 Source Data 5. Gene Ontology (GO) terms for RBMXL2 splicing targets identified. GO terms were obtained from the Mouse Genome Database (Smith et al., 2018).

Figure 3 Source Data 6. MGI mouse phenotypes reported after whole gene knockout for RBMXL2 splicing targets. Phenotypes were obtained from the Mouse Genome Database (Smith et al., 2018).

References


ncb2426 [pii]


ddn179 [pii]


PGENETICS-D-11-00956 [pii]


nrg3052 [pii]


gkt1168 [pii]


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 

LIU, N., ZHOU, K. I., PARISIEN, M., DAI, Q., DIATCENKO, L. & PAN, T. 2017. N6-
methyladenosine alters RNA structure to regulate binding of a low-complexity protein. 
Nucleic Acids Res, 45, 6051-6063. doi:10.1093/nar/gkx141

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-
doi:10.1371/journal.pgen.1000707

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, 
Family with Rna-Binding Protein Homology - Candidates for the Azoospermia Factor 
A zf Controlling Human Spermatogenesis. Cell, 75, 1287-1295. doi:Doi 10.1016/0092-
8674(93)90616-X

Unmasking alternative splicing inside protein-coding exons defines exitrons and their 

MARTINEZ, G., KHERRAF, Z. E., ZOUARI, R., FOURATI BEN MUSTAPHA, S., SAUT, A., 
PERNET-GALLAY, K., BERTRAND, A., BIDART, M., HOGRAINDEUR, J. P., 
AMIRI-YEKTA, A., KAROUF, M., KARAOUZENE, T., THIERRY-MIEG, N., 
DACHEUX-DESCAMPS, D., SATRE, V., BONHIVERS, M., TOURE, A., 
identifies mutations in FSIP2 as a recurrent cause of multiple morphological 
doi:10.1093/humrep/dey264

MATSUNAGA, S., TAKATA, H., MORIMOTO, A., HAYASHIHARA, K., HIGASHI, T., 
AKATSUCHI, K., MIZUSAWA, E., YAMAKAWA, M., ASHIDA, M., MATSUNAGA, T. 
maintenance and centromeric protection of sister chromatid cohesion. Cell Rep, 1, 
299-308. doi:10.1016/j.celrep.2012.02.005

RNA, 24, 761-768. doi:10.1261/rna.065508.117

mode of hnRNP G extends its role in SMN2 splicing regulation. Nucleic Acids Res, 
42, 6659-72. doi:10.1093/nar/gku244

gku244 [pii]

NASIM, M. T., CHERNOVA, T. K., CHOWDHURY, H. M., YUE, B. G. & EPERON, I. C. 
2003. HnRNP G and Tra2beta: opposite effects on splicing matched by antagonism 

NISHIYAMA, T., LADURNER, R., SCHMITZ, J., KREIDL, E., SCHLEIFFER, A., 
BHASKARA, V., BANDO, M., SHIRAHIGE, K., HYMAN, A. A., MECHTLER, K. & 
PETERS, J. M. 2010. Sororin mediates sister chromatid cohesion by antagonizing 

alternative splicing complexity in the human transcriptome by high-throughput 
sequencing. Nat Genet, 40, 1413-5. doi:10.1038/ng.259

g.259 [pii]

SCHMID, R., GRELLSCHEID, S. N., EHRMANN, I., DALGLIESH, C., DANILENKO, M., 
PARONETTO, M. P., PEDROTTI, S., GRELLSCHEID, D., DIXON, R. J., SETTE, C., 
EPERON, I. C. & ELLIOTT, D. J. 2013. The splicing landscape is globally

gkt811 [pii]


e11752 [pii]

10.1093/hmg/ddi233


nature07509 [pii]


gah042 [pii]


