Y chromosomal noncoding RNA regulates autosomal gene expression via piRNAs in mouse testis

Hemakumar M. Reddy,[1,2,11] Rupa Bhattacharya,[1,3,11] 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,[10] Mahadevan Sivaramakrishnan,[1,8] Anurag Chaturvedi,[1,9] Nandini Rangaraj[1], Shiju Michael Thomas[1], Badanapuram Sridevi,[1] Sachin Kumar,[1] Ram Reddy Dereddi[1], Sunayana Rayabandla[1], Rachel A. Jesudasan[1]*

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

Present address:

[2]Division of Paediatric Neurology, Department of Paediatrics, University of Florida College of Medicine, Gainesville, FL, USA

[3]Lawrenceville, New Jersey 08648, USA

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

[5]Department of Medical and Clinical Genetics, Goteborg University, Goteborg, Sweden

[6]National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India

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

[9]Laboratory of Biodiversity and Evolutionary Genomics, Charles Deberiotstraat, Leuven, Belgium.

[10]Deceased.

[11]These authors contributed equally to the work

*Correspondence: Email ID: rachellike@gmail.com; rachel@csirccmb.org

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

Running title: Y-ncRNA regulates autosomal gene expression

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

Significance

Our study adds new dimensions to the repertoire of functions that Y-heterochromatin has with respect to autosomal gene regulation in male fertility. We describe a novel sex and species-specific ncRNA, *Pirmy*, transcribed from mouse Y-chromosome, with unprecedented alternative splicing. Mice with deletions of Y-long arm (Yq) were reported to exhibit differential sterility and sperm abnormalities. We identify sperm proteins that are deregulated in Yq-deleted mice; corresponding genes map to autosomes and display sequence homology to piRNAs derived from *Pirmy* in their UTRs. Thus, piRNAs from species-specific repeats appear to regulate autosomal genes involved in reproduction. The male sterility phenotypes of Yq-deleted mice and cross-species hybrids are comparable, indicating a larger role for species-specific repeats from Y-chromosome in speciation and evolution.

ABSTRACT

Repeats from the male-specific long arm of mouse Y-chromosome (MSYq) transcribe protein coding and noncoding RNAs. Majority of genes expressed during spermatogenesis are autosomal. Mice with different deletions of Yq show sub-fertility and sterility, depending on the length of deletion. The connection between Yq deletion and autosomal gene regulation was not well studied before. We describe a novel multi-copy mouse Yq-derived long noncoding RNA, *Pirmy* (piRNA from mouse Y

chromosome), which shows unprecedented alternative splicing in testis. Further, *Pirmy* acts as a template for several piRNAs in mouse testis. Mice with deletions of Yq including the region of *Pirmy* exhibit differential male fertility, aberrant sperm morphology, motility and sex ratio skewed towards females. We identified deregulation of ten autosome-encoded sperm proteins in a Y-deleted strain of mouse. *Pirmy* splice 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-encoded *Pirmy*. Our study provides novel and interesting insights into possible role of MSY-derived noncoding RNAs in male fertility and reproduction. Finally, sperm phenotypes from 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 noncoding RNAs 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 (1-7). 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 (8-11).

Previously published data have described two different strains of mice with partial deletions of the long arm of the Y chromosome (Yq) (9, 12). 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 (9). 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 (13). This suggested the presence of multicopy spermiogenesis gene(s) on mouse Yq (9, 14). 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 (9, 13, 15). 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 (13). The next major gene to be discovered on mouse Yq was the multicopy *Sly*. As SLY interacts with a histone acetyl transferase and an acrosomal protein, the authors suggested that *Sly* could control transcription and acrosome functions (16). Further, Coccquet 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 (16).

Vast majority of the genes required for spermatogenesis and spermiogenesis are non-Y-linked (17-19). There is one report of interaction of an autosomal gene with a human Y chromosome lncRNA from our lab (3). We hypothesized more interactions of the kind wherein there would exist a complex interplay between protein coding genes and noncoding RNAs from the Y chromosome. Altogether, we conjectured structural or

regulatory elements in the biology of spermatogenesis from the Yq region, that are yet to be discovered.

