Y chromosomal noncoding RNAs regulate autosomal gene expression via piRNAs in mouse testis

Hemakumar M. Reddy[1,2,16], Rupa Bhattacharya[1,3,16], Zeenath Jehan[1,4], Kankadeb Mishra[1,5], Pranatharthi Annapurna[1,6], Shrish Tiwari[1], Nissankararao Mary Praveena[1], Jomini Liza Alex[1], Vishnu M Dhople[1,7], Lalji Singh[8], Mahadevan Sivaramakrishnan[1,9], Anurag Chaturvedi[1,10], Nandini Rangaraj[1], Thomas Michael Shiju[1,11], Badanapuram Sridevi[1], Sachin Kumar[1], Ram Reddy Dereddi[1,12], Sunayana M Rayabandla[1,13], Rachel A. Jesudasan[1,14*, 15*]

[1]Centre for Cellular and Molecular Biology (CCMB), Uppal Road, Hyderabad, Telengana – 500007, India.

Present address:

[2]Brown University BioMed Division, Department of Molecular Biology, Cell Biology and Biochemistry, 185 Meeting Street room 260, Sidney Frank Life Sciences Building, Providence, RI 02912, USA.

[3]Pennington NJ 08534, USA.

[4]Department of Genetics and Molecular Medicines, Vasavi Medical and Research Centre, 6-1-91 Khairatabad, Hyderabad 500 004 India.

[5]Department of Cell Biology, Memorial Sloan Kettering Cancer Centre, Rockefeller Research Laboratory, 430 East 67th Street, RRL 445, New York, NY 10065, USA.

[6]Departments of Orthopaedic Surgery & Bioengineering, University of Pennsylvania, 376A Stemmler Hall, 36th Street & Hamilton Walk, Philadelphia, PA 19104.USA.

[7]Ernst-Moritz-Arndt-University of Greifswald Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, Friedrich-Ludwig-Jahn-Straße 15 a, 17487 Greifswald, Germany.

[8] Deceased.

[9]Jubilant Biosystems Ltd., #96, Industrial Suburb, 2nd Stage, Yeshwantpur, Bangalore-560022, Karnataka, India.

[10]Department of Ecology and Evolution, University of Lausanne, Biophore Building, 1015, Lausanne, Switzerland.

[11]Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio 44120. USA

[12] Institute for Anatomy and Cell Biology, building-307, Heidelberg, Germany.

[13]Telangana Social Welfare Residential Degree college for Women, Suryapet-508213. Telangana. India.

[14]Department of Genetics, Osmania University, Hyderabad – 500007. Telangana. India.

[15]Inter University Centre for Genomics & Gene Technology, Karyavattom Campus, University of Kerala, Trivandrum, Kerala, India.

[16]These authors contributed equally to the work.

*Correspondence: Email ID: rachellike@gmail.com

Running title: Y-ncRNAs regulate autosomal gene expression

Key Words: Mouse Y chromosome; noncoding RNA; alternative splicing; piRNA; *Pirmy*; male sterility.

ABSTRACT

Majority of the genes expressed during spermatogenesis are autosomal. Mice with different deletions of Yq show sub-fertility, sterility and sperm abnormalities. The connection between Yq deletion and autosomal gene regulation is not well understood. We describe a novel mouse Yq-derived long noncoding RNA, *Pirmy*, which shows unprecedented number of splice variants in testis. Further, *Pirmy* transcript variants act as templates for several piRNAs. We identified ten differentially expressed autosome-encoded sperm proteins in mutant mice. *Pirmy* transcript variants have homology to 5'/3'UTRs of these deregulated autosomal genes. Thus, subfertility in Y-deleted mice appears to be a polygenic phenomenon that is partially regulated epistatically by the Y-chromosome. Our study provides novel insights into possible role of MSY-derived ncRNAs in male fertility and reproduction. Finally, sperm phenotypes from the Y-deleted mice seem to be similar to that reported in inter-specific male-sterile hybrids. Taken together, this study provides novel insights into possible role of Y-derived ncRNAs in male sterility and speciation.

INTRODUCTION

Y chromosome has come a long way from a single-gene male determining chromosome to one that houses a few protein-coding genes besides sequences crucial for spermatogenesis and fertility (Bellott et al., 2014; Jehan et al., 2007; Kuroda-Kawaguchi et al., 2001; Moretti et al., 2017; Piergentili, 2010; Soh et al., 2014; Tiepolo & Zuffardi, 1976). Earlier studies have shown that genes involved in sex determination and spermatogenesis are present on the short arm. Several lines of evidence indicate that the male-specific region on the long arm of the Y chromosome (MSYq) in mouse is replete with highly repetitive mouse-specific

sequences that are expressed in spermatids (Burgoyne & Mitchell, 2007; Conway et al., 1994; Prado, Lee, Zahed, Vekemans, & Nishioka, 1992; Toure et al., 2005).

Previously published data have described two different strains of mice with partial deletions of the long arm of the Y chromosome (Yq) (Conway et al., 1994; Józefa Styrna, Bili, & Krzanowska, 2002). Mice from both the genetic backgrounds exhibit male sterile phenotypes such as subfertility, sex ratio skewed towards females, reduced number of motile sperms, aberrant sperm motility and sperm head morphological abnormalities (Conway et al., 1994). Mice with partial deletions of Yq show sperm abnormalities with less severe phenotype whereas mice with total deletion of the Yq have extensive sperm morphological aberrations and are sterile (Toure et al., 2004). This suggested the presence of multicopy spermiogenesis gene(s) on mouse Yq (Burgoyne, Mahadevaiah, Sutcliffe, & Palmer, 1992; Conway et al., 1994). Subsequently multicopy transcripts such as Y353/B, spermiogenesis-specific transcript on the Y *(Ssty)* and Sycp3-like, Y-linked *(Sly)* from the mouse Yq were projected as putative candidate genes for male sterility and spermiogenesis in mice (Cocquet et al., 2009; Conway et al., 1994; Riel et al., 2013; Toure et al., 2004). As mice with partial deletions of Yq (XY^{RIII}qdel, 2/3rd interstitial deletion of Yq) show reduced expression of *Ssty* and impaired fertility, this gene (present on Yq) was implicated in spermatogenesis (Toure et al., 2004). The next major gene to be discovered on mouse Yq was the multicopy *Sly*. As SLY interacts with a histone acetyl transferase and is an acrosomal protein, the authors suggested that *Sly* could control transcription and acrosome functions (Reynard, Cocquet, & Burgoyne, 2009). Further, Cocquet and colleagues observed major problems in sperm differentiation when they disrupted functions of the *Sly* gene by transgenic delivery of siRNA to the gene. Therefore, *Sly* was conjectured as a putative candidate gene for spermiogenesis (Reynard et al., 2009). However, subsequently Ward and

colleagues showed that *Sly* expression alone is not sufficient for spermiogenesis (Riel et al., 2019).

Vast majority of the genes required for spermatogenesis and spermiogenesis are non-Y-linked (Choi et al., 2007; Schultz, Hamra, & Garbers, 2003; Xiao, Tang, Yu, Gui, & Cai, 2008). Deletions of the Y-chromosome leading to different degrees of male infertility prompted us to hypothesise interactions between Y-derived transcripts and autosomal genes in male fertility. Earlier, studies in the lab reported a testis-specific chimeric transcript, generated by trans-splicing between the *CDC2L2* mRNA from chromosome number 1 and a long noncoding RNA from the Y chromosome in human (Jehan et al., 2007). We hypothesized more such interactions between protein coding genes on autosomes and noncoding RNAs from the Y chromosome. In this context, we studied a mutant mouse, which had a partial deletion of the Y-chromosome heterochromatic block XY^{RIII}qdel (Conway et al., 1994) to look for regulatory elements, if any, in the deleted region.

Previous studies in the lab identified 300-400 copies of a mouse Y chromosome specific genomic clone, M34 (DQ907163.1) using slot blot, sequencing and bioinformatics analyses (Bajpai, Sridhar, Reddy, & Jesudasan, 2007; Singh, Panicker, Nagaraj, & Majumdar, 1994). There is a reduction of M34 copies in the XY^{RIII}qdel mice that exhibit multiple sperm abnormalities. As deletions of the Y chromosome show sperm abnormalities, we reasoned that these repeat sequences could have important functional role(s) in the multistep developmental process of sperm production. The XY^{RIII}qdel mice also showed reduced transcription of M34 compared to its wild type XY^{RIII} mice in testis. In order to understand putative functions of this sequence, first of all we identified a transcript corresponding to M34, *Pirmy,* from mouse testis. Subsequent experiments identified multiple splice variants of this transcript. Parallel experiments identified deregulated proteins in the sperm

proteome of the XY^{RIII}qdel mice. Interestingly, genes corresponding to all these proteins localized to different autosomes. Further, we showed that the UTRs of these genes bear homology to piRNAs derived from *Pirmy*. Thus, our results demonstrate for the first time (i) a novel noncoding RNA (*Pirmy*) on mouse Y long arm (ii) large number of splice variants of *Pirmy*, and the generation of piRNAs from it in mouse testis and (iii) their putative role in regulation of autosomal genes involved in male fertility and reproduction.

RESULTS

M34 is transcribed in mouse testis

To address the precise function of the M34 transcript we confirmed the localization of the sex and species-specific repeat M34 on mouse Y long arm again by fluorescence in situ hybridization (FISH) (Fig. 1*A*). BLAST analysis of M34 sequence against mouse whole genome also showed maximum similarity to Y chromosome (>97% identity) with few hits on the X (Fig. 1*B,* NCBI build 38.1). M34 was then analysed for expression in adult mouse testis. FISH using M34 as a probe revealed abundant transcription in testis (Fig. 1*C*). Pretreatment with RNase abolished these signals confirming the presence of RNA (Fig. 1*D*). Expression profiling by FISH in testes showed the presence of M34 transcripts in 18-day embryos, newborns and 1-month old mice (30 days postpartum) (Fig. S1), suggesting transcription of this repeat in mouse testis from early developmental stages.

Isolation of a novel polyadenylated noncoding transcript from mouse Y chromosome

Sequence analysis of the 9.51 kb M34 clone using Tandem Repeats Finder (TRF) and RepeatMasker identified different simple sequence repeats and partial mid-repetitive sequences like LINEs, SINEs and LTR elements which constitute ~35% of the total M34

96 sequence (Fig. 1*E*). A number of gene prediction programs like GENSCAN, GrailEXP, MZEF,
97 and GeneMark did not predict any genes within M34 with consistency. BLAST analysis of
98 M34 sequence against the EST database of NCBI (NCBI build 36.1) identified 5 ESTs at the 3'
99 end of M34 sequence (Fig. 1*E*). Expression of these ESTs was observed in embryos from at
100 least 13.5d onwards (data not shown). Two of these ESTs showed male-specific expression
101 by Reverse transcription PCR (RT-PCR) analysis.
102 In order to identify a cDNA corresponding to M34 in testis, one of the male specific ESTs
103 (CA533654) was used to screen a mouse testis cDNA library. A 1395 nt long polyadenylated
104 Y-specific cDNA was isolated and this was named *Pirmy* - piRNA from mouse Y chromosome
105 (DQ907162.1). FISH on to mouse metaphase spreads showed that *Pirmy* is present only on
106 the Y chromosome in multiple copies (Fig. 1*F*), similar to that of the genomic clone M34.
107 BLAST of *Pirmy* against the nucleotide database of NCBI picked up only mouse sequences
108 with statistically significant alignments (e-value $<4e^{-04}$). This suggests that *Pirmy* is specific to
109 mouse. BLAST analysis using *Pirmy* against the mouse genome plus transcriptome database
110 showed homology to *Sly* transcript in exons 1-5. The exons 2-5 were identical in *Pirmy* and
111 *Sly* (Fig. 1*G*). The region of homology is less than a quarter of *Pirmy*, and is confined to the
112 5'end. The entire sequence of *Pirmy* localizes to the Y chromosome at 4341127-4381724
113 (GRCm39). BLAST search against the nucleotide database identified a reference sequence
114 NR164186, which was annotated as a long noncoding RNA using evidence data for transcript
115 exon combination from *Pirmy*. The exon-intron junctions of *Pirmy* are shown in Figure 1H.

116 **Splice variants of *Pirmy* in mouse testis**

117 *Pirmy* was analyzed further by RT-PCR. Two rounds of amplification using primers to the two
118 ends of *Pirmy* yielded multiple products in testis and brain (Fig. 2*B*). PCR products were
119 cloned. Sanger sequencing of more than 1000 clones from testis yielded 107 transcripts

(NCBI accession numbers FJ541075-FJ541181), besides the one obtained by screening testis-cDNA library (Figs. 2A, 3). BLAST analysis of these transcripts against the NCBI genome database (Build 39) showed that 28 of these transcripts (FJ541075- FJ541102; Fig. 3) localise to multiple regions on the mouse Y chromosome. The remaining 79 transcripts (FJ541103-FJ541181) and *Pirmy* were present at a single locus on the Y chromosome at 4341127-4381724 (GRCm39) (Fig. 2). Thus, the intron-exon organization, found in *Pirmy* and the 79 splice variants is present only at a single locus on the Y chromosome. These 79 transcripts could therefore be alternatively spliced isoforms of *Pirmy.* Each exon from *Pirmy* is present in multiple copies on the Y chromosome. The 28 transcript variants reveal that different combinations of these exons and introns are there at multiple loci on the Y. The splice variants of *Pirmy* exhibited the full spectrum of splicing patterns like exon skipping, alternative 5' and 3' splice sites, mutually exclusive alternative exons, intron retention and combination of different splicing events (Fig. S2). The splice variants of *Pirmy* contained consensus splice signal sequences at all the intron-exon junctions (Data sheet).

Splice variants of *Pirmy* in mouse brain

Many transcripts that are expressed in testis are also known to be expressed in brain; therefore, we studied the expression of *Pirmy* in mouse brain also. RT-PCR amplification (Fig. 2*B*), cloning and sequencing of *Pirmy* products from mouse brain yielded 12 transcripts. All these transcripts localize to the single locus, to which *Pirmy* and its splice variants localized (4341127-4381724 (GRCm39)). In fact, comparison of the exons of the splice variants from the brain and testis showed that the same splice isoforms were present in both the tissues (Fig. 4).