<table>
<thead>
<tr>
<th>Reagent type (species) or resource</th>
<th>Designation</th>
<th>Source or reference</th>
<th>Identifiers</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene (M. musculus)</td>
<td>Rbmxl2</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>strain, strain background (M. musculus)</td>
<td>C57Bl/6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>genetic reagent (M. musculus)</td>
<td>Rbmxl2</td>
<td>this publication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cell line (H. sapiens)</td>
<td>HEK293</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>antibody</td>
<td>Mouse monoclonal anti-tubulin</td>
<td>Sigma</td>
<td>catalog number T6793</td>
<td>(1:1000)</td>
</tr>
<tr>
<td>antibody</td>
<td>Rabbit polyclonal anti-GAPDH</td>
<td>Abgent, San Diego CA</td>
<td>catalog number AP7873b</td>
<td>(1:1000)</td>
</tr>
<tr>
<td>antibody</td>
<td>Rabbit polyclonal anti-caspase 3</td>
<td>Proteintech</td>
<td>catalog number 25546-1-AP</td>
<td>(1:100)</td>
</tr>
<tr>
<td>antibody</td>
<td>Rabbit polyclonal anti-Sycp3</td>
<td>Life span</td>
<td>Cat# LS-B175, RRID:AB_2197351; lot#73619</td>
<td>(1:500)</td>
</tr>
<tr>
<td>antibody</td>
<td>Mouse monoclonal anti-Sycp3 [Cor 10G11/7]</td>
<td>Abcam</td>
<td>Cat#ab97672 RRID:AB_10678841; lot#GR282926-1</td>
<td>(1:500)</td>
</tr>
<tr>
<td>Antibody</td>
<td>Manufacturer</td>
<td>Catalog Number</td>
<td>RRID</td>
<td>Dilution</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------</td>
<td>----------------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>Guinea Pig polyclonal anti-Sycp1</td>
<td>Laboratory of Howard Cooke</td>
<td>NA</td>
<td>(1:200)</td>
<td></td>
</tr>
<tr>
<td>Mouse monoclonal anti-phospho H2AX JBW301</td>
<td>Millipore</td>
<td>Cat# 05-636, RRID:AB_309864; lot#1997719</td>
<td>(1:3000)</td>
<td></td>
</tr>
<tr>
<td>Rabbit polyclonal anti-H3 trimethyl K9</td>
<td>Abcam</td>
<td>Cat# ab8898, RRID:AB_306848; lot#GR250850-2</td>
<td>(1:500)</td>
<td></td>
</tr>
<tr>
<td>Rabbit polyclonal anti-H3 phosphorylation (Ser10)</td>
<td>Abcam</td>
<td>Cat# ab14955, RRID:AB_443110; lot#GR192274-8</td>
<td>(1:1000)</td>
<td></td>
</tr>
<tr>
<td>Human polyclonal anti-Centromere Protein</td>
<td>Antibodies Incorporated</td>
<td>Cat#15-235-0001</td>
<td>(1:50)</td>
<td></td>
</tr>
<tr>
<td>Mouse monoclonal anti-AcSmc3</td>
<td>Katsuhiko Shirahige, University of Tokyo. Nishiyama et al., (2010) Cell, 143, 737-49</td>
<td>N/A</td>
<td>(1:1000)</td>
<td></td>
</tr>
<tr>
<td>Goat polyclonal anti-Mouse Alexa Fluor 488</td>
<td>Invitrogen</td>
<td>Cat# A11001</td>
<td>(1:500)</td>
<td></td>
</tr>
<tr>
<td>Goat polyclonal anti-Rabbit Alexa Fluor 488</td>
<td>Invitrogen</td>
<td>Cat# A11008</td>
<td>(1:500)</td>
<td></td>
</tr>
<tr>
<td>Goat polyclonal anti-Human Alexa Fluor 488</td>
<td>Invitrogen</td>
<td>Cat# A11013</td>
<td>(1:500)</td>
<td></td>
</tr>
<tr>
<td>Goat polyclonal anti-Guinea Pig Alexa Fluor 488</td>
<td>Invitrogen</td>
<td>Cat# A11073</td>
<td>(1:500)</td>
<td></td>
</tr>
<tr>
<td>Goat polyclonal anti-Mouse Alexa Fluor 594</td>
<td>Invitrogen</td>
<td>Cat# A11005</td>
<td>(1:500)</td>
<td></td>
</tr>
<tr>
<td>Goat polyclonal anti-Rabbit Alexa Fluor 594</td>
<td>Invitrogen</td>
<td>Cat# A11012</td>
<td>(1:500)</td>
<td></td>
</tr>
<tr>
<td>Goat polyclonal anti-Rabbit Alexa Fluor 647</td>
<td>Invitrogen</td>
<td>Cat# A32733</td>
<td>(1:500)</td>
<td></td>
</tr>
<tr>