Previous studies in the lab identified 300-400 copies of the M34 clone (DQ907163.1) on mouse Yq using slot blot, sequencing and bioinformatics analyses (20, 21). This region is deleted in the XY^{RIII}qdel mice with sperm abnormalities. However, the function of these sequences remains undefined. As sperm development is a polygenic phenomenon 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 showed a decrease in expression of M34 in testis compared to the XY^{RIII} mice. In order to understand putative functions of this sequence, first of all we identified a transcript corresponding to M34 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) novel noncoding RNA (*Pirmy*) that is present in multiple copies on mouse Y long arm and is expressed from the embryonic stages of testicular development (ii) extensive alternative splicing of *Pirmy*, and the generation of piRNAs from it in mouse testis and (iii) their 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 - (3)) (Fig. 1*A*). BLAST analysis of M34 sequence against mouse whole genome (NCBI Build 38.1) also showed maximum similarity to Y chromosome (>97% identity) with few hits on the X (Fig. 1*B*). M34 was then analyzed for expression in adult mouse testis. FISH using M34 as a probe revealed abundant expression 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 (30 days postpartum) (Fig. S1) suggesting expression of this repeat in mouse testis from early developmental stages.

Figure 1

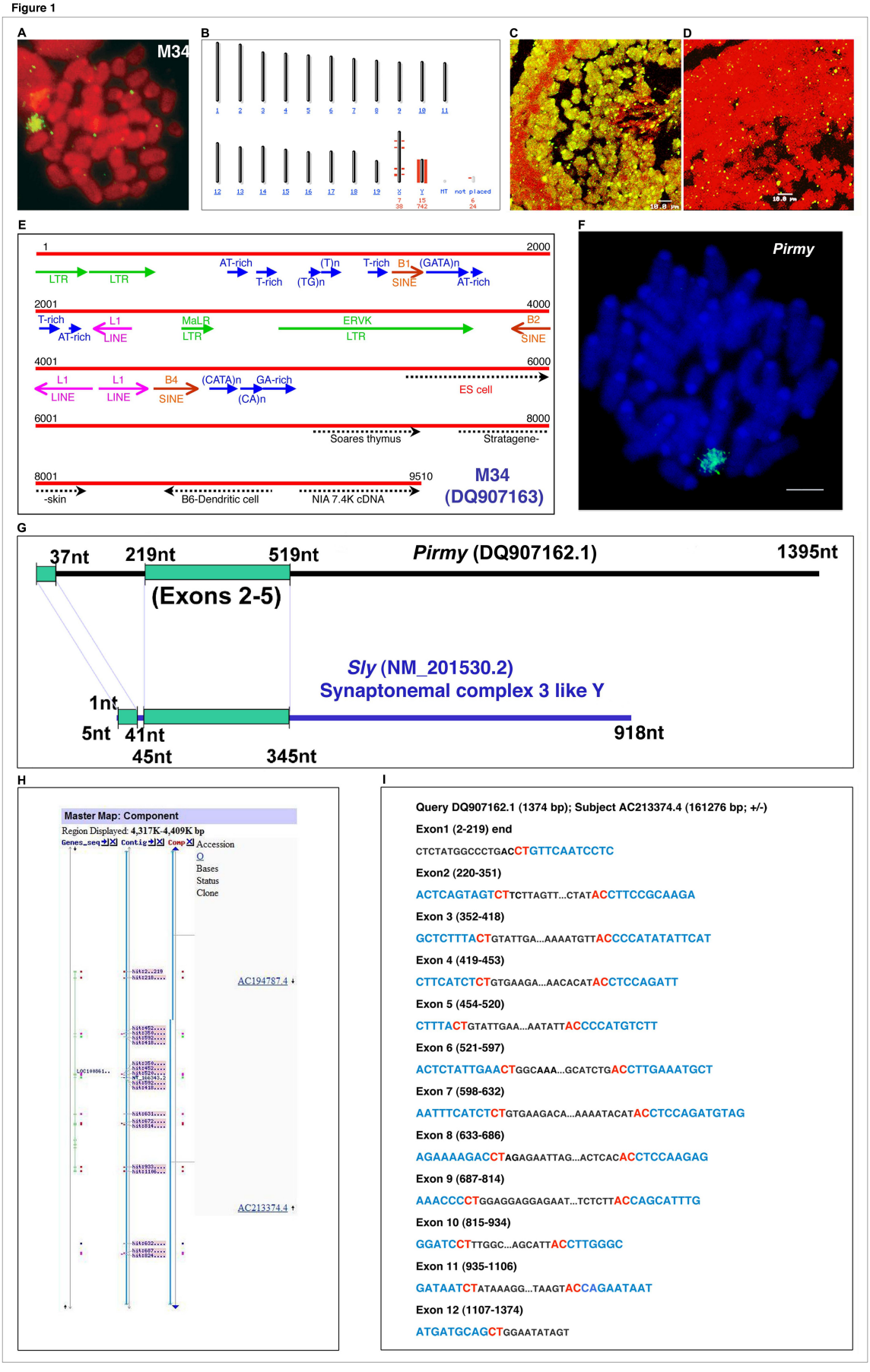


Figure 1. Analysis of M34 (DQ907163) and identification of a novel noncoding RNA. (*A*) Localization by FISH of the genomic clone, M34 to Mouse Y long arm in multiple copies spanning its entire length. (*B*) Mouse genome map view of M34 BLAST hits, showing Y-chromosomal localization further indicating 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*) NCBI BLAST map view of *Pirmy* (DQ907162). All the exons in DQ907162 match to the locus LOC100861. Complete sequence of DQ907162 localizes to the BAC clone AC213374. (*I*) 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).

Isolation of novel polyadenylated noncoding repeat transcript from mouse Y chromosome

Sequence analysis of the 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 sequence (Fig. 1*E*). A number of gene prediction programs like GENSCAN, GrailEXP, MZEF, and GeneMark did not predict any genes within M34 with consistency. BLAST analysis of M34 sequence against the EST database of NCBI (NCBI build 36.1) identified 5 ESTs at the 3' end of M34 sequence (Fig. 1*E*). Expression of these ESTs was observed in embryos from at least 13.5d onwards (data not shown). Two of these ESTs showed male-specific expression by Reverse transcription PCR (RT-PCR) analysis.