Expression of M34 in XY^{RIII}qdel mice

143 Metaphase spreads from the XY^{RIII}qdel mice showed a reduction in copy number of M34 on
144 the Y chromosome showing that it localizes to the deleted region of the Y chromosome in
145 the above mice (Fig. S3*B*). Therefore, expression of M34 was then checked in testis and
146 sperms of XY^{RIII} and XY^{RIII}qdel mice by FISH. Dramatic reduction in fluorescence intensity was
147 observed in testis and sperm of XY^{RIII}qdel mice. However, sperms from epididymis of both
148 XY^{RIII} and XY^{RIII}qdel mice showed faint fluorescence intensity (Fig. S3*B*). The subclones of
149 M34 (Fig. S3A) when used as probes on testis sections showed reduction in fluorescence
150 intensity in XY^{RIII}qdel mice (Fig. 3*C*). We also checked the copy number of *Pirmy* in genomic
151 DNA isolated from the wild type and XY^{RIII}qdel mice by Real-Time PCR using primers to exon
152 7; a significant reduction in copy number was observed in the XY^{RIII}qdel genome (Fig. S4C).

153 **Many proteins coded by autosomal genes are deregulated in XY^{RIII}qdel sperm proteome**

154 We then analysed the motility profile of sperms from XY^{RIII} and XY^{RIII}qdel mice and identified
155 a stark difference in motility patterns (Movies S1, S2 respectively). Spermatozoa from XY^{RIII}
156 mice show linear progressive motion whereas sperms from XY^{RIII}qdel mice show rapid
157 flagellar movement with non-linear and non-progressive motion. Most of the spermatozoa
158 from XY^{RIII}qdel mice stall at the same position with no linear displacement. Studies from
159 Burgoyne's lab reported morphological abnormalities in sperms from XY^{RIII}qdel mice
160 (Conway et al., 1994). In order to understand the connection between the Y-deletion and
161 sperm abnormalities, we performed comparative sperm proteome analysis between normal
162 mice and the XY^{RIII}qdel mice by 2D-PAGE and mass spectrometry using protocols
163 standardized in the laboratory (Bhattacharya, Devi, Dhople, & Jesudasan, 2013).

164 This analysis identified 8 protein spots that were differentially expressed in the pI ranges of
165 4-7 and 5-8 (Fig. 5*A*). Surprisingly four of these i.e. calreticulin (D1), Cu/Zn superoxide
166 dismutase (SOD (D4)), fatty acid binding protein 9 (FABP9 (D5)), Serine Peptidase Inhibitor

167 (Kazal type II (SPINK2)/Acrosin-Trypsin inhibitor variant 2 (D2) were upregulated in XY^{RIII}qdel
168 sperms compared to XY^{RIII} sperms (Fig. 5*A*). A novel shorter isoform of SPINK2 - SPINK2
169 variant3 (D3, Q8BMY) which was shorter by 27 amino acids at the N-terminal end was
170 downregulated in XY^{RIII}qdel sperms (Fig. 5*A*, Fig. S4B). Three proteins were not detectable in
171 XY^{RIII}qdel sperms in the pI range of 5-8 (Fig. 5A). Two of these were reported as hypothetical
172 proteins in the NCBI database (1700001L19 Riken cDNA and 1700026L06 Riken cDNA. As we
173 have identified these proteins, they have now been deposited in the Uniprot database with
174 accession numbers Q9DAR0 and, Q7TPM5 (MAST (Bhattacharya et al., 2013) respectively.
175 The third one was Stromal cell derived factor 2 like 1 (SDF2L1 (B)). Expression of four of the
176 eight differentially expressed proteins was also confirmed by western blotting in testis and
177 sperms (Fig. 5C).
178

179 Calreticulin, SOD and FABP9 showed upregulation in XY^{RIII}qdel sperms by both 2D PAGE and
180 western blot analyses; MAST was not detectable by both the techniques in sperms. Protein
181 expression of calreticulin, SOD, FABP9 and MAST did not vary significantly between testes of
182 XY^{RIII} and XY^{RIII}qdel mice. Thus, in sperms from the XY^{RIII}qdel, four proteins were upregulated
183 while one was downregulated. Three proteins were not visible in the XY^{RIII}qdel sperm
184 proteome. Surprisingly, all the eight genes corresponding to the differentially expressed
185 sperm proteins localized to different autosomes (Fig. 5B).

186 Next, we analysed the expression of the transcripts corresponding to the protein spots in
187 testis by Real-Time PCR/Northern blot analysis (Fig. S4A, B). Although the proteins Q9DAR0,
188 SDF2L1 and MAST were not detectable in sperms of X^{YRIII}qdel, the corresponding RNAs were
189 present in testis. The transcripts of SDF2L1, MAST, calreticulin and SPINK2 variant 2 proteins

were upregulated in XY^{RIII}qdel mice testis. In contrast transcripts of Q9DAR0 and Spink2 variant 3 did not show significant quantitative difference between the two (Fig. S4B).

UTRs of deregulated autosomal genes show homology to 108 transcripts

The fact that a few autosomal genes were deregulated when there was a deletion on the Y chromosome prompted us to investigate the mechanism behind this puzzling observation. We hypothesized that *Pirmy* transcript variants that show reduction in copy number in the genome and reduced transcription in XY^{RIII}qdel, could be regulating autosomal gene expression in testis. BLAST analysis of these transcripts against the 3' and 5' UTRs of the deregulated genes revealed short stretches of homology ranging in size from 10-16 nucleotides, in either +/+ or +/- orientations. BLAST analysis of the 108 transcripts against the UTRs of deregulated genes identified 21 different hits. Of these 11 were from *Pirmy* splice variants and 10 from the other transcript variants respectively (Fig 6). There are as many as 7 *Pirmy* hits in the 3' UTR of Q9DAR0 (Fig. 6). Homology between Y-derived transcripts and UTRs of deregulated autosomal genes indicates interactions between genes on the Y chromosome and autosomes in mouse testis. A BLAST against the entire transcripts of the deregulated genes showed homology to the coding regions also; however, more stringent BLAST parameters, showed homology to the UTRs alone.

Furthermore, we performed BLAST analysis of all the 108 transcripts (ncRNAs) against the entire UTR database. This identified small stretches of homology in the UTRs of a number of genes across different species. The homologous sequences (10-22 nt) localized to both exons and exon-exon junctions of the 108 transcripts. Representation of some UTR homologies are shown in Fig. 7A. We identified 372 unique homologous stretches in the 108 transcripts. Of these 302 (81.19%) localized to the exons and 70 (18.82%) to exon-exon

junctions (Fig. 7B). The autosomal genes bearing homology to these ncRNAs in their UTRs, are expressed in multiple tissues including testis/epididymis (Table S1).

Identification of ~30nt RNAs from *Pirmy* transcript variants

To check if these short stretches of homologies corresponded to small RNAs, few representative oligonucleotide sequences from the ncRNAs with homology to different UTRs were used as probes (Fig. S5) on small RNA northern blots. Two genes were chosen from the deregulated proteins Q9DAR0 (Spot A) and superoxide dismutase (SOD); Four genes (butyrylcholinesterase (*Bche*), phospholipase A2, group XIIB *(PLA2G12B) and* sialophorin *(Spn)*)were chosen from the BLAST output against UTR database. All the above probes elicited approximately 26-30 nt long testis-specific signals, of the size of piRNAs (Fig. 7*C*).

To confirm that these homologous sequences are indeed piRNAs, different experiments were designed. As the antiparallel strands of DNA are reported to express different levels of piRNA, differential expression from the antiparallel strands were studied using Sense (S) and antisense (AS) probes designed to homologous stretches in the 3' UTRs of *Sod* and *Bche*. Identical experimental conditions showed differential expression of these 30 nt species of RNAs from the two strands of DNA (Fig. 7*D*), further indicating that these short RNAs could be piRNAs.

As piRNAs are PIWI/MIWI binding small RNAs, Electrophoretic Mobility Shift Assay (EMSA) using the *Pirmy*-derived oligonucleotides and recombinant MIWI protein was done to check if the sequences with homologies to UTRs of different genes are indeed piRNAs. Representative gel shifts using oligonucleotides from UTRs of Q9DAR0 and *Sod* are depicted in Fig. 7E and F respectively. piR1 (Girard, Sachidanandam, Hannon, & Carmell, 2006), a known piRNA served as the positive control. Specificity of binding was indicated by the use

237 of corresponding cold competitors as described in the legend to Fig. 7E and F. The piRNA
238 derived oligonucleotides competed out binding of piR1 to MIWI protein and vice versa. This
239 confirmed that these oligonucleotides are indeed MIWI-binding RNAs and therefore piRNAs.
240 The mobility shift was also competed out by MIWI antibody while Argonaute 3 antibody did
241 not alter the mobility of the gel-shifted band obtained with MIWI indicating specificity of
242 binding. These experiments provide further evidence that these short RNA sequences are
243 piRNAs.

244 ***Pirmy* identifies piRNAs from NCBI Archives database**

245 Next line of evidence to the fact these short stretches of homologies are piRNAs came from
246 the NCBI Archives database for piRNAs. BLAST analysis using the 108 transcripts identified a
247 total of 1445 small RNAs in the Sequence Read Archives SRP001701 and SRP000623 with
248 >95% identity. These transcripts identified 352 MIWI2 - MILI-associated reads in SRP000623.
249 Of these 310 had homology to midrepetitive sequences (SINEs, LINEs and LTRs) and 5 had
250 homology to X chromosome. Thus, the remaining 37 piRNAs could be specifically derived
251 from the Y chromosome.

252 **Antagopirs downregulate reporter gene expression**

253 Complementary oligonucleotides synthesized to piRNA sequences present in UTRs of *Sod*
254 and *PLA2G12B* were designated as antagopirs (Fig. S5). Cloning of these UTRs 3' to the
255 Luciferase gene reduced its expression. Treatment with increasing concentrations of
256 antagopirs, i. e. 5 nM, 10 nM and 20 nM caused further concentration dependent reduction
257 in Luciferase expression (Fig. 7*H*). The antagopirs to *Sod* and *PLA2G12B* did not have an
258 effect when the UTR from a non-target gene (*Cdc2l1)* was cloned 3' to the Luciferase gene.
259 Fig. 7G is a schematic representation of 3'UTR reporter construct. Thus, the use of
260 antagopirs to piRNAs could modulate gene expression in vitro.

This study therefore, paved the way for a series of novel and exciting observations. We have identified, a novel, polyadenylated lncRNA (*Pirmy*) expressed from mouse Yq in testis, that shows the large number of splice variants in testis. These Y-derived transcripts harbour piRNAs that have homology to the UTRs of a few autosomal genes expressed in mouse testis. The proteins expressed from these autosomal genes are deregulated in sperms of Y-deleted mice and appear to be regulated by piRNAs generated from MSYq derived lncRNAs. Consolidation of the results from the proteomics analysis and the molecular studies suggested that piRNAs generated from male-specific mouse Yq regulate expression of multiple autosomal genes in testis. Partial deletion of Yq resulted in deregulation of these proteins leading to sperm anomalies and subfertility. Thus, subfertility in mice appears to be a polygenic phenomenon that is regulated epistatically by the Y chromosome.

DISCUSSION

Y-chromosomes harbour genes for male determination and male fertility. Yet the role of Y chromosomal repeats in male fertility and spermatogenesis remains enigmatic. In this study, we elucidate the role in male fertility of a species-specific repeat, M34, from mouse Y long arm, which is transcribed in mouse testis. Transcription from repeats on mouse Y chromosome has been reported earlier. Testis-specific transcription of a family of poly(A)RNAs from the mouse Y chromosome was first reported by Bishop and Hatat using a multicopy Y-derived probe, pY353/B (Bishop & Hatat, 1987; Conway et al., 1994). Subsequently more transcripts were identified in mouse testis, using repeat sequences localizing to mouse Y chromosome (Prado et al., 1992; Toure et al., 2005). The report of *Pirmy* transcript variants described in this study adds to the repertoire.