<td>Goat polyclonal anti-Guinea Pig Alexa Fluor 647</td>
<td>Invitrogen</td>
<td>Cat# A21450</td>
<td>(1:500)</td>
<td></td>
</tr>
<tr>
<td>Dye</td>
<td>DAPI (4',6-diamidino-2-phenylindole)</td>
<td>Biotium</td>
<td>Cat# 40043; lot# 15D1117</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------------------------------------</td>
<td>---------</td>
<td>------------------------</td>
<td></td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>PXJ41 Kdm4d</td>
<td>this paper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sequence-based reagent</td>
<td>primers</td>
<td>this paper</td>
<td>designed using Primer3 <a href="http://primer3.ut.ee/">http://primer3.ut.ee/</a></td>
<td></td>
</tr>
<tr>
<td>sequence-based reagent</td>
<td>qPCR primers</td>
<td>this paper</td>
<td>designed using Primer3 <a href="http://primer3.ut.ee/">http://primer3.ut.ee/</a></td>
<td></td>
</tr>
<tr>
<td>commercial assay or kit</td>
<td>Rnaeasy plus kit</td>
<td>Qiagen</td>
<td>catalog number 74134</td>
<td></td>
</tr>
<tr>
<td>commercial assay or kit</td>
<td>DNA free</td>
<td>Ambion</td>
<td>catalog number AM1906</td>
<td></td>
</tr>
<tr>
<td>commercial assay or kit</td>
<td>Superscript Vilo</td>
<td>Invitrogen</td>
<td>catalog number 11754</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>Bouins</td>
<td>Sigma</td>
<td>catalog number HT10132</td>
<td></td>
</tr>
<tr>
<td>software, algorithm</td>
<td>Graphpad prism</td>
<td><a href="https://graphpad.com">https://graphpad.com</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>software, algorithm</td>
<td>Sroogle</td>
<td><a href="http://sroogle.tau.ac.il/">http://sroogle.tau.ac.il/</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Generation of KO mice.**
A targeting construct in which the *Rbmxl2* open reading frame was flanked by *LoxP* sites was made using standard molecular biology techniques, and electroporated into ES129 cells. Positive clones were injected into blastocyst to create chimaeras, and bred to yield agouti pups heterozygous for the targeted locus (Ozgene, Perth, Australia). The original mice containing the *Neomycin* gene (Supplementary Figure 1A) were crossed to *FlpE* mice to remove the *Neo* gene and to generate the *Rbmxl2* *LoxP* conditional allele (Supplementary Figure 1B). Mice containing the *Rbmxl2* *LoxP* conditional allele were crossed with mice expressing *PGKCre*, resulting in deletion of the *Rbmxl2* reading frame (Supplementary Figure 1C). Genetic structures of the wild type and targeted alleles were confirmed by Southern blotting. Genomic DNA was prepared from mouse tails, cut with *BamHI*, run on an agarose gel and blotted onto a Hybond N nylon membrane (GE Healthcare, Little Chalfont, UK). Blots were probed using an internal probe Enp (generated using primers enpF 5′-ACTGTTGATTTCCCCCTTCAAC-3′ and enp R 5′-ACTCCCTGCTGTGGTCTTATC-3′, and α-32P labelled by random priming). The correct insertion of the targeting vector in the genome was confirmed by cutting genomic DNA with *PshA1* and *EcoRV* and hybridizing with 5′ (generated from genomic DNA using 5′-agcattcagcagactcagcagctc-3′ and 5′-ttaaaactgagggagactgc-3′) and 3′ (generated from genomic DNA using 5′-actgcatagttgtagccatc-3′ and 5′-tgcaactctttagcacttc-3′) probes respectively.
Animal work
Animal research was carried with the approval of the Newcastle University animal research ethics committee and the UK Government Home Office (Home Office project Licence Number PIL 60/4455).