In order to identify a cDNA corresponding to M34 in testis, one of the male specific ESTs (CA533654) was used to screen a mouse testis cDNA library. A 1395 nt long polyadenylated Y-specific cDNA was isolated and this was named *Pirmy* - piRNA from mouse Y chromosome (NCBI Accession number DQ907162.1). FISH on to mouse metaphase spreads showed that *Pirmy* is present only on the Y chromosome in multiple copies (Fig. 1*F*) similar to that of the genomic clone M34. BLAST of *Pirmy* against nr nucleotide database of NCBI picked up only mouse sequences with statistically significant alignments (e-value $<4e^{04}$). This suggests that *Pirmy* is specific to mouse. BLAST analysis using *Pirmy* against the mouse genome plus transcriptome database showed homology to *Sly* transcript in exons 1-5. The exons 2-5 were identical in *Pirmy* and *Sly* (Fig. 1*G*). The entire sequence of *Pirmy* localizes to the BAC clone AC213374 from the Y chromosome (Fig. 1*H*). Computational translation using different programs like ExPASy-Translate and NCBI ORF Finder did not predict ORFs of significant length in any of the reading frames of *Pirmy* reiterating that *Pirmy* is a novel mouse Y specific noncoding RNA. The exon-intron junctions of *Pirmy* are shown in Figure 1I.

Splice variants of *Pirmy* in mouse testis

Pirmy was analyzed further by RT-PCR. Two rounds of amplification using primers to the two ends of *Pirmy* yielded multiple products in testis and brain (Fig. 2*B*). Cloning and sequencing of the PCR products from testis identified 107 splice variants of *Pirmy* (Figs. 2A, 3A), (NCBI accession numbers FJ541075-FJ541181) besides the one obtained by screening testis-cDNA library. Sanger sequencing of more than 1000 clones yielded the 107 splice variants. BLAST analysis of these transcripts against the NCBI genome database showed that they are present in multiple copies on the mouse Y chromosome. There were 17 exons in the splice variants of *Pirmy* when exons were arranged in the order of occurrence at the corresponding genomic loci in NCBI build 38.1. These splice-variants clustered into two groups based on two sets of mutually exclusive exons. In Group 1 exons 11-16 were absent whereas in Group II exons 7-10 were absent (Fig. S2).