284 The 108 *Pirmy* transcript variants in the present study were discovered serendipitously by
285 cloning and sequencing of the multiple RT-PCR products obtained using primers to the initial
286 and terminal exons. As the primers were restricted to just two of the exons, it is possible
287 that we might discover more transcript variants using primers to different combinations of
288 exons for RT-PCR amplification. Alternative splicing has been reported in ncRNAs (Pang et
289 al., 2005), yet, such extensive splicing as observed in our study has not been reported for
290 any of them. Very few polymorphically spliced genes have been described earlier from sex
291 chromosomes, particularly the Y (Ellis, Ferguson, Clemente, & Affara, 2007; Wegmann,
292 Dupanloup, & Excoffier, 2008). Thus, the 80 splice variants of *Pirmy* in this study appears to
293 be by far the maximum number of isoforms characterized from a single ncRNA by
294 alternative splicing. Elucidation of identical splice isoforms in mouse brain and testis in this
295 study (Figs. 2, 4) shows that these alternative splicing events are precise and not random
296 because the same splicing events take place in both the tissues. The consensus splice signal
297 sequences present at the intron-exon junctions in all the splice variants further reaffirm
298 programmed splicing events.
299 The identification of deregulated proteins, for which the corresponding genes localise to
300 autosomes, in a Y-deletion mutant was a surprise. Identification of small stretches of
301 homology in the UTRs of these transcripts from the *Pirmy* transcript variants established the
302 connection between the two. Observation of 27-32 nt long signals on small RNA northern
303 blots using the short homologous sequences as probes, which also bind to MIWI protein and
304 show differential expression from the antiparallel strands of DNA indicated that these
305 probes identify piRNAs. Some of these also identified piRNAs in the SRA databases. Small
306 RNAs of the size of ~27-32 nucleotides that bind PIWI protein are classified as piRNAs. These
307 are also known to be differentially transcribed from the two strands of DNA. The presence

308 of piRNAs in the UTRs of genes corresponding to deregulated proteins, suggests the
309 regulation of these autosomal genes by Y-derived piRNAs. The concentration dependent
310 reduction of Luciferase expression by antagopirs corroborates the regulation of these genes
311 by Y-derived piRNAs. Proteins from three other autosomal genes are also deregulated in
312 different strains of Y-deleted mice besides the ones identified in the proteomics screen;
313 caldendrin is upregulated in XY^{RIII}qdel sperm (Bhattacharya et al., 2013) and acrosin is
314 downregulated in B10.BR-Y^{del} sperm. (J. Styrna, Klag, & Moriwaki, 1991). Aromatase is
315 overexpressed in both the Y-deleted strains of mice i.e. XY^{RIII}qdel (our unpublished
316 observation), and B10BR-Y^{del} (Kotula-Balak, Grzmil, Styrna, & Bilinska, 2004). Therefore, it is
317 not surprising to find *Pirmy* homologous sequences in the UTRs of caldendrin, acrosin and
318 aromatase genes, suggesting Y- mediated regulation for these genes as well. Ellis and
319 colleagues also observed up or down regulation of genes from the X-chromosome and
320 autosomes in testes of mice with deletions of Y heterochromatin using a microarray
321 approach (Ellis et al., 2005). Homology between UTRs of some of the above genes and the
322 108 ncRNAs (Table S2) further strengthens the hypothesis of putative regulation of genes
323 located elsewhere in the genome by Y chromosomal repeats.
324 piRNAs are known to regulate gene expression at the levels of both transcription and
325 translation. Our study also suggests regulation at both the transcriptional and post-
326 transcriptional levels by the piRNAs derived from the 108 transcript variants. Genes
327 corresponding to three proteins that were missing from the mutant sperm proteome, were
328 indeed transcribed in testis. The fact that there is comparable expression of calreticulin and
329 SPINK2 variant2 proteins in XY^{RIII} and XY^{RIII}qdel testes by western blot analysis, despite
330 significantly upregulated expression of these transcripts in testis, could indicate regulation
331 at the level of both transcription and translation for these genes. piRNAs regulate

translation in early embryos and gonads besides containing transposable elements (Aravin, Hannon, & Brennecke, 2007; Deng & Lin, 2002).

We also propose that the deregulated proteins identified in the current study are at least partially responsible for the sperm phenotype observed in XY^{RIII}qdel mice. For example, FABP9 that is upregulated in XY^{RIII}qdel spermatozoa, localizes to perforatorium, the subacrosomal region in spermatozoa with falciform head shapes (Korley, Pouresmaeili, & Oko, 1997; Oko & Morales, 1994; Pouresmaeili, Morales, & Oko, 1997). It is the most abundant protein of the perforatorium, and putatively has a role in shaping the unique sperm head structure of rodents (Oko & Morales, 1994; Pouresmaeili et al., 1997). The deregulated expression of Spink2 isoforms also could contribute to sperm head morphological abnormalities and reduced sperm motility; Kherraf and colleagues (Kherraf et al., 2017) observed grossly misshapen sperm heads and reduced motility in Spink2 knock out mice.

Calreticulin overexpressed in XY^{RIII}qdel spermatozoa is a calcium store associated with sperm functions such as hyperactivated motility, capacitation and acrosome reaction (Yanagimachi, 1982). The subsequent cascade of events could result in subfertility. Caldendrin yet another protein that is upregulated in XY^{RIII}qdel sperm (Bhattacharya et al., 2013), localizes to acrosome in rats and is considered to be a stimulus dependent regulator of calcium (Redecker, Kreutz, Bockmann, Gundelfinger, & Boeckers, 2003). SPINK2 variant3 localizes to the acrosome in mouse spermatozoa (our unpublished results). The physiological function of SPINK2 is believed to be blocking of deleterious degradation of proteins released by acrosin from spermatozoa during acrosome reaction (Moritz, Lilja, & Fink, 1991). Being a putative sperm acrosin inhibitor a role in fertility can be envisaged. *Sdf2l1* is an ER stress-inducible gene, induced by the unfolded protein response pathway

356 (Fukuda et al., 2001). The XY^{RIII}qdel spermatozoa may be more susceptible to stress induced
357 damages as they lack stress response protein, SDF2L1.
358 MAST a novel protein that is not visible in the XY^{RIII}qdel sperm proteome, but localizes to
359 both the acrosome and sperm tail, indicating a role in sperm motility and penetration of the
360 egg (Bhattacharya et al., 2013). Bioinformatic analysis of the novel protein Q9DAR0 predicts
361 it as a cilia related gene (McClintock, Glasser, Bose, & Bergman, 2008). Based on this
362 information putative role of this protein in sperm motility can be envisaged. SOD, a protein
363 upregulated in XY^{RIII}qdel, has been found to be positively associated with sperm count and
364 overall motility (Lu, Huang, & Lu, 2010). The irreversible conversion of androgens to
365 estrogens is catalyzed by aromatase transcribed by *Cyp19* gene (Simpson et al., 1994).
366 Biology behind the skewed sex ratio towards females in the progeny sired by XY^{RIII}qdel
367 males could be explained by the upregulated expression of aromatase in testes of these
368 mice. Acrosin plays a crucial role in acrosome exocytosis and egg zona pellucida penetration
369 (Yamagata et al., 1998). Spermatozoa lacking acrosin (acrosin$^{-/-}$) exhibit delayed fertilization
370 as both the processes of acrosome exocytosis and egg penetration are deferred (Adham,
371 Nayernia, & Engel, 1997; Mao & Yang, 2013).
372 Functions of Y chromosome have been elucidated using different deletions of the
373 chromosome in the past. Naturally occurring deletions in the euchromatic long arm of Y
374 chromosome in azoospermic men showed the involvement of this region in human male
375 infertility (Vogt, 1998). *Drosophila melanogaster* males with deletions of different regions of
376 the Y chromosome show absence of several sperm axoneme proteins (Goldstein, Hardy, &
377 Lindsley, 1982). Mice with partial or total deletions of Y heterochromatin show deregulation
378 of testicular gene expression and subfertility/sterility (Cocquet et al., 2009; Ellis et al., 2005).
379 Previous studies in the lab elucidated another example of an intronless Y-derived ncRNA

380 mediated regulation of an autosomal gene, *CDC2L2* via trans-splicing in human testis (Jehan
381 et al., 2007). The role of mouse Y heterochromatin in the current study therefore reveals a
382 novel pathway for the regulation of autosomal genes by Y chromosome, mediated by
383 piRNAs, in male reproduction. Therefore, consolidation of the observations from mouse and
384 human studies therefore shows that Y chromosome regulates autosomal genes expressed in
385 testis using distinct mechanisms in different species.
386 Comparative sperm proteomics analysis in our study portrays involvement of multiple
387 autosomal genes in subfertility. The regulation of autosomal gene expression appears to be
388 relaxed in sperms of Yq-deleted mice. This reflects a connection between the Y
389 chromosome and autosomes. In fact, as suggested by Piergentili, Y chromosome could be a
390 major modulator of gene expression (Piergentili, 2010). Our results seem to provide
391 explanation for some of the earlier classical observations of mice with different Y
392 chromosomal deletions exhibiting subfertility/sterility along with sperm morphological
393 abnormalities, fewer motile sperms, sex ratio skewed toward females etc. Similar
394 phenotypes are also observed in cross-species male sterile hybrids of *Drosophila* and mouse
395 (Albrechtova et al., 2012; Campbell, Good, Dean, Tucker, & Nachman, 2012; Heikkinen &
396 Lumme, 1998; Johnson, Hollocher, Noonburg, & Wu, 1993; Piergentili, 2010; Tao, Zeng, Li,
397 Hartl, & Laurie, 2003; Vigneault & Zouros, 1986; White, Stubbings, Dumont, & Payseur,
398 2012). Y-chromosome has also been implicated in the male sterility phenotype of these
399 interspecies hybrids (Campbell et al., 2012; Carvalho, Vaz, & Klaczko, 1997; Lamnissou,
400 Loukas, & Zouros, 1996; Vigneault & Zouros, 1986; Zouros, Lofdahl, & Martin, 1988). Thus,
401 the phenotypes observed in cross species male sterile hybrids and the Y-deletion mutants
402 are comparable. Introduction of Y-chromosomes into different genetic backgrounds of
403 *Drosophila* resulted in deregulated expression of hundreds of genes localizing to the X-

404 chromosome and autosomes (David et al., 2005; Lemos, Araripe, & Hartl, 2008). It has also
405 been proposed that incompatibility between the Y-chromosomes and different autosomes
406 could result in the hybrid dysgenesis of sperm related phenotypes observed in *Drosophila*
407 (Zouros et al., 1988). Zouros and colleagues also suggested the presence of epistatic
408 networks in interspecies hybrids, based on the fact that homospecific combination of alleles
409 at a given set of loci could sustain normal development, but heterospecific combinations
410 could not (Davis, Noonburg, & Wu, 1994; Lamnissou et al., 1996; Wu & Palopoli, 1994). This
411 early hypothesis seems to be amply supported by our study. Further, our results elucidate
412 the Y - derived piRNAs as the genetic basis of epistatic interactions between Y chromosome
413 and autosomes in mouse. Our results also suggest for the first time, the mechanism of
414 piRNA mediated regulation of autosomal genes involved in spermiogenesis and male
415 fertility. This is, to our knowledge is the first report on possible regulation of autosomal
416 genes involved in male fertility and spermiogenesis, mediated by Y-encoded small
417 RNAs/piRNAs in any species.
418 In brief, the XY^{RIII}qdel mutant strain of mouse, where there is a partial deletion of long arm
419 of the Y chromosome, exhibit sperm morphological and motility related aberrations and
420 subfertility (Conway et al., 1994). A comparative sperm proteomic profiling of the XY^{RIII} and
421 XY^{RIII}qdel mice captured few differentially expressed proteins that could partially account for
422 the aberrant sperm phenotype. Surprisingly, genes corresponding to the deregulated
423 proteins localized to autosomes and not to the deleted region of the Y chromosome. Earlier
424 we demonstrated an event of trans-splicing between a Y-ncRNA and a protein coding
425 autosomal mRNA in human testis for putative translational regulation. A search for the Y-
426 autosome connections in mouse led to the identification of novel ncRNAs from mouse Y
427 long arm that subsequently was shown to regulate the genes expressed in testis via piRNAs.

428 Thus, adopting a top-down approach we have established a novel mode of regulation of
429 autosomal genes expressed in mouse testis by the Y chromosome and the biology behind
430 the aberrant sperm phenotype in Y-deleted mice.
431 Finally, evolutionary impact of novel genetic interactions or regulatory mechanisms such as
432 those reported in this study could be significant. The generation of piRNAs from species-
433 specific repeats on mouse Y-chromosome that apparently regulate autosomal gene
434 expression in testis raises more questions in the field of speciation and evolution. Do
435 mutations in the Y chromosomal repeats collapse the poise of the species? Are species-
436 specific repeats on the Y chromosome the fulcrum on which rests the fine balance between
437 species identity and evolution?
438
439 **METHODS**
440 **Animals and Reagents**
441 The XY^{RIII} strain (wild type) and the XY^{RIII}qdel strain (Y-deletion mutant) of mice used in the
442 study was a gift from Prof. Paul S Burgoyne, MRC, UK. All the animals used in the
443 experiments were bred and reared in the in-house animal facility of our institute (CCMB), in
444 accordance with the guidelines from Indian Science Academy under CPCSEA (Committee for
445 the Purpose of Control and Supervision of Experimental Animals). This study was approved
446 by the Institutional Animal Ethics Committee (IAEC 65/28/2006). The recombinant construct
447 in pAAV-IRES-hrGFP from which MIWI protein was isolated, was a gift from Arvind Kumar,
448 CCMB, Hyderabad.
449 **Identification of cDNA using M34 (DQ907163):**
450 An amplified mouse testis cDNA library (Mouse testis MATCHMAKER cDNA Library –
451 Clonetech) was screened with the male specific ES cell EST (CA533654) with sequence

452 homology to the genomic clone DQ907163 (Fig. 1*E*) as per manufacturer's protocol. Mouse
453 testis cDNA Library was screened according to standard protocol at a stringency of 2 x SSC,
454 0.1% SDS at 65^0C for 10 min. A total of 2 x 10^5 colonies were screened to obtain 18 clones
455 after tertiary screening. Male-specificity of the positive clones from the library was
456 determined using Southern blots containing mouse male and female DNA. All 18 clones
457 gave the same sequence. The sequence was submitted to the NCBI database and is named
458 as *Pirmy* (DQ907162)**.**

459 **Identification of Splice Variants:** Total RNA (1 μg) isolated from brain and testes tissues
460 each of XY^{RIII} strain of mouse were reverse transcribed with the (SuperscriptII, Reverse
461 Transcriptase enzyme, Invitrogen), using oligo (dT) primers and random hexamers. The RT-
462 mixes were concentration equalized using GAPDH primers. DQ907162 was amplified after
463 two rounds of PCR using primers to the first and last exons. For the first round, forward
464 (GTGTGACAGGGTGGGGAATC) and reverse primers (TTCCTGAAGATAGCACTTGTG), and the
465 following conditions were used - initial denaturation 95^0C, 1min, cycle denaturation 95^0C, 1
466 min, annealing 62^0C, 30 sec and extension 72^0C for 2min (35cycles), final extension 72^0C for
467 7 min. The second round of amplification was done using nested primers
468 GAGGACCGTATTCATGGAAGAG (forward) and GCAAATGGCTCACATCAGTGG (reverse) using
469 initial denaturation at 95^0C for 1 min, cycle denaturation at 95^0C, 1 min, annealing at 66^0C
470 for 30 sec and extension at 72^0C for 2 min (38 cycles) and final extension at 72^0C for 7 min.
471 Annealing temperatures up to 70^0C yielded multiple products. The multiple products
472 obtained after two rounds of PCR from both testis and brain (Fig. 2*B*) were cloned into pCR
473 TOPO vector using Invitrogen TOPO TA cloning kit. Approximately 1000 clones were
474 sequenced on 3730 DNA Analyser (ABI Prism) using sequencing kit BigDye Terminator V3.1
475 Cycle Sequencing kit. These yielded 108 unique transcripts. BLAST analysis of these