Analysis of mouse germ cell development in vivo
Testis/body weight ratios, and sperm counts were measured on a C57Bl6 background. In order to determine sperm counts, the cauda epididymis was dissected in PBS and the sperm were counted in a haemocytometer.

Preparation of RNA for RNA-Seq
Mice were back crossed onto the C57Bl6 background for 8 generations, and we used back crossed male mice for subsequent analysis. Testes were homogenized in RLTplus buffer from Qiagen before purification with the Rneasy plus kit (Qiagen, 74134). The RNA was then re-purified with the kit before Dnase digestion (Ambion).

RNA-Seq Analysis
Paired-end sequencing was done for six samples in total (three biological replicates of 18dpp wild-type and \( \text{Rbmxl2}^{-/-} \) testis) using 75bp reads. Libraries were prepared using TruSeq Stranded mRNA Library Prep Kit (Illumina) and sequenced; 75 bp single reads were sequenced on a NextSeq 500 (Illumina, using the mid output v2 150 cycles kit). The base quality of the raw sequencing reads were checked using FastQC. Trimmmomatic (v0.32) was used to remove adapters and to trim the first twelve bases and the last base at position 76 (Bolger et al., 2014). Reads were aligned to the UCSC D3c. 2011 (GRCm38/mm10) assembly of the mouse genome using STAR (v2.5.2b) (Dobin et al., 2013). Alternative splicing events were assessed using MAJIQ and VOILA software packages (Vaquero-Garcia et al., 2016). Briefly, uniquely mapped, junction-spanning reads were used by MAJIQ to construct splice graphs for transcripts from a custom Ensembl transcriptome annotation and to quantify PSI (within conditions) and ΔPSI (between conditions) for all local splicing variations (LSVs). The captured LSVs include classical alternative splicing events (e.g. cassette exons, alternative 5’ splice sites, etc.) as well as more complex variations (Vaquero-Garcia et al., 2016). LSVs with an expected change of greater than 10% were then visualized using VOILA to produce splice graphs, violin plots representing PSI and ΔPSI quantifications, and interactive HTML outputs for changes between wild type and \( \text{Rbmxl2}^{-/-} \).

In silico analysis of mouse cryptic exons
Splicing patterns of identified target genes were analysed using the UCSC and IGV mouse genome browsers (Karolchik et al., 2014, Thorvaldsdottir et al., 2013). A panel of 10 cryptic exons from the \textit{Catsperb} (2 exons), \textit{Ccduc60}, \textit{Kdm4d}, \textit{Lrccc63}, 4933409G03Rik, \textit{Tcte2}, \textit{Ube2e2}, \textit{Fam178b} genes) and their immediately downstream exons were monitored for strength of splicing sequences using Sroogle (Schwartz et al., 2009). A panel of 19 exons were analysed for length and translation termination codons content (1700074P13Rik, 4833439L19Rik, ccd60, Dnah12, Hmga2, Map2k5, Rbks, Tcte2, Ube2e2, 4933409G03Rik, 4930500J02Rik, 4932414N04Rik, Fam178b, Ccdc144b, Ston1, Usp54, Lca5l, and Catsperb and Lca5l).
Detection of splice isoforms in mouse testis RNA

The levels of different mRNA splice forms were detected in DNAse treated total RNA isolated from 18dpp mouse testes. cDNA was synthesised using superscript Vilo (Invitrogen) and PCR amplified, followed by either capillary gel electrophoresis (Qiagen) or agarose gel electrophoresis (Grellscheid et al., 2011). Gene-specific primers specific for different splice isoforms are given in Figure 3 –Source Data 4.