In silico analysis of the splice variants against NCBI build 38.1 confirmed that they localized to two different contigs, NT_166343 and NT_187045. Of the 108 splice variants 80 matched the contig NT_166343 at NCBI (Fig. 2*C*) and localized to a single locus at chrY: 4,337,941-4,389,324 on the UCSC Genome Browser on Mouse genome (Fig. S3). The rest of the 28 splice variants mapped to NT_187045 (Fig. 3*A*); 25 of these mapped to chrY: 32,548,222-32,631,777 on the UCSC Browser Mouse Genome (Fig. 3*B*). These alternatively spliced transcripts 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. S4). The 108 splice variants of *Pirmy* contained consensus splice signal sequences at all the intron-exon junctions with the exception of four junctions in four different isoforms i. e. splice variants (SV) 4, 8, 19 and 22 (Data sheet).

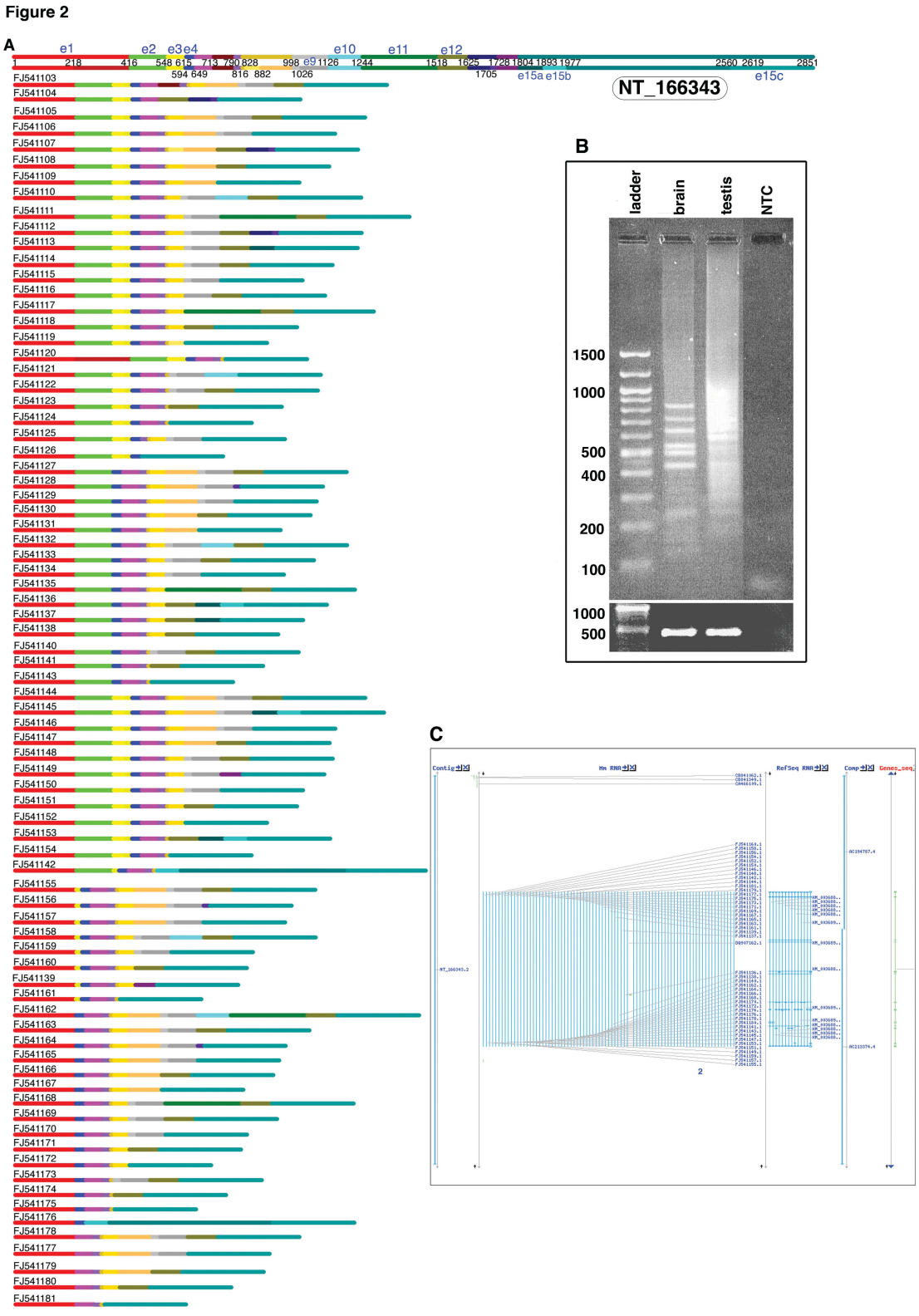


Figure 2. Identification of multiple splice-variants of *Pirmy***.** Autoassembler program of ABI-Prism identified cDNAs that differed from one another. (*A*) Color coded line diagram showing extensive alternative splicing of *Pirmy*. The splice variants depicted

here show homology to contig NT_166343. Each exon is represented by the same color in different isoforms. Sizes of the exons are to scale. Top two lines show the representation of all exons present in this contig together 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 80 splice variants of *Pirmy* to NT_166343.