476 transcripts against the genomic sequences (GRCm39) at a stringency of >97% localised 80 of
477 these to a single locus on the Y chromosome (4341127-4381724) and the rest to multiple
478 sites on the Y.
479 Therefore, the 80 transcripts that localised to the single locus are splice variants of *Pirmy*.
480 **Sperm Proteome Analysis:** Sperm lysates (1 mg of cell weight per 5 µl of lysis buffer – [Urea
481 8 M, CHAPS 4% (w/v), Tris 40 mM, Biolyte (3-10) 0.2% (w/v) and TBP (1µl/100 µl]). The
482 suspension was incubated on ice for one hour to allow buffer to permeabilise and lyse the
483 sample. Further, the sample was briefly sonicated on ice. The lysate was centrifuged for 15
484 min at 13,000 rpm at 4^0C. The supernatant was collected and was further taken for ultra-
485 centrifugation at 55,000 rpm for one hour at 4^0C. The clear lysate was collected in fresh
486 tube and PMSF added to a final concentration of 1mM. The protein concentration in the cell
487 lysate was estimated by Bicinchoninic acid assay (Pierce, Rockford IL), following
488 manufacturer's instructions in a micro titre plate. BSA was used as the standard for
489 estimation. The proteins were separated in the first dimension on 4-7 and 5-8 IPG strips
490 (Bio-Rad). The strips were then loaded onto 8-20% gradient PAGE for separation on the
491 second dimension. Protein spots were visualized by Coomassie Blue staining. Spot to spot
492 matches were done to identify differences. Analysis of five sets of gels after normalization
493 with control spots using PDQUEST software version 6.0 (Bio-Rad) identified the differential
494 proteins. Measuring the optical density of these differentially expressed proteins in arbitrary
495 units validated the quantitative differences (Fig. S7). These values were subjected to
496 nonparametric Kruskal-Wallis H test and the levels of confidence determined by the Chi-
497 squared test (75-95% degrees of confidence). Trypsin digested spots were processed to
498 obtain the protein tags by MS analysis on Hybrid Quadrupole TOF mass spectrometer (API
499 QSTAR PULSAR i, PE SCIEX).

500 **Small RNA Isolation**

501 Total RNA was extracted from male kidney, testis and brain of mice using TRIZOL reagent
502 (Invitrogen). Total RNA was denatured at 65°C for 10 min, incubated with 10% PEG-8000
503 (Sigma-Aldrich) and 5 M NaCl for 30 min on ice and centrifuged at 7,000 rpm for 7-10 min.
504 The supernatant containing small RNA was precipitated overnight with 3 volumes of
505 absolute alcohol and centrifuged at 13,000 rpm for 30 min. The small RNA pellet was
506 washed with 80% ethanol and resuspended in RNase free Gibco water. The quality of small
507 RNA was checked on 12% Urea PAGE (Biorad) and quantitated using Nanodrop V-1000
508 (Thermo Fisher Scientific).

509 **Small RNA Northern Blotting**

510 20-50 µg of small RNA from each tissue was resolved on a 12% Urea PAGE gel in 0.5 x TBE
511 running buffer and transferred onto Hybond N^{+} membrane. Decade marker (Ambion) was
512 labelled and loaded according to the manufacturer's instructions. 10-25 µM of each LNA-
513 oligonucleotide probe (Exiqon), was end labelled for use as probes (hybridization buffer - 5 x
514 SSC, 5 x Denhardt's and 1% SDS). Blots were hybridized at 37^{0}C and washed from 37^{0}C to
515 65^{0}C in 2 x SSC, 0.2% SDS depending on the intensity of the signal. U6 was used as the
516 loading control. Fig. S6 shows the location of the LNA (locked nucleic acid) probes used for
517 small RNA northern blots on the corresponding splice variants.

518 **Real-time PCR analysis**

519 **Copy number estimation of *Pirmy***

520 Genomic DNA was isolated from testes of XY^{RIII} and XY^{RIII}qdel mice using phenol-chloroform
521 method and quantified using a Nanodrop (NANODROP 2000, ThermoScientific).
522 Quantitative Real-time PCR (LightCycler 480, Roche) was performed using SYBR green
523 master mix (Qiagen) with 2 ng of genomic DNA and a primer concentration of 200 nM per

524 reaction. The primers used were as follows: *Pirmy* (exon 7)*:* forward 5′-GTG CGG TTG TGA
525 AGG TGT TC– 3′, reverse 5′-CCT CCA CCT TCC ATT CAC CC-3′; *Gapdh:* forward 5′-ACG GGA
526 AGC TCA CTG GCA TGG -3′, reverse 5′-CAA CAG CGA CAC CCA CTC CTC–3′. PCR conditions
527 included an initial denaturation for 5 min at 95°C followed by 45 cycles of denaturation at
528 95°C for 10 secs, annealing at 58°C for 20 sec and elongation at 72°C for 30 sec. The
529 amplification of specific product was confirmed by melting curve profile (cooling the sample
530 to 65°C for 1 min and heating slowly with an increase in temperature of 5°C at each step till
531 95°C, with continuous measurement of fluorescence). The relative fold change in *Pirmy* was
532 analyzed based on Livak method ($2^{-\Delta\Delta Ct}$).

533 **Real-time PCR analysis of differentially expressed genes**

534 The total RNA was extracted from testes using Trizol (Invitrogen). One µg of RNA was
535 reverse transcribed to cDNA using Verso cDNA synthesis kit (ThermoScientific). qPCR was
536 performed using SYBR green master mix (Qiagen) and analysed in Roche Light Cycler LC480.
537 The primer sequences corresponding to mouse cDNA used were as follows: forward 5′-CGA
538 GGG CCA GAC AGG GAT TG–3′ and reverse 5′-CCC ATA GAC AGA GGA CAT CAG- 3′ for Riken
539 cDNA 1700001L19; forward 5′-ACT TCC CGT CGC CGC TAT C-3′ and reverse 5′-TGA CCG ACA
540 GGA ACA CAG AGG-3′ for *Sdf2l1*; forward 5′-CAG CAT CGA GCA GAA GTA TAA GC-3′ and
541 reverse 5′-TGG GTG GAG TTA TTG CAG TAG-3′ for *Mast*; forward 5′-GGA AAC CAC GTC AAA
542 TTG -3′ and reverse 5′-GGT GAT GAG GAA ATT GTC-3′ for Calreticulin; forward 5′-GGC TAC
543 TTG ACC ACT GC-3′ and reverse 5′-TTT GAG AAT CGG AAG AGT C-3′ for *Spink2* Variant 2;
544 forward 5′-TTC CGA ACA CCA GAC TG-3′ and reverse 5′-ATG GCT ACC GTC CTC C-3′ for
545 *Spink2* Variant 3; forward 5′-TGA AGT CGC AGG AGA CAA CCT-3′ and reverse 5′-ATG GCC
546 TTC CGT GTT CCT A-3′ for *Gapdh*. PCR conditions included an initial denaturation for 5 min
547 at 95°C followed by 45 cycles of denaturation at 95°C for 10 secs, annealing at 58°C for 20

548 sec and elongation at 72°C for 30 sec. The amplification of specific product was confirmed
549 by melting curve profile. The relative fold change in expression was estimated based on
550 Livak method ($2^{-\Delta\Delta Ct}$).

551 **Electrophoretic mobility shift assay** RNA oligonucleotides corresponding to
552 GAAGCAGAUGAGUAUAUG from *Sod* and UCAUUGGACAUAAACUGAAUUUUCCA from the
553 gene for hypothetical protein spot A (Q9DAR0) were end labelled with γ-^{32}P ATP and column
554 purified using G-25 Sephadex (Sigma-Aldrich) and quantitated on a scintillation counter.
555 EMSA reactions were set up in a total volume of 25 µl using binding buffer (20 mM HEPES, 3
556 mM MgCl2, 40 mM KCl, 5% Glycerol, 2 mM DTT and RNase inhibitor (4U)), with MIWI
557 protein (5 µg per reaction). MIWI was over-expressed from a recombinant construct in
558 pAAV-IRES-hrGFP vector and purified using the FLAG tag. Competitors i.e., unlabeled
559 oligonucelotides (30 x concentration of hot oligo), MIWI antibody (90 ng) and Argonaute 3
560 antibody (100 ng) were added to the reaction, incubated for 1 h on ice, before addition of
561 radio-labeled oligonucleotide (7000-10,000 cpm) and the entire mix was incubated on ice
562 for another 30 min. EMSA was done on 5% native PAGE and image captured using FUJI
563 phosphor Imager (FUJIFILM FLA-3000). A known piRNA, piR1, was used as the positive
564 control, Argonaute 3 antibody served as the antibody control.

565 **Luciferase assay**

566 Either Luciferase gene alone or luciferase along with the UTR was cloned into pcDNA3.1
567 expression vector for assaying the effect of antagopirs (Figure 7H) on Luciferase expression.
568 Co-transfection experiments were done using the GC-1spg cell line (ATCC CRL-2053) and
569 lipofectamine 2000 (Invitrogen) using protocols specified by the manufacturer. Cells were
570 seeded in 48-well plates, 24 hrs prior to transfection to obtain approximately 80%
571 confluency. Each well was transfected with 50 ng of pcDNA3.1 plasmid containing either

572 Luciferase gene alone or along with the cloned UTRs, 50 ng of the β-gal plasmid and varying
573 concentrations of oligonucleotides complementary to the piRNA along with 0.5µl of
574 lipofectamine 2000 in antibiotic and serum free DMEM (GIBCO). The complementary
575 oligonucleotide to the piRNA has been designated as antagopirs. The antagopirs (*Sod* -
576 5'GAAGCAGAUGAGUAUAUG3'; *PLA2G12B* - 5'CCAAACUGUUGGAAGAAGGAAU3') were
577 procured as RNA oligonucleotides from Eurofins Genomics India Pvt. Ltd, Bangalore, India.
578 Different concentrations of antagopirs (0 nM, 10 nM, 20 nM and 40 nM/ well) were tested
579 in the assay for their effect on the UTRs. Five hrs post transfection, the medium was
580 replaced with complete growth medium. The cell extracts were prepared 24 hrs post
581 transfection using Reporter Lysis Buffer (Promega) and assayed for Luciferase activity in
582 EnSpire 2300 multimode plate reader (Perkin Elmer). The Luciferase activity was normalized
583 using β galactosidase. Three independent sets of experiments were done in triplicates.

584 **DATA ACCESS**

585 All sequences from this study have been deposited in NCBI database. With the accession
586 numbers: DQ907162.1, FJ541075-FJ541181, Q7TPM5, Q9DAR0, Q8BMY7.

587 **ACKNOWLEDGEMENTS**

588 We would like to dedicate this manuscript to Prof. Lalji Singh and Prof. Burgoyne both of
589 whom have left us for their heavenly abode and whom we miss greatly at this point of time.
590 We gratefully acknowledge the gift of the XY^{RIII} and XY^{RIII}qdel mice by Prof. Paul S Burgoyne,
591 MRC, Edinburgh, UK. This study could not have been done without the gift of these mice.
592 The GC-1spg cell line was gifted by Prof. MRS Rao, JNCASR, Bangalore, India. We thank Dr.
593 Dinesh Kumar and Professor B. K. Thelma for reading the manuscript and giving useful
594 inputs and Mr. Sivarajan Karunanithi for help with partial *in silico* analysis.

The funding by Department of Science and Technology, India (SP/SO/B70/2001) and Department of Biotechnology, India (BT/PR 10707/AGR/36/596/2008), intramural funding from Council of Scientific and Industrial Research (CSIR), India to RAJ and fellowships by CSIR, India to HMR, RB are acknowledged.

Author contributions

HMR, KM did experiments and partial bioinformatics analysis. RB, ZJ, PA, NMP, VMD, MS, BS, JLA, SMT, RRD, SMR performed experiments. ST, AC, SK contributed to *in silico* analysis. NR helped with confocal imaging. LS gave the probe and inputs. RAJ conceived, guided the work and wrote the manuscript.

DISCLOSURE DECLARATION

The authors declare that they have no competing interests

REFERENCES:

Adham, I. M., Nayernia, K., & Engel, W. (1997). Spermatozoa lacking acrosin protein show delayed fertilization. *Mol Reprod Dev, 46*(3), 370-376. doi:10.1002/(SICI)1098-2795(199703)46:3<370::AID-MRD16>3.0.CO;2-2

Albrechtova, J., Albrecht, T., Baird, S. J., Macholan, M., Rudolfsen, G., Munclinger, P., . . . Pialek, J. (2012). Sperm-related phenotypes implicated in both maintenance and breakdown of a natural species barrier in the house mouse. *Proc Biol Sci, 279*(1748), 4803-4810. doi:10.1098/rspb.2012.1802

Aravin, A. A., Hannon, G. J., & Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science, 318*(5851), 761-764. doi:10.1126/science.1146484

Bajpai, A., Sridhar, S., Reddy, H. M., & Jesudasan, R. A. (2007). BRM-Parser: a tool for comprehensive analysis of BLAST and RepeatMasker results. *In Silico Biol, 7*(4-5), 399-403. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18391232

Bellott, D. W., Hughes, J. F., Skaletsky, H., Brown, L. G., Pyntikova, T., Cho, T. J., . . . Page, D. C. (2014). Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature, 508*(7497), 494-499. doi:10.1038/nature13206

Bhattacharya, R., Devi, M. S., Dhople, V. M., & Jesudasan, R. A. (2013). A mouse protein that localizes to acrosome and sperm tail is regulated by Y-chromosome. *BMC Cell Biol, 14*(1), 50. doi:10.1186/1471-2121-14-50

Bishop, C. E., & Hatat, D. (1987). Molecular cloning and sequence analysis of a mouse Y chromosome RNA transcript expressed in the testis. *Nucleic Acids Res, 15*(7), 2959-2969. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3562242

Burgoyne, P. S., Mahadevaiah, S. K., Sutcliffe, M. J., & Palmer, S. J. (1992). Fertility in mice requires XY pairing and a Y-chromosomal "spermiogenesis" gene mapping to the long arm. *Cell, 71*(3), 391-398.