HITS-CLIP

HITS-CLIP was performed as previously described (Licatalosi et al., 2008) using an antibody specific to RBMXL2 (Ehrmann et al., 2008). In short, for the CLIP analysis a single mouse testis was sheared in PBS and UV crosslinked. After lysis, the whole lysate was treated with DNase and RNase, followed by radiolabelling and linker ligation. After immunoprecipitation with purified antisera specific to RBMXL2 (Ehrmann et al., 2008), RNA bound RBMXL2 was separated on SDS-PAGE. A thin band at the size of 65 kDa (RBMXL2 migrates at around 50 kDa and MW of 50 nt RNA is about 15 kDa) was cut out and subject to protein digestion. RNA was recovered and subject to Illumina sequencing, and the reads were aligned to the mouse genome. HITS-CLIP tags mapped to 22,438 protein coding genes. HITS-CLIP data was further analysed to extract most frequently occurring pentamer nucleotide sequences within genes (ncRNA, ORF, 3’UTR, 5’UTR and intron). A window between +50 and -50 of the cross link site was analysed (to avoid crosslinking bias), and after correction for average pentamer occurrence in control sequences (as described in (Wang et al., 2010)). Each pentamer was only counted once in the analysed window.

Kdm4d minigene analysis

The Kdm4d cryptic exon and flanking intron sequences were PCR-amplified from mouse genomic DNA using the cloning primers Kdm4dF (5’-GAATTCCCACACAGCAAAACCCTCCTC-3’) and Kdm4dR (5’-GAATTCGCCACCTTTTCTATTCTTTCCTT-3’). The PCR products were digested with EcoR1 restriction enzyme and cloned into the pXJ41 vector using the Mfe1 site midway through the 757 nucleotide β-globin intron (Bourgeois et al., 1999). Analysis of splicing patterns transcribed from minigenes was carried out in HEK293 cells as previously described (Grellscheid et al., 2011, Venables et al., 2005) using primers within the β-globin exons of pXJ41, pXJ41F 5’-GCTCCGGATCGATCCTGAGAACT-3’ and pXJ41R (5’-GCTGCAATAACACAAGTTCTGCT-3’).

Detection of proteins in mouse testis

For protein detection by immunohistochemistry, testes were fixed in Bouins, embedded in paraffin wax, sectioned and attached to glass slides. Sections were prepared and immunohistochemistry carried out as previously described (Grellscheid et al., 2011). Primary antibodies were specific to caspase 3, p17-specific (Proteintech Europe Ltd). Protein detection by Western was as previously described (Grellscheid et al., 2011) using primary antibodies specific to RBMXL2 (Ehrmann et al., 2008), Meioc (Abby et al., 2016), and control antibodies specific to tubulin (Sigma T6793) and GAPDH (Abgent, San Diego CA, AP7873b).
Spermatocyte Spreading
The tunica was removed from testes, and seminiferous tubules macerated with razor blades in a small volume of DMEM media. DMEM was added to 4ml, and debris allowed to settle for 10 minutes. Cells were then pelleted from the suspension at 233g for 5 minutes. The pellet was resuspended in 1-2ml DMEM per animal. 5 drops of 4.5% sucrose were added to the centre of each clean glass slide using a Pasteur pipette. A single drop of the cell suspension was added to each slide from a height of 15-20cm, followed by one drop of 0.05% Triton-X-100. The slides were incubated in a humid chamber for 10 minutes, then eight drops of fixative (2% PFA, 0.02% SDS, pH 8) were added to each slide, and incubated in humid chamber for 20 minutes. Slides were dipped in water to wash and allowed to air dry before storing at -80°C (Crichton et al., 2017).

Antibody Staining and Image Capture
Slides were blocked for one hour in 0.15% BSA, 0.1% Tween-20, 5% goat serum PBS, then incubated with primary antibodies for three hours at room temperature or 4°C overnight. Primary antibodies were used at the following concentrations, diluted in block solution: rabbit anti-Sycp3 (1:500, Lifespan Cat# LS-B175 RRID:AB_2197351), mouse anti-Sycp3 (1:500, Abcam Cat# ab97672 RRID:AB_10678841), guinea pig anti-Sycp1 (1:200, from Howard Cooke), mouse anti-γH2AX (1:3000; Millipore Cat# 05-636, RRID:AB_309864), rabbit anti-H3K9me3 (1:500; Abcam Cat# ab8898, RRID:AB_306848), mouse anti-phospho H3 (1:1000; Abcam Cat# ab14955, RRID:AB_443110), human anti-centromere (1:50; Antibodies Incorporated, Cat#15-235-0001), mouse anti-AcSmc3 (1:1000, a gift from Katsuhiko Shirage (Nishiyama et al., 2010)). Slides were washed three times for five minutes in PBS before incubating with 1:500 secondary antibodies and ng/ul DAPI (4',6-diamidino-2-phenylindole) for one hour at room temperature in darkness. Slides washed again three times and mounted in 90% glycerol, 10% PBS, 0.1% p-phenylenediamine.