Figure 3

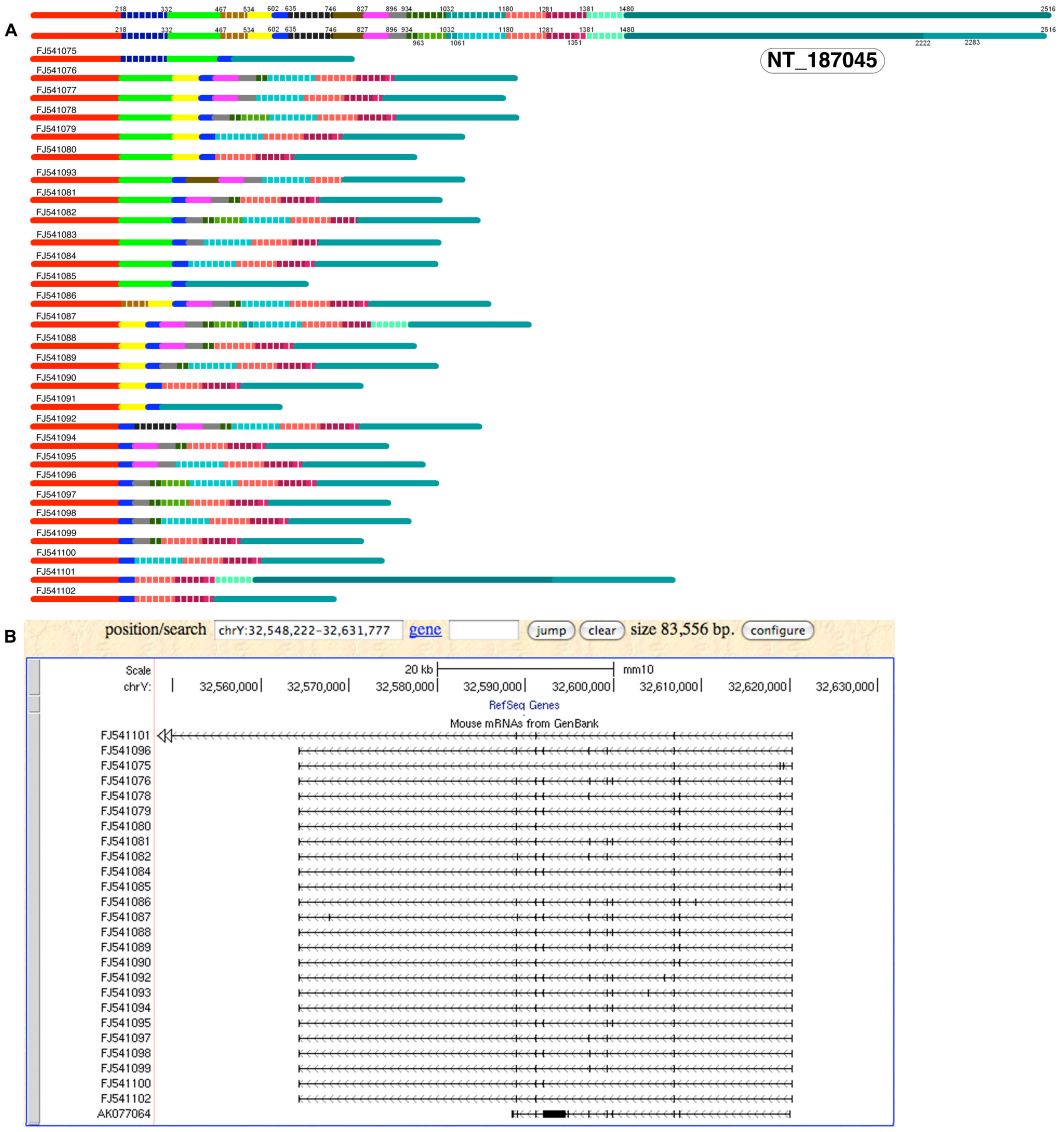


Figure 3. Localization of 28 splice-variants of *Pirmy*. (*A*) Top two lines show the representation of all exons present together according to their order in NT_187045 as e1, e2 etc. Line 2 indicates the nucleotide positions in a scenario wherein all the exons are present. Different splice variants have been arranged in linear order. Exons in dashed lines are specific to NT_187045, whereas other exons are common to both the sets of splice isoforms. (*B*) BLAST against the mouse genome on the UCSC browser shows localization of 25 of the 28 isoforms at the same locus chrY: 32,548,222-32,631,777.

Splice variants of *Pirmy* in mouse brain

RT-PCR amplification (Fig. 2*B*), cloning and sequencing of *Pirmy* products from mouse brain yielded 12 splice variants. Comparison of the exons of the splice variants from the brain and testis showed that the same isoforms were present in both (Fig. 4). All the splice variants from brain localized to the genomic contig NT_166343, to which the 80 splice variants from testis also mapped.