Burgoyne, P. S., & Mitchell, M. J. (2007). The roles of mouse Y chromosome genes in spermatogenesis. *Y chromosome and male germ cell biology. Hackensack (New Jersey): World Scientific Publishers*, 27-45.

Campbell, P., Good, J. M., Dean, M. D., Tucker, P. K., & Nachman, M. W. (2012). The contribution of the Y chromosome to hybrid male sterility in house mice. *Genetics, 191*(4), 1271-1281. doi:10.1534/genetics.112.141804

Carvalho, A. B., Vaz, S. C., & Klaczko, L. B. (1997). Polymorphism for Y-linked suppressors of sex-ratio in two natural populations of Drosophila mediopunctata. *Genetics, 146*(3), 891-902. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9215895

Choi, E., Lee, J., Oh, J., Park, I., Han, C., Yi, C., . . . Cho, C. (2007). Integrative characterization of germ cell-specific genes from mouse spermatocyte UniGene library. *BMC Genomics, 8*, 256. doi:10.1186/1471-2164-8-256

Cocquet, J., Ellis, P. J., Yamauchi, Y., Mahadevaiah, S. K., Affara, N. A., Ward, M. A., & Burgoyne, P. S. (2009). The multicopy gene Sly represses the sex chromosomes in the male mouse germline after meiosis. *PLoS Biol, 7*(11), e1000244. doi:10.1371/journal.pbio.1000244

Conway, S. J., Mahadevaiah, S. K., Darling, S. M., Capel, B., Rattigan, A. M., & Burgoyne, P. S. (1994). Y353/B: a candidate multiple-copy spermiogenesis gene on the mouse

Y chromosome. *Mamm Genome, 5*(4), 203-210. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7516747

David, J. R., Araripe, L. O., Chakir, M., Legout, H., Lemos, B., Petavy, G., . . . Moreteau, B. (2005). Male sterility at extreme temperatures: a significant but neglected phenomenon for understanding Drosophila climatic adaptations. *Journal of evolutionary biology, 18*(4), 838-846. doi:10.1111/j.1420-9101.2005.00914.x

Davis, A. W., Noonburg, E. G., & Wu, C. I. (1994). Evidence for complex genic interactions between conspecific chromosomes underlying hybrid female sterility in the Drosophila simulans clade. *Genetics, 137*(1), 191-199. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8056310

Deng, W., & Lin, H. (2002). miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Developmental cell, 2*(6), 819-830. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12062093

Ellis, P. J., Clemente, E. J., Ball, P., Toure, A., Ferguson, L., Turner, J. M., . . . Burgoyne, P. S. (2005). Deletions on mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids. *Hum Mol Genet, 14*(18), 2705-2715. doi:ddi304 [pii]

10.1093/hmg/ddi304

Ellis, P. J., Ferguson, L., Clemente, E. J., & Affara, N. A. (2007). Bidirectional transcription of a novel chimeric gene mapping to mouse chromosome Yq. *BMC Evol Biol, 7*, 171. doi:1471-2148-7-171 [pii]

10.1186/1471-2148-7-171

Fukuda, S., Sumii, M., Masuda, Y., Takahashi, M., Koike, N., Teishima, J., . . . Kamiya, K. (2001). Murine and human SDF2L1 is an endoplasmic reticulum stress-inducible gene and encodes a new member of the Pmt/rt protein family. *Biochem Biophys Res Commun, 280*(1), 407-414. doi:10.1006/bbrc.2000.4111

Girard, A., Sachidanandam, R., Hannon, G. J., & Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature, 442*(7099), 199-202. doi:10.1038/nature04917

Goldstein, L. S., Hardy, R. W., & Lindsley, D. L. (1982). Structural genes on the Y chromosome of Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America, 79*(23), 7405-7409. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6818544

Heikkinen, E., & Lumme, J. (1998). The Y chromosomes of Drosophila lummei and D. novamexicana differ in fertility factors. *Heredity, 81 (Pt 5)*, 505-513. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9881450

Jehan, Z., Vallinayagam, S., Tiwari, S., Pradhan, S., Singh, L., Suresh, A., . . . Jesudasan, R. A. (2007). Novel noncoding RNA from human Y distal heterochromatic block (Yq12) generates testis-specific chimeric CDC2L2. *Genome Res, 17*(4), 433-440. doi:10.1101/gr.5155706

Johnson, N. A., Hollocher, H., Noonburg, E., & Wu, C. I. (1993). The effects of interspecific Y chromosome replacements on hybrid sterility within the Drosophila simulans clade. *Genetics, 135*(2), 443-453. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8244006

Kherraf, Z. E., Christou-Kent, M., Karaouzene, T., Amiri-Yekta, A., Martinez, G., Vargas, A. S., . . . Arnoult, C. (2017). SPINK2 deficiency causes infertility by inducing sperm defects in heterozygotes and azoospermia in homozygotes. *EMBO Mol Med, 9*(8), 1132-1149. doi:10.15252/emmm.201607461

Korley, R., Pouresmaeili, F., & Oko, R. (1997). Analysis of the protein composition of the mouse sperm perinuclear theca and characterization of its major protein constituent. *Biol Reprod, 57*(6), 1426-1432. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9408250

Kotula-Balak, M., Grzmil, P., Styrna, J., & Bilinska, B. (2004). Immunodetection of aromatase in mice with a partial deletion in the long arm of the Y chromosome. *Acta histochemica, 106*(1), 55-64. doi:10.1016/j.acthis.2003.09.001

Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L. G., Minx, P. J., Cordum, H. S., Waterston, R. H., . . . Page, D. C. (2001). The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet, 29*(3), 279-286. doi:10.1038/ng757

Lamnissou, K., Loukas, M., & Zouros, E. (1996). Incompatibilities between Y chromosome and autosomes are responsible for male hybrid sterility in crosses between Drosophila virilis and Drosophila texana. *Heredity, 76 (Pt 6)*, 603-609. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8801221

Lemos, B., Araripe, L. O., & Hartl, D. L. (2008). Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science, 319*(5859), 91-93. doi:10.1126/science.1148861

Lu, J. C., Huang, Y. F., & Lu, N. Q. (2010). [WHO Laboratory Manual for the Examination and Processing of Human Semen: its applicability to andrology laboratories in China]. *Zhonghua Nan Ke Xue, 16*(10), 867-871. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21243747

Mao, H. T., & Yang, W. X. (2013). Modes of acrosin functioning during fertilization. *Gene, 526*(2), 75-79. doi:10.1016/j.gene.2013.05.058

McClintock, T. S., Glasser, C. E., Bose, S. C., & Bergman, D. A. (2008). Tissue expression patterns identify mouse cilia genes. *Physiol Genomics, 32*(2), 198-206. doi:10.1152/physiolgenomics.00128.2007

Moretti, C., Serrentino, M. E., Ialy-Radio, C., Delessard, M., Soboleva, T. A., Tores, F., . . . Cocquet, J. (2017). SLY regulates genes involved in chromatin remodeling and interacts with TBL1XR1 during sperm differentiation. *Cell Death Differ, 24*(6), 1029-1044. doi:10.1038/cdd.2017.32

Moritz, A., Lilja, H., & Fink, E. (1991). Molecular cloning and sequence analysis of the cDNA encoding the human acrosin-trypsin inhibitor (HUSI-II). *FEBS Lett, 278*(1), 127-130. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1704312

Oko, R., & Morales, C. R. (1994). A novel testicular protein, with sequence similarities to a family of lipid binding proteins, is a major component of the rat sperm perinuclear theca. *Dev Biol, 166*(1), 235-245. doi:10.1006/dbio.1994.1310

Pang, K. C., Stephen, S., Engstrom, P. G., Tajul-Arifin, K., Chen, W., Wahlestedt, C., . . . Mattick, J. S. (2005). RNAdb--a comprehensive mammalian noncoding RNA database. *Nucleic acids research, 33*(Database issue), D125-130. doi:10.1093/nar/gki089

Piergentili, R. (2010). Multiple roles of the Y chromosome in the biology of Drosophila melanogaster. *The Scientific World Journal, 10*, 1749-1767.

Pouresmaeili, F., Morales, C. R., & Oko, R. (1997). Molecular cloning and structural analysis of the gene encoding PERF 15 protein present in the perinuclear theca of the rat spermatozoa. *Biol Reprod, 57*(3), 655-659. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9283004

Prado, V. F., Lee, C. H., Zahed, L., Vekemans, M., & Nishioka, Y. (1992). Molecular characterization of a mouse Y chromosomal repetitive sequence that detects

transcripts in the testis. *Cytogenet Cell Genet, 61*(2), 87-90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1395731

Redecker, P., Kreutz, M. R., Bockmann, J., Gundelfinger, E. D., & Boeckers, T. M. (2003). Brain synaptic junctional proteins at the acrosome of rat testicular germ cells. *J Histochem Cytochem, 51*(6), 809-819. doi:10.1177/002215540305100612

Reynard, L. N., Cocquet, J., & Burgoyne, P. S. (2009). The multi-copy mouse gene Sycp3-like Y-linked (Sly) encodes an abundant spermatid protein that interacts with a histone acetyltransferase and an acrosomal protein. *Biol Reprod, 81*(2), 250-257. doi:10.1095/biolreprod.108.075382

Riel, J. M., Yamauchi, Y., Ruthig, V. A., Malinta, Q. U., Blanco, M., Moretti, C., . . . Ward, M. A. (2019). Rescue of Sly Expression Is Not Sufficient to Rescue Spermiogenic Phenotype of Mice with Deletions of Y Chromosome Long Arm. *Genes (Basel), 10*(2). doi:10.3390/genes10020133

Riel, J. M., Yamauchi, Y., Sugawara, A., Li, H. Y., Ruthig, V., Stoytcheva, Z., . . . Ward, M. A. (2013). Deficiency of the multi-copy mouse Y gene Sly causes sperm DNA damage and abnormal chromatin packaging. *J Cell Sci, 126*(Pt 3), 803-813. doi:10.1242/jcs.114488

Schultz, N., Hamra, F. K., & Garbers, D. L. (2003). A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc Natl Acad Sci U S A, 100*(21), 12201-12206. doi:10.1073/pnas.1635054100

Simpson, E. R., Mahendroo, M. S., Means, G. D., Kilgore, M. W., Hinshelwood, M. M., Graham-Lorence, S., . . . et al. (1994). Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis. *Endocr Rev, 15*(3), 342-355. doi:10.1210/edrv-15-3-342

Singh, L., Panicker, S. G., Nagaraj, R., & Majumdar, K. C. (1994). Banded krait minor-satellite (Bkm)-associated Y chromosome-specific repetitive DNA in mouse. *Nucleic acids research, 22*(12), 2289-2295. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8036156

Soh, Y. Q., Alfoldi, J., Pyntikova, T., Brown, L. G., Graves, T., Minx, P. J., . . . Page, D. C. (2014). Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell, 159*(4), 800-813. doi:10.1016/j.cell.2014.09.052

Styrna, J., Bili, B., & Krzanowska, H. (2002). The effect of a partial Y chromosome deletion in B10. BR-Ydel mice on testis morphology, sperm quality and efficiency of fertilization. *Reproduction, Fertility and Development, 14*(2), 101-108.

Styrna, J., Klag, J., & Moriwaki, K. (1991). Influence of partial deletion of the Y chromosome on mouse sperm phenotype. *Journal of reproduction and fertility, 92*(1), 187-195. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2056490

Tao, Y., Zeng, Z. B., Li, J., Hartl, D. L., & Laurie, C. C. (2003). Genetic dissection of hybrid incompatibilities between Drosophila simulans and D. mauritiana. II. Mapping hybrid male sterility loci on the third chromosome. *Genetics, 164*(4), 1399-1418. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12930748

Tiepolo, L., & Zuffardi, O. (1976). Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum Genet, 34*(2), 119-124. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1002136

Toure, A., Clemente, E. J., Ellis, P., Mahadevaiah, S. K., Ojarikre, O. A., Ball, P. A., . . . Affara, N. A. (2005). Identification of novel Y chromosome encoded transcripts by testis

transcriptome analysis of mice with deletions of the Y chromosome long arm. *Genome Biol, 6*(12), R102. doi:gb-2005-6-12-r102 [pii]

10.1186/gb-2005-6-12-r102

Toure, A., Szot, M., Mahadevaiah, S. K., Rattigan, A., Ojarikre, O. A., & Burgoyne, P. S. (2004). A new deletion of the mouse Y chromosome long arm associated with the loss of Ssty expression, abnormal sperm development and sterility. *Genetics, 166*(2), 901-912. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15020475

Vigneault, G., & Zouros, E. (1986). The genetics of asymmetrical male sterility in Drosophila mojavensis and Drosophila arizonensis hybrids: interactions between the Y-chromosome and autosomes. *Evolution*, 1160-1170.

Vogt, P. H. (1998). Human chromosome deletions in Yq11, AZF candidate genes and male infertility: history and update. *Molecular human reproduction, 4*(8), 739-744. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9733430

Wegmann, D., Dupanloup, I., & Excoffier, L. (2008). Width of gene expression profile drives alternative splicing. *PLoS One, 3*(10), e3587. doi:10.1371/journal.pone.0003587

White, M. A., Stubbings, M., Dumont, B. L., & Payseur, B. A. (2012). Genetics and evolution of hybrid male sterility in house mice. *Genetics, 191*(3), 917-934. doi:10.1534/genetics.112.140251

Wu, C. I., & Palopoli, M. F. (1994). Genetics of postmating reproductive isolation in animals. *Annual review of genetics, 28*, 283-308. doi:10.1146/annurev.ge.28.120194.001435

Xiao, P., Tang, A., Yu, Z., Gui, Y., & Cai, Z. (2008). Gene expression profile of 2058 spermatogenesis-related genes in mice. *Biol Pharm Bull, 31*(2), 201-206. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18239273

Yamagata, K., Murayama, K., Okabe, M., Toshimori, K., Nakanishi, T., Kashiwabara, S., & Baba, T. (1998). Acrosin accelerates the dispersal of sperm acrosomal proteins during acrosome reaction. *J Biol Chem, 273*(17), 10470-10474. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9553106

Yanagimachi, R. (1982). Requirement of extracellular calcium ions for various stages of fertilization and fertilization related phenomena in the hamster. *Gamete Research, 5*(4), 323-344. doi:10.1002/mrd.1120050404

Zouros, E., Lofdahl, K., & Martin, P. (1988). Male hybrid sterility in Drosophila: interactions between autosomes and sex chromosomes in crosses of D. mojavensis and D. arizonensis. *Evolution*, 1321-1331.