Images were captured with Micromanager imaging software using an Axioplan II fluorescence microscope (Carl Zeiss) equipped with motorised colour filters. Centromere and H3K9me3 staining was imaged by capturing z-stacks using a piezoelectrically-driven objective mount (Physik Instrumente) controlled with Volocity software (PerkinElmer). These images were deconvolved using Volocity, and a 2D image generated in Fiji.

Nuclei were staged by immunostaining for the axial/lateral element marker SYCP3. Pachytene nuclei were identified in this analysis by complete co-localisation of SYCP3 and either the transverse filament marker SYCP1 on all nineteen autosomes, or by the presence of nineteen bold SYCP3 axes in addition to two paired or unpaired sex chromosome axes of unequal length. Asynapsed pachytene nuclei were judged to be nuclei containing at least one fully synapsed autosome, and at least one autosome unsynapsed along >50% of its length (Crichton et al., 2017).

Quantification and Statistical Analysis
Quantification and statistical analyses were done using Graphpad (http://www.graphpad.com/).

Spermatocyte Staining Analysis
All scoring of immunostained chromosome spreads was performed blind on images after randomisation of filenames by computer script. Three Rbmxl2−/− and three wild type control animals were analysed for each experiment. Meiotic substaging analysis involved 179 nuclei from control animals, and 138 from Rbmxl2−/− mice. Diplotene phospho-H3 analysis involved 128 control and 157 Rbmxl2−/− nuclei, and the proportion of nuclei with bold centromeric signal was analysed using a Fisher’s test. Pachytene asynapsis was measured in 108
control and 92 $Rbmxl2^{-/-}$ nuclei, and the data analysed using a Fisher’s test. Centromere positioning and H3K9me3 staining was assessed across 73 control and 65 $Rbmxl2^{-/-}$ pachytene nuclei, and 66 diplotene control and 66 $Rbmxl2^{-/-}$ diplotene nuclei. These data were analysed using Wilcoxon Rank Sum test.

**Accession codes:** RNAseq data are available via the Gene Expression Omnibus (GEO) using GEO accession number GSE101511.
**Gene family:**
- **RBMX** (all vertebrates)
- **RBMXL2** (mammals)
- **RBMY** (mammals)

**RNA binding site:**
- **RBMX:** AAN
- **RBMXL2:** unknown
- **RBMY:** C(U/A)CAA

**Expression:**
- **RBMX:** Ubiquitous except during /after meiosis
- **RBMXL2:** Testis (only in and after meiosis)
- **RBMY:** Testis (pre-meiosis only)

**Global function:**
- **RBMX:** Unknown
- **RBMXL2:** Unknown
- **RBMY:** Unknown

**Ancestral RBMX gene:**
- 65 MYA
- 180-210 MYA

**Protein interactions:**
- Splicing regulators including Tra2β

**Protein domains:**
- RRM
- Protein interaction domain

**Figure 1**
Figure 2

(A) Testis morphology

(B) Graph showing testis weight/body weight ratio for different genotypes. +/+ and +/- genotypes are compared, with p<0.0001 indicating statistical significance. n.s. denotes non-significant results.

(C) Haematoxylin stained adult testis sections showing differences in Spd, Rtd, Spc, and Spg staining patterns between Rbmxl2 +/+ and Rbmxl2 -/- genotypes.

(D) Haematoxylin stained epididymal sections of Rbmxl2 +/+ and Rbmxl2 -/- genotypes.

(E) Sperm count per million in WT mice for Rbmxl2 +/+ and Rbmxl2 -/- genotypes. +/- genotypes result in a statistically significant decrease in sperm count.