Figure 4

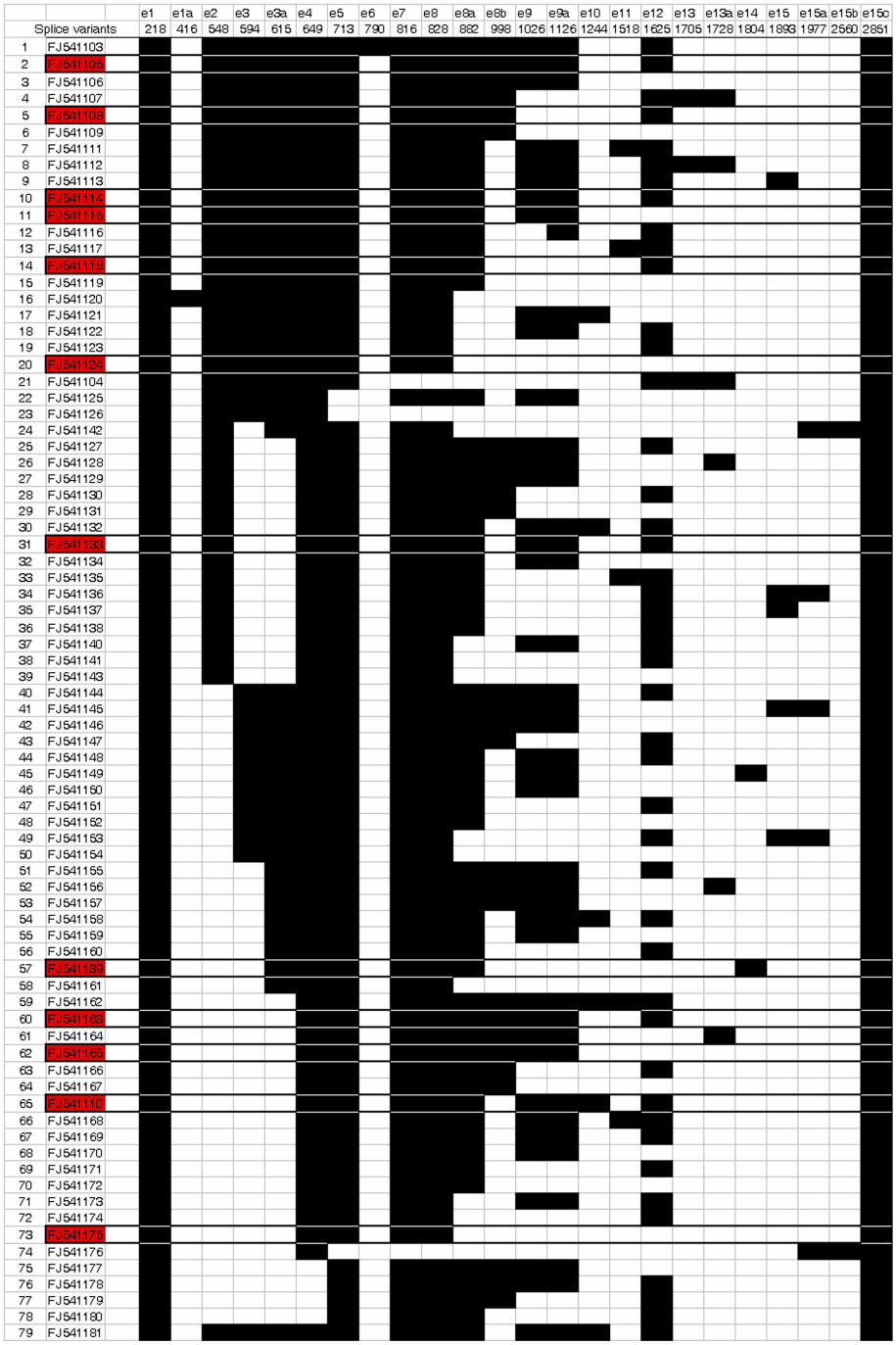


Figure 4. Comparison of splice-variants from testis and brain. The sequential exons from *Pirmy* splice-variants in testis have been represented in an Excel data sheet. Highlighted in red are the splice variants that are transcribed from male brain, which were also found in testis. The representation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

Expression of M34 in Y-deleted (XY^{RIII}qdel) mice

Metaphase spreads from the XY^{RIII}qdel mice showed a reduction in copy number of M34 on the Y chromosome showing that it localizes to the deleted region of the Y chromosome in the above mice (Fig. S5*B*). Therefore, expression of M34 was then checked in testis and sperms of XY^{RIII} and XY^{RIII}qdel mice by FISH. Dramatic reduction in fluorescence intensity was observed in testis and sperm of XY^{RIII}qdel mice. However, sperms from epididymis of both XY^{RIII} and XY^{RIII}qdel mice showed faint fluorescence intensity (Fig. S5*B*). The subclones of M34 (Fig. *S5A*) when used as probes on testis sections showed reduction in fluorescence intensity in XY^{RIII}qdel mice (Fig. S5*C*).

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

We then analyzed the motility profile of sperms from XY^{RIII} and XY^{RIII}qdel mice and identified a stark difference in motility patterns (Movies S1, S2). Spermatozoa from XY^{RIII} mice show linear progressive motion whereas sperms from XY^{RIII}qdel mice show rapid flagellar movement with non-linear and non-progressive motion. Most of the spermatozoa from XY^{RIII}qdel mice stall at the same position with no linear displacement. Studies from Burgoyne's lab reported morphological abnormalities in sperms from XY^{RIII}qdel mice (9). In order to understand the connection between the Y-deletion and sperm abnormalities, we performed comparative sperm proteome analysis between normal mice and Y-long arm deletion mutant XY^{RIII}qdel mice by 2D-PAGE and mass spectrometry using protocols standardized in the laboratory (22).