FIGURE LEGENDS

Figure 1. Analysis of M34 (DQ907163) and identification of a novel noncoding RNA. (*A*) Localization by FISH of the genomic clone, M34 to Mouse Y long arm in multiple copies spanning its entire length. (*B*) Mouse genome map view of M34 BLAST hits, showing Y-chromosomal localization further indicating male-specificity of these repeats. (*C*) Fluorescent in situ hybridization (FISH) using M34 shows intense signals in adult mouse

testis. (*D*) Hybridization of M34 onto RNase treated testis sections does not elicit signals, indicating that the signals in panel C are due to the presence of M34 derived RNA (*E*) Sequence analysis of the 9.5kb M34 shows presence of incomplete copies of different repeats like long terminal repeats (LTRs), long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), endogenous retroviral sequences (ERVK) and simple sequence repeats in both direct and reverse orientations in the clone. ESTs matching to M34 are marked as dotted arrows at the 3'end. (*F*) Shows the Y-specific localization of *Pirmy* cDNA on a mouse metaphase spread by FISH. (*G*) Partial homology between *Sly* and *Pirmy* (DQ907162) indicating identification of a novel cDNA. Homology region is highlighted in green rectangles. (*H*) Shows the consensus splice signal sequences at all intron- exon junctions. Since DQ907162 localizes to the complementary strand, the splice site consensus is seen as CT/AC instead AG/GT (see also Data Sheet).

Figure 2. Identification of multiple splice-variants of *Pirmy***.** Autoassembler program of ABI-Prism identified cDNAs that differed from one another. (*A*) Color coded line diagram showing extensive alternative splicing of *Pirmy*. The splice variants depicted here localise to Y: 4341127-4381724 (GRC m39). Each exon is represented by the same colour in different isoforms. Sizes of the exons are to scale. Top two lines show the representation of all exons present at this locus according to their order in the genomic sequence as e1, e2 etc. Line 2 indicates the nucleotide positions of each exon in a scenario where all the exons present. The exons have been arranged in linear order. (*B*) RT-PCR amplification of *Pirmy* showed multiple amplicons in both testis and brain, with many more amplicons in testis compared to brain. The RT-PCR products from brain and testis were cloned and sequenced to identify the splice variants. NTC is the non-template control. (*C*) BLAST analysis against mouse genome localizes the splice variants of *Pirmy* to NT_166343.2.

Figure 3. Localization of 28 transcript variants. (*A*) Top two lines show the representation of all exons present together according to their order in the transcript variants as e1, e2 etc. Putative nucleotide positions in a scenario wherein all the exons are present are indicated. Different splice variants have been arranged in linear order. Exons in dashed lines are specific to these 28 transcript variants, whereas other exons are common to *Pirmy* splice variants (Figure 2) and the transcript variants.

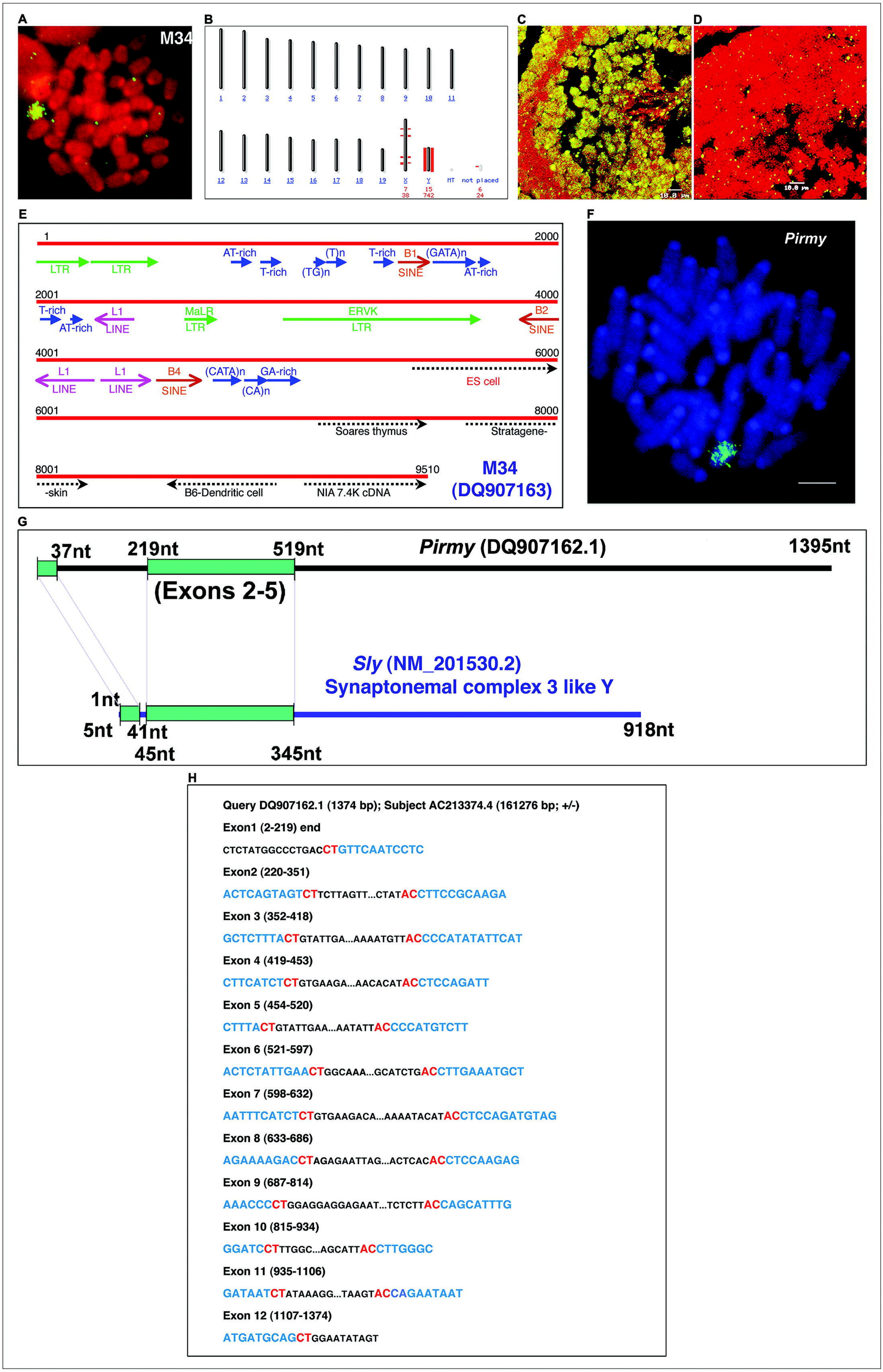
Figure 4. Comparison of splice-variants from testis and brain. The sequential exons from *Pirmy* splice-variants in testis have been represented in an Excel data sheet. Highlighted in red are the splice variants that are transcribed from male brain, which were also found in testis. The representation of e1, e2 etc. and the nucleotide positions is same as in Figure 2A.

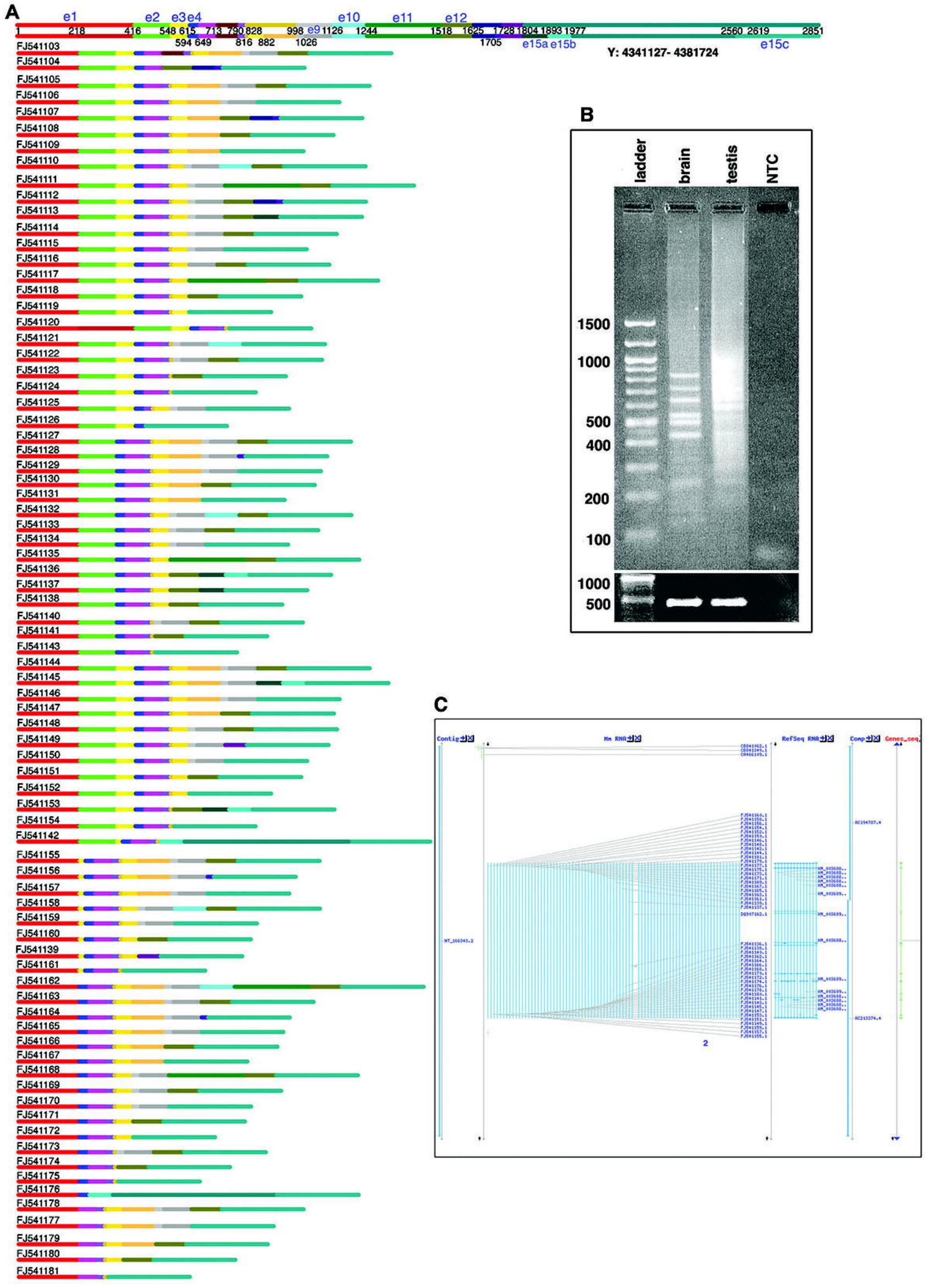
Figure 5. Sperm proteins are deregulated in XY^{RIII}qdel mice. (*A*) Sperm lysates from the wild type XY^{RIII} strain and the mutant XY^{RIII}qdel mice were separated by 2D-PAGE in the pI ranges of 4-7 and 5-8 on 8-20% gradient gels. Five differentially expressed proteins (D1 to D5) were identified by Mass Spectrometry analysis in the 4-7 pI range, of which 4 were upregulated (upward arrow) and one downregulated (downward arrow). Three proteins (A, B & C) were not detectable in the 5-8 pI range in XY^{RIII}qdel compared to the XY^{RIII} sperm lysate. (*B*) The list of all the differentially expressed proteins identified in the proteomics screen is given in the table. The genes corresponding to all the differentially expressed proteins in XY^{RIII}qdel localized to autosomes. (*C*) confirms the expression levels of four proteins (D1, D4, D5 and C) identified on 2D-gels by western blotting; the differential expression is observed in sperms but not in testis (RT - XY^{RIII} testis, YT - XY^{RIII}qdel testis, RS - XY^{RIII} sperm, YS - XY^{RIII}qdel sperm). The lower sub-panel for all blots is the loading control using Tubulin.