Genotypes of nuclear spreads:
- Rbmxl2 +/+
- Rbmxl2 -/-
Figure 3

(A) Mean of normalised counts

(B) E(PSI) WT vs. E(PSI) KO

(C) Pie charts showing transcription and splicing

(D) HITS-CLIP top 10 pentamers

<table>
<thead>
<tr>
<th>Pentamer</th>
<th>Z score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAAT</td>
<td>19.837</td>
</tr>
<tr>
<td>AAATA</td>
<td>17.217</td>
</tr>
<tr>
<td>ATAAAA</td>
<td>17.127</td>
</tr>
<tr>
<td>AATAA</td>
<td>16.821</td>
</tr>
<tr>
<td>CAAAA</td>
<td>15.39</td>
</tr>
<tr>
<td>TAAAA</td>
<td>15.183</td>
</tr>
<tr>
<td>AAAAA</td>
<td>14.934</td>
</tr>
<tr>
<td>TAAAT</td>
<td>13.958</td>
</tr>
<tr>
<td>AACAA</td>
<td>13.67</td>
</tr>
<tr>
<td>AAACAA</td>
<td>13.302</td>
</tr>
</tbody>
</table>

Figure 3
(A) Diagram showing the location of the cryptic exon relative to the Kdm4d and beta globin exons.

(B) Graph showing the percentage of PSI (Processing Site Index) with and without the cryptic exon, and the corresponding p-values for each condition.

(C) Graph showing the percentage of co-transfected proteins with and without the cryptic exon, and the corresponding p-values for each condition.

Figure 6
Event types controlled by RBMXL2 (N=87 events)

- Annotated intron retentions
- Annotated terminal exons
- Annotated skipped exons
- Annotated alternative splice sites
- Cryptic intron retentions
- Cryptic terminal exons
- Cryptic skipped exons
- Cryptic alternative splice sites

Meiotic cryptic splicing repression
1. RBMXL2 inhibits splicing activator Tra2β
2. RBMXL2 binds some splice site sequences

Meiotic cryptic splicing
Absence of RBMXL2 leads to cryptic splice site selection
Figure 1 - figure supplement 2
Figure 2 - figure supplement 1

(A) Adult testis section PAS

(B) Rbmxl2 genotype

Round spermatids/tubule

p=0.0013

+B+/+  -/- Rbmxl2
Figure 2 - figure supplement 2
Figure 2 - figure supplement 3
Genotypes of nuclear spreads

<table>
<thead>
<tr>
<th></th>
<th>+/+</th>
<th>-/-</th>
<th>+/+</th>
<th>-/-</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K9me3 domains</td>
<td>+/+</td>
<td>-/-</td>
<td>+/+</td>
<td>-/-</td>
</tr>
<tr>
<td>pachytene (p=0.00081)</td>
<td>5.62e-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>diplotene (p=1.39e-12)</td>
<td>5.62e-16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Percentage sex bodies labelled with H3K9me3

<table>
<thead>
<tr>
<th></th>
<th>+/+</th>
<th>-/-</th>
</tr>
</thead>
<tbody>
<tr>
<td>pachytene (p=0.21)</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>diplotene (p&lt;0.0001)</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

Percentage asynapsis

<table>
<thead>
<tr>
<th></th>
<th>+/+</th>
<th>-/-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotypes</td>
<td>+/+</td>
<td>-/-</td>
</tr>
<tr>
<td>n.s.</td>
<td>+/+</td>
<td>-/-</td>
</tr>
</tbody>
</table>

Figure 2 - figure supplement 4
(A) RNAseq reads

(B) qPCR analysis

Fsip2 gene

Rbmxl2 +/- day 18

Rbmxl2 -/- day 18

RNAseq reads

CLIP tags

NCBI RefSeq genes, curated subset (NM_*, NR_*, and YP_*) - Annotation Release GCF_000001635.25_GRCm38.p5 (2017-08-04)

UCSC annotations of RefSeq RNAs (NM_* and NR_*).

NCBI RefSeq genes, predicted subset (XM_* and XR_*) - Annotation Release GCF_000001635.25_GRCm38.p5 (2017-08-04)

60 vertebrates conservation by PhastCons.