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

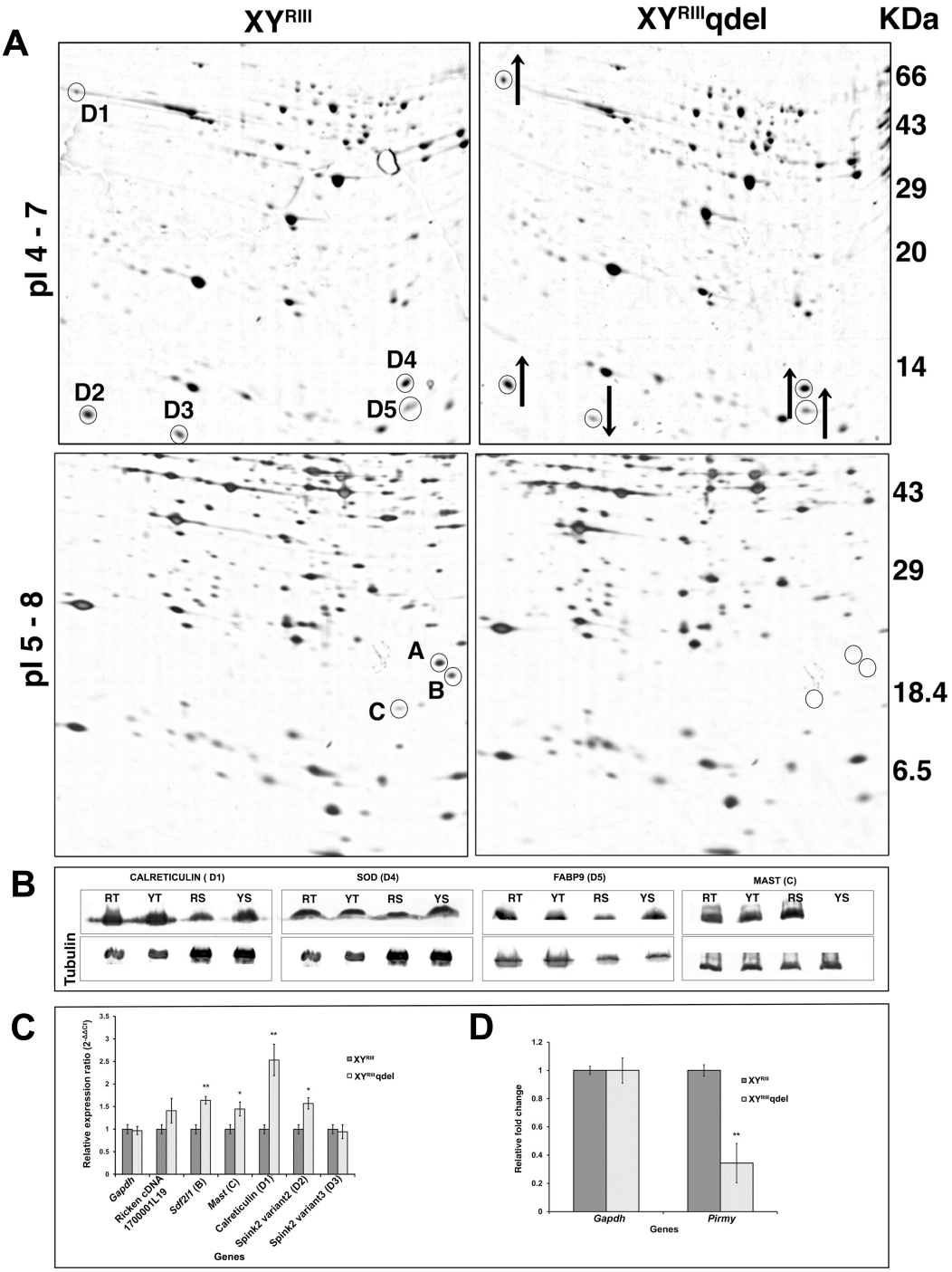
Peptidase Inhibitor (Kazal type II (SPINK2)/Acrosin-Trypsin inhibitor (D2) were upregulated in XY^{RIII}qdel sperms compared to XY^{RIII} sperms (Fig. 5*A*). A variant of SPINK2 (D3), with a 30-amino acid N-terminal truncation (Q8BMY7) was downregulated in XY^{RIII}qdel sperms (Fig. 5*A*, Fig. S7). Three proteins were not detectable in XY^{RIII}qdel sperms in the pI range of 5-8 (Fig. 5*A*). Two of these were reported as hypothetical proteins in the NCBI database (1700001L19 Riken cDNA [Q9DAR0 (A)], 1700026L06 Riken cDNA [MAST, Q7TPM5 (C)]) and the third one was Stromal cell derived factor 2 like 1 (SDF2L1 (B)). Expression of four of the eight differentially expressed proteins was also confirmed by western blotting (Fig. 5*B*). Calreticulin, SOD and FABP9 showed upregulation in XY^{RIII}qdel sperms by both 2D PAGE and western blot analyses. However, MAST was not detectable by both the techniques in sperms