Figure 6. Localization of *Pirmy* homologous sequences to UTRs of deregulated genes. Panel (*A*) shows the UTR regions of the deregulated genes with the sequences homologous to *Pirmy* highlighted in red. Both +/+ and +/- homologies are observed. The splice isoforms of *Pirmy* are indicated in brown and the gene names in green color. (*B*) Couple of transcripts (aromatase and caldendrin) were identified independent of the proteomics screen (*C*) acrosin (literature survey) also harbor homologies to *Pirmy.*

Figure 7. Identification of piRNAs in the transcript variants. (*A*) Short stretches of homology identified in the UTRs by BLAST analysis against UTR database are represented on the set of 28 transcript variants. The boxes highlighted in black match within exons and boxes highlighted in red match at the exon-exon junctions. Line 1 indicates the nucleotide positions as in Fig. 3*A*. (*B*) Pie chart shows ~81% of the matches in exons and ~19% at exon-exon junctions. (*C*) Use of sequences homologous to the 3' UTRs of a hypothetical protein (SpotA4), *Sod, Bche, PLA2G12B* and *Spn* as probes on small RNA northern blots shows testis-specific signals of ~30 nt size (indicated by arrows), which correspond to the size of piRNAs. Control blot using a scrambled oligonucleotide as probe shows no signal at ~30 nt size. (*D*) Hybridization using probes from the sense (S) and antisense (AS) strands of *Sod* and *Bche* shows differential transcription from the two strands under the same conditions. Lower panels (C, D) show loading control using U6 probe. (*E, F*) EMSA using chemically synthesized RNA oligonucleotide sequences from *FJ541077* (E) and *FJ541118* (F) that have homology to the UTRs of the genes of hypothetical protein spot A (A4) and *Sod* respectively. These oligonucleotides and piR1 (Girard et al. 2006), a known piRNA, showed the shift in mobility with recombinant MIWI protein. The gel-shifts obtained with RNA-oligonucleotides from FJ541077 (spotA4) and FJ541118 (*Sod*) were competed out by cold piR1 and vice versa, i.e., the gelshift observed with piR1 was competed out by cold RNA-oligonucleotides from

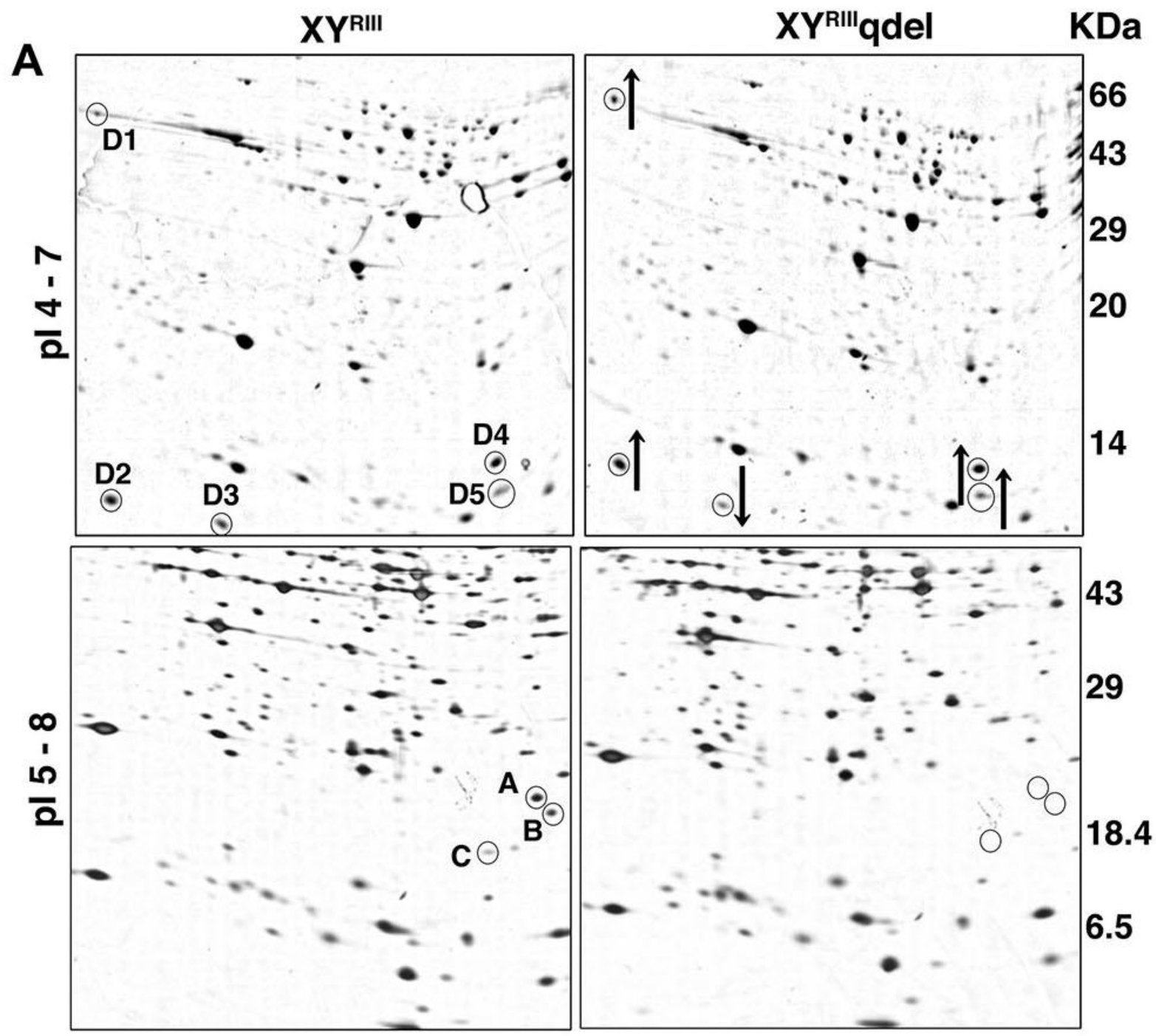
FJ541077 and FJ541118. Pre-incubation with MIWI antibody abolished the gelshift whereas pre-incubation with argonaute 3 antibody (ARG) did not - indicating specificity of binding. (*G*) Schematic representation of 3'UTR reporter constructs. (*H*) Results of reporter assay: Antagopirs to piRNA homologous sequences present in 3'UTRs of two genes, *Sod* and *PLA2G12B*, bring about concentration dependent reduction in luciferase expression. Use of these oligos against a non-target 3'UTR (*Cdc2l1*) did not affect luciferase expression.


A
M34
B
1 2 3 4 5 6 7 8 9 10 11
12 13 14 15 16 17 18 19 X Y MT not placed
7 38
15 742
6 24
C
10.0 µm
D
10.0 µm
E
1
2000
LTR
LTR
AT-rich
T-rich
(TG)n
(T)n
T-rich
B1
SINE
(GATA)n
AT-rich
2001
4000
T-rich
AT-rich
L1
LINE
MaLR
LTR
ERVK
LTR
B2
SINE
4001
6000
L1
LINE
L1
LINE
B4
SINE
(CATA)n
(CA)n
GA-rich
ES cell
6001
8000
Soares thymus
Stratagene-
8001
9510
-skin
B6-Dendritic cell
NIA 7.4K cDNA
M34
(DQ907163)
F
Pirmy
G
37nt
219nt
519nt
Pirmy (DQ907162.1)
1395nt
(Exons 2-5)
Sly (NM_201530.2)
Synaptonemal complex 3 like Y
1nt
5nt
41nt
45nt
345nt
918nt
H
Query DQ907162.1 (1374 bp); Subject AC213374.4 (161276 bp; +/-)
Exon1 (2-219) end
CTCTATGGCCCTGACCTGTTCAATCCTC
Exon2 (220-351)
ACTCAGTAGTCTTCTTAGTT...CTATACCTTCCGCAAGA
Exon 3 (352-418)
GCTCTTTACTGTATTGA...AAAATGTTACCCCATATATTCAT
Exon 4 (419-453)
CTTCATCTCTGTGAAGA...AACACATACCTCCAGATT
Exon 5 (454-520)
CTTTACTGTATTGAA...AATATTACCCCATGTCTT
Exon 6 (521-597)
ACTCTATTGAACTGGCAAA...GCATCTGACCTTGAAATGCT
Exon 7 (598-632)
AATTTCATCTCTGTGAAGACA...AAAATACATACCTCCAGATGTAG
Exon 8 (633-686)
AGAAAAGACCTAGAGAATTAG...ACTCACACCTCCAAGAG
Exon 9 (687-814)
AAACCCCTGGAGGAGGAGAAT...TCTCTTACCAGCATTTG
Exon 10 (815-934)
GGATCCTTTGGC...AGCATTACCTTGGGC
Exon 11 (935-1106)
GATAATCTATAAAGG...TAAGTACCAGAATAAT
Exon 12 (1107-1374)
ATGATGCAGCTGGAATATAGT


A
e1
e2
e3
e4
e10
e11
e12
e9
e15a
e15b
e15c
Y: 4341127- 4381724
B
ladder
brain
testis
NTC
1500
1000
500
400
200
100
1000
500
C
Contig
Mm RNA
RefSeq RNA
Comp
Genes_seq

218
332
467
534
602
635
746
827
896
934
1032
1180
1281
1381
1480
2516
963
1061
1351
2222
2283
FJ541075
FJ541076
FJ541077
FJ541078
FJ541079
FJ541080
FJ541126
FJ541093
FJ541081
FJ541082
FJ541083
FJ541084
FJ541085
FJ541086
FJ541087
FJ541088
FJ541089
FJ541090
FJ541091
FJ541092
FJ541094
FJ541095
FJ541096
FJ541097
FJ541098
FJ541099
FJ541100
FJ541101
FJ541102

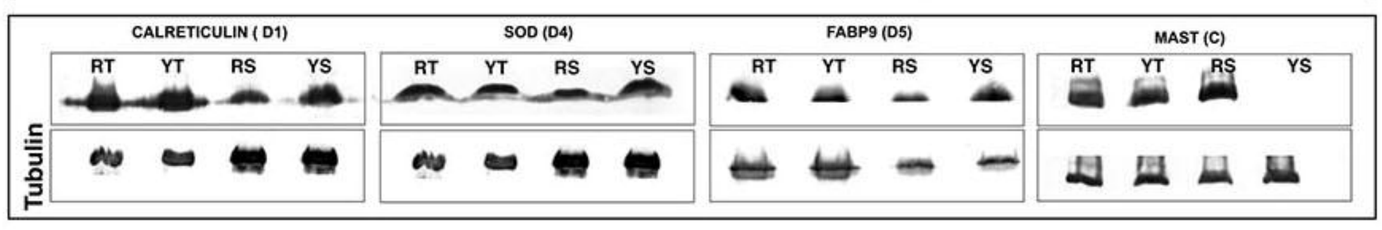
		e1	e1a	e2	e3	e3a	e4	e5	e6	e7	e8	e8a	e8b	e9	e9a	e10	e11	e12	e13	e13a	e14	e15	e15a	e15b	e15c
Splice variants		218	416	548	594	615	649	713	790	816	828	882	998	1026	1126	1244	1518	1625	1705	1728	1804	1893	1977	2560	2851
1	FJ541103	■		■	■	■	■	■	■	■	■	■	■	■	■			■							■
2	FJ541105	■		■	■	■	■	■		■	■	■	■	■	■			■							■
3	FJ541106	■		■	■	■	■	■		■	■	■	■	■	■										■
4	FJ541107	■		■	■	■	■	■		■	■	■	■					■	■	■					■
5	FJ541108	■		■	■	■	■	■		■	■	■	■					■							■
6	FJ541109	■		■	■	■	■	■		■	■	■	■												■
7	FJ541111	■		■	■	■	■	■		■	■	■		■	■		■	■							■
8	FJ541112	■		■	■	■	■	■		■	■	■		■	■			■	■	■					■
9	FJ541113	■		■	■	■	■	■		■	■	■		■	■			■				■			■
10	FJ541114	■		■	■	■	■	■		■	■	■		■	■			■							■
11	FJ541115	■		■	■	■	■	■		■	■	■		■	■										■
12	FJ541116	■		■	■	■	■	■		■	■	■			■			■							■
13	FJ541117	■		■	■	■	■	■		■	■	■					■	■							■
14	FJ541118	■		■	■	■	■	■		■	■	■						■							■
15	FJ541119	■		■	■	■	■	■		■	■	■													■
16	FJ541120	■	■	■	■	■	■	■		■	■														■
17	FJ541121	■		■	■	■	■	■		■	■			■	■	■									■
18	FJ541122	■		■	■	■	■	■		■	■			■	■			■							■
19	FJ541123	■		■	■	■	■	■		■	■							■							■
20	FJ541124	■		■	■	■	■	■		■	■														■
21	FJ541104	■		■	■	■	■	■										■	■	■					■
22	FJ541125	■		■	■	■	■			■	■	■		■	■										■
23	FJ541126	■		■	■	■	■																		■
24	FJ541142	■		■		■	■	■		■	■												■	■	■
25	FJ541127	■		■			■	■		■	■	■	■	■	■			■							■
26	FJ541128	■		■			■	■		■	■	■	■	■	■					■					■
27	FJ541129	■		■			■	■		■	■	■	■	■	■										■
28	FJ541130	■		■			■	■		■	■	■	■					■							■
29	FJ541131	■		■			■	■		■	■	■	■												■
30	FJ541132	■		■			■	■		■	■	■		■	■	■		■							■
31	FJ541133	■		■			■	■		■	■	■		■	■			■							■
32	FJ541134	■		■			■	■		■	■	■		■	■										■
33	FJ541135	■		■			■	■		■	■	■					■	■							■
34	FJ541136	■		■			■	■		■	■	■						■				■	■		■
35	FJ541137	■		■			■	■		■	■	■						■				■			■
36	FJ541138	■		■			■	■		■	■	■						■							■
37	FJ541140	■		■			■	■		■	■			■	■			■							■
38	FJ541141	■		■			■	■		■	■							■							■
39	FJ541143	■		■			■	■		■	■														■
40	FJ541144	■			■		■	■		■	■	■	■	■	■			■							■
41	FJ541145	■			■		■	■		■	■	■	■	■	■							■	■		■
42	FJ541146	■			■		■	■		■	■	■	■	■	■										■
43	FJ541147	■			■		■	■		■	■	■	■					■							■
44	FJ541148	■			■		■	■		■	■	■		■	■			■							■
45	FJ541149	■			■		■	■		■	■	■		■	■						■				■
46	FJ541150	■			■		■	■		■	■	■		■	■										■
47	FJ541151	■			■		■	■		■	■	■						■							■
48	FJ541152	■			■		■	■		■	■	■													■
49	FJ541153	■			■		■	■		■	■							■				■	■		■
50	FJ541154	■			■		■	■		■	■														■
51	FJ541155	■				■	■	■		■	■	■	■	■	■			■							■
52	FJ541156	■				■	■	■		■	■	■	■	■	■					■					■
53	FJ541157	■				■	■	■		■	■	■	■	■	■										■
54	FJ541158	■				■	■	■		■	■	■		■	■	■		■							■
55	FJ541159	■				■	■	■		■	■	■		■	■										■
56	FJ541160	■				■	■	■		■	■	■						■							■
57	FJ541139	■				■	■	■		■	■	■									■				■
58	FJ541161	■				■	■	■		■	■														■
59	FJ541162	■					■	■		■	■	■	■	■	■	■	■	■							■
60	FJ541163	■					■	■		■	■	■	■	■	■			■							■
61	FJ541164	■					■	■		■	■	■	■	■	■					■					■
62	FJ541165	■					■	■		■	■	■	■	■	■										■
63	FJ541166	■					■	■		■	■	■	■					■							■
64	FJ541167	■					■	■		■	■	■	■												■
65	FJ541110	■					■	■		■	■	■		■	■	■		■							■
66	FJ541168	■					■	■		■	■	■		■	■		■	■							■
67	FJ541169	■					■	■		■	■	■		■	■			■							■
68	FJ541170	■					■	■		■	■	■		■	■										■
69	FJ541171	■					■	■		■	■	■						■							■
70	FJ541172	■					■	■		■	■	■													■
71	FJ541173	■					■	■		■	■			■	■			■							■
72	FJ541174	■					■	■		■	■							■							■
73	FJ541175	■					■	■		■	■														■
74	FJ541176	■					■																■	■	■
75	FJ541177	■						■		■	■	■	■	■	■										■
76	FJ541178	■						■		■	■	■	■	■	■			■							■
77	FJ541179	■						■		■	■	■	■					■							■
78	FJ541180	■						■		■	■	■						■							■
79	FJ541181	■		■	■	■	■	■		■	■	■		■	■	■		■							■

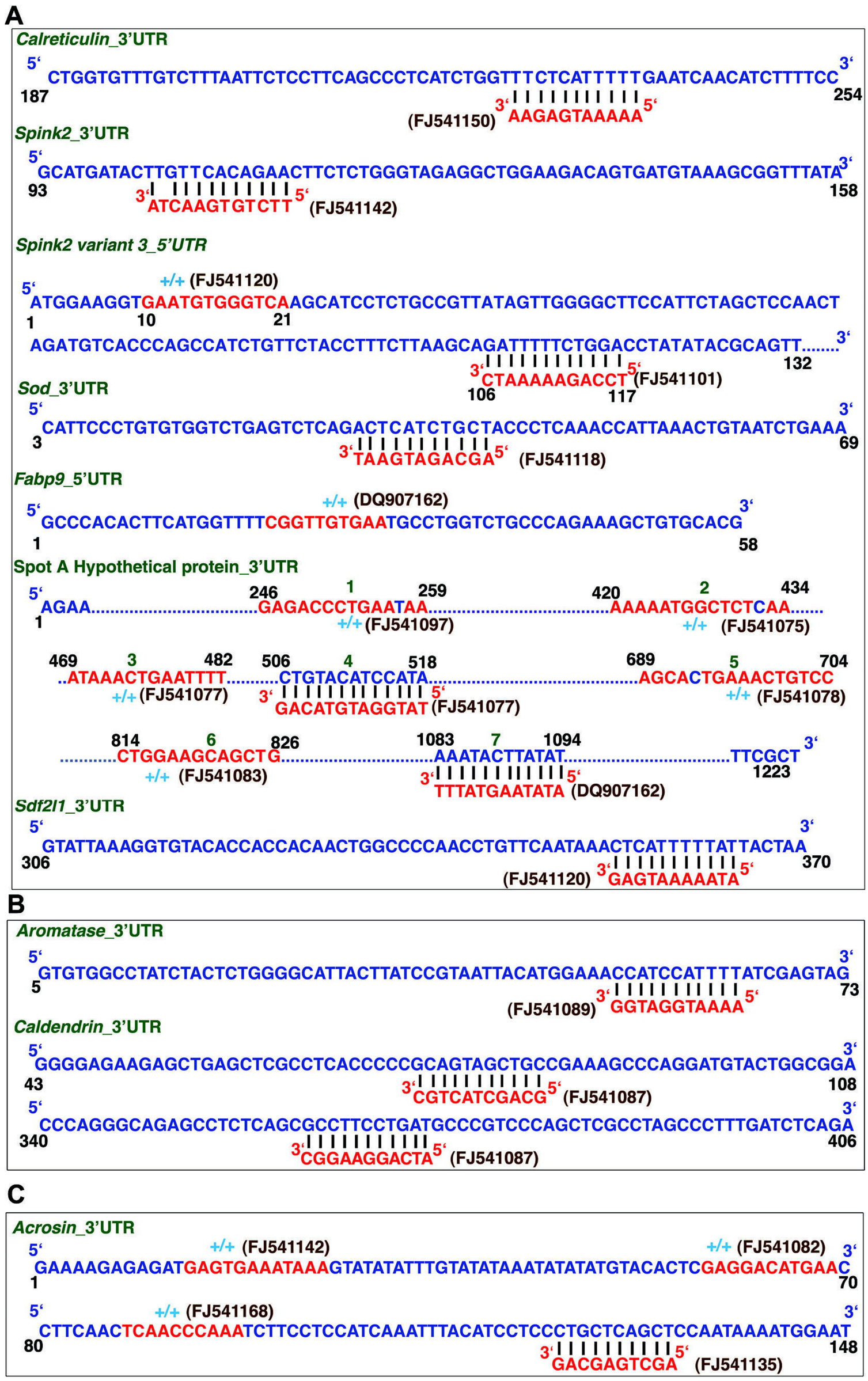


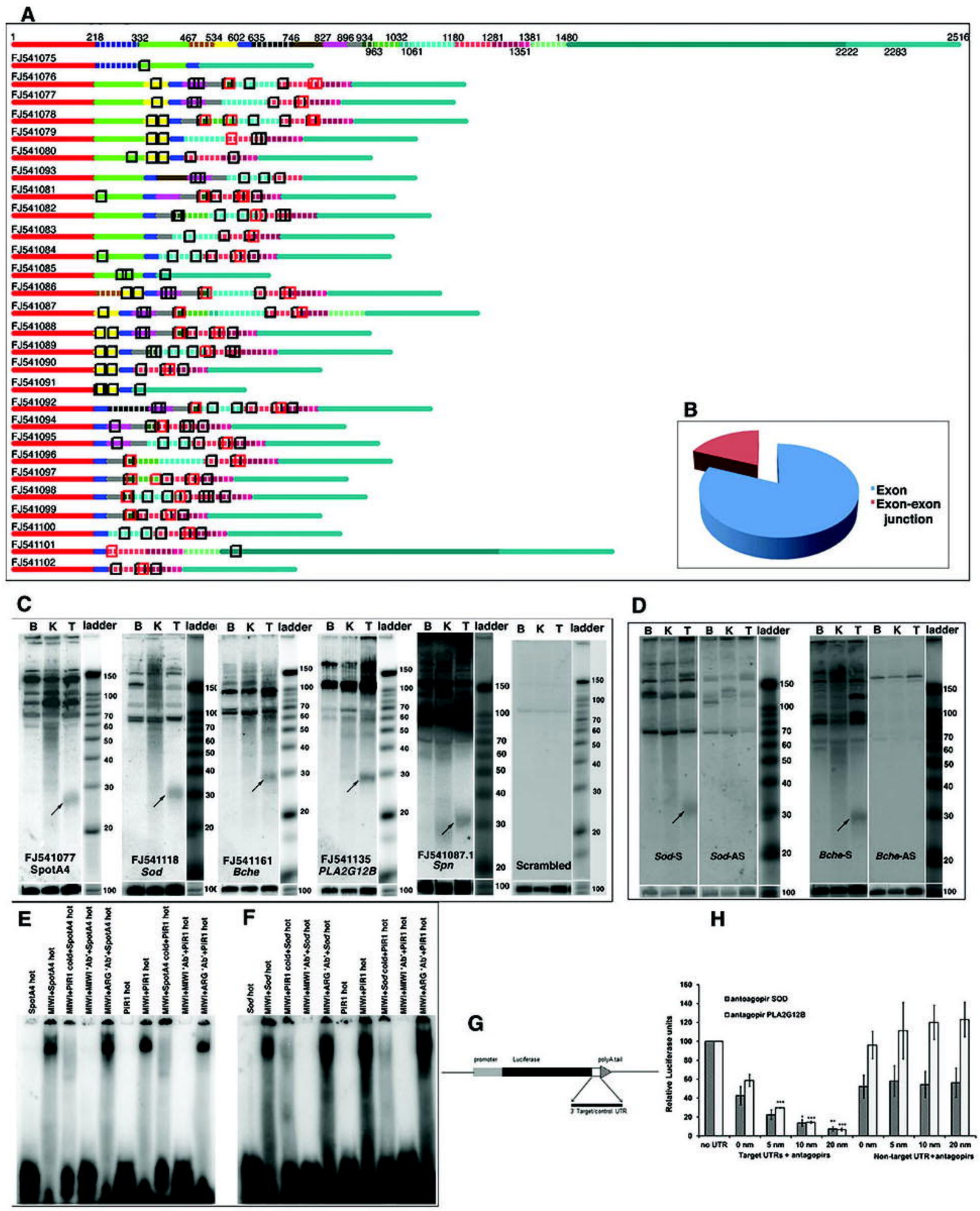
B

Serial number	Protein spot	MS-tags and N-terminal sequences	Protein identity & Gene ID	Protein expression in XY^RIIIqdel	Chromosome localization
1	D1	**DPAIY** (N-terminal sequence)	CALRETICULIN **12317**	Upregulated	8
2	D2	(198.0975) **DCGHFDFPACPR** (156.1027) (269.1519) **EDGSHINII** (629.1989)	SPINK2 **69982**	Upregulated	5
3	D3	(198.08) **DCGHFDFPACPR** (159.10) **SQFRTPDCGH** (N-terminal sequence)	Q8BMY7 **69982**	Down regulated	5
4	D4	(115.0652) **GVANVSIED** (156.1003) (99.1182) **ISLSGEGSIIG** (156.0945)	SOD **20655**	Upregulated	16
5	D5	(1469.6) **ISSENFENYV** (156.0826) (131.0266) **DIQAGSAC** (156.0985) (113.0616) **GEEFEETTADN** (156.1044) (230.1680) **VECTMNNVVST** (156.1044)	FABP9 **21884**	Upregulated	3
6	A	(398.4527) **HINEEEQE** (155.9080)	Q9DAR0 **69315**	Absent	13
7	B	(218.1635) **EGLFLQ** (1192.3747)	**SDF2L1**	Absent	16
8	C	(214.1065) **VSQSNDLSSL** (156.0965)	Q7TPM5 **69987**	Absent	2

C




A
Calreticulin_3'UTR
5' CTGGTGTTTGTCTTTAATTCTCCTTCAGCCCTCATCTGGTTTCTCATTTTTGAATCAACATCTTTTCC 3'
187 254
3' AAGAGTAAAAA 5' (FJ541150)
Spink2_3'UTR
5' GCATGATACTTGTTCACAGAACTTCTCTGGGTAGAGGCTGGAAGACAGTGATGTAAAGCGGTTTATA 3'
93 158
3' ATCAAGTGTCTT 5' (FJ541142)
Spink2 variant 3_5'UTR
+/+ (FJ541120)
5' ATGGAAGGTGAATGTGGGTCAAGCATCCTCTGCCGTTATAGTTGGGGCTTCCATTCTAGCTCCAACT
1 10 21
AGATGTCACCCAGCCATCTGTTCTACCTTTCTTAAGCAGATTTTTCTGGACCTATATACGCAGTT..... 3'
132
3' CTAAAAAGACCT 5' (FJ541101)
106 117
Sod_3'UTR
5' CATTCCCTGTGTGGTCTGAGTCTCAGACTCATCTGCTACCCTCAAACCATTAAACTGTAATCTGAAA 3'
3 69
3' TAAGTAGACGA 5' (FJ541118)
Fabp9_5'UTR
+/+ (DQ907162)
5' GCCCACACTTCATGGTTTTCGGTTGTGAATGCCTGGTCTGCCCAGAAAGCTGTGCACG 3'
1 58
Spot A Hypothetical protein_3'UTR
5' AGAA.....GAGACCCTGAATAA.....AAAAATGGCTCTCAA.....
1 246 1 259 420 2 434
+/+ (FJ541097) +/+ (FJ541075)
469 3 482 506 4 518 689 5 704
..ATAAACTGAATTTT.....CTGTACATCCATA.....AGCACTGAAACTGTCC
+/+ (FJ541077) 3' GACATGTAGGTAT 5' (FJ541077) +/+ (FJ541078)
814 6 826 1083 7 1094
.....CTGGAAGCAGCTG.....AAATACTTATAT.....TTCGCT 3'
+/+ (FJ541083) 3' TTTATGAATATA 5' (DQ907162) 1223
Sdf2l1_3'UTR
5' GTATTAAAGGTGTACACCACCACAACTGGCCCCAACCTGTTCAATAAACTCATTTTTATTACTAA 3'
306 370
(FJ541120) 3' GAGTAAAAATA 5'
B
Aromatase_3'UTR
5' GTGTGGCCTATCTACTCTGGGGCATTACTTATCCGTAATTACATGGAAACCATCCATTTTATCGAGTAG 3'
5 73
(FJ541089) 3' GGTAGGTAAAA 5'
Caldendrin_3'UTR
5' GGGGAGAAGAGCTGAGCTCGCCTCACCCCCGCAGTAGCTGCCGAAAGCCCAGGATGTACTGGCGGA 3'
43 108
3' CGTCATCGACG 5' (FJ541087)
5' CCCAGGGCAGAGCCTCTCAGCGCCTTCCTGATGCCCGTCCCAGCTCGCCTAGCCCTTTGATCTCAGA 3'
340 406
3' CGGAAGGACTA 5' (FJ541087)
C
Acrosin_3'UTR
5' +/+ (FJ541142) +/+ (FJ541082) 3'
GAAAAGAGAGATGAGTGAAATAAAGTATATATTTGTATATAAATATATATGTACACTCGAGGACATGAAC
1 70
5' +/+ (FJ541168) 3'
CTTCAACTCAACCCAAATCTTCCTCCATCAAATTTACATCCTCCCTGCTCAGCTCCAATAAAATGGAAT
80 148
3' GACGAGTCGA 5' (FJ541135)


A
B
Exon
Exon-exon junction
C
FJ541077 SpotA4
FJ541118 Sod
FJ541161 Bche
FJ541135 PLA2G12B
FJ541087.1 Spn
Scrambled
D
Sod-S
Sod-AS
Bche-S
Bche-AS
E
F
G
promoter
Luciferase
polyA tail
3' Target/control UTR
H
antagopir SOD
antagopir PLA2G12B
Relative Luciferase units
Target UTRs + antagopirs
Non-target UTR+antagopirs