(blocked expression of Fsip2 exon 16)

Genotype

Fold change

p=0.0008
Figure 3 - figure supplement 3
<table>
<thead>
<tr>
<th>GO term</th>
<th>GO term</th>
<th>GO term</th>
<th>MGI infertility phenotype</th>
<th>MGI lethal phenotype</th>
<th>MGI development phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spermatogenesis</td>
<td>Meiotic cell cycle process</td>
<td>Germ cell development</td>
<td>Brca2</td>
<td>Kdm4d</td>
<td>Ppox</td>
</tr>
<tr>
<td>Catsperb</td>
<td></td>
<td></td>
<td>Catsperb</td>
<td>Ccdc62</td>
<td>Mylk</td>
</tr>
<tr>
<td>Hmga2</td>
<td></td>
<td></td>
<td>Hmga2</td>
<td>Pcsk4</td>
<td>Hmga2</td>
</tr>
<tr>
<td>Odf2</td>
<td>4930447C04Rik</td>
<td>4930447C04Rik</td>
<td>Alms1</td>
<td>Zfp148</td>
<td>Alms1</td>
</tr>
<tr>
<td>Tlk2</td>
<td>Mlh3</td>
<td>Alms1</td>
<td>Meioc</td>
<td>Ald5h1</td>
<td>Plec</td>
</tr>
<tr>
<td>Pcsk4</td>
<td></td>
<td></td>
<td>Gba2</td>
<td>a</td>
<td>Tmpo</td>
</tr>
<tr>
<td>Lrguk</td>
<td>Brca2</td>
<td>Gabbp1</td>
<td>Nxn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4930447C04Rik</td>
<td>Poln</td>
<td>Clec16a</td>
<td>Med1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slc9c1</td>
<td>Pcsk4</td>
<td>Tln1</td>
<td>Tln1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alms1</td>
<td>Zbtb16</td>
<td>Brca2</td>
<td>Brca2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zfp148</td>
<td>Ebp41</td>
<td>Vit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mlh3</td>
<td>Tnrc6a</td>
<td>Slc39a13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gabpa</td>
<td>Furin</td>
<td>Furin</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nectin3</td>
<td>Wdpcp</td>
<td>Wdpcp</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slc9c1</td>
<td>Gabpa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lnpk</td>
<td>Slc39a8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tmed10</td>
<td>Med1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clip1</td>
<td>Slc7a1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Brd4</td>
<td>Lnpk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hip1</td>
<td>Psmc3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lrrc1</td>
<td>Tmed10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Myo7a</td>
<td>Coq2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Odf2</td>
<td>Brd4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spag6l</td>
<td>Hip1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4930447C04Rik</td>
<td>Nxn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dnal1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seh1l</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plec</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mapk25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Atp8b2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Elac2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cep67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
E(PSI) of quantifiable putative cryptic junctions

Mouse tissue

(A)
Figure 4- figure supplement 2

(A) 

Rbmxl2  +/+ day 18

Rbmxl2  -/- day 18

(B) 

Fold change

Genotype

p=0.0821

Kdm4d gene

Rbmxl2  +/+ day 18

Rbmxl2  -/- day 18
(A)

**Rbmxl2** +/+ day 18

**Rbmxl2** -/- day 18

Region containing cryptic splice sites

(B)

**Age:** 10dpp 12dpp 13dpp 13dpp 14dpp 14dpp 16dpp 16dpp 18dpp 18dpp 21dpp

**Key:**
- ↓ splice site

### Region containing cryptic splice sites

#### Rbmxl2 +/+ day 18

---

#### Rbmxl2 -/- day 18

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Splice site score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotated 3’ splice site</td>
<td>tag↓GACC</td>
</tr>
<tr>
<td>Cryptic 5’ splice site</td>
<td>AG↓gccagta</td>
</tr>
<tr>
<td>Cryptic 3’ splice site</td>
<td>tttcaag↓ACCAA</td>
</tr>
<tr>
<td>Annotated 5’ splice site</td>
<td>AAG↓gtaa</td>
</tr>
</tbody>
</table>

---

*Figure 5 - figure supplement 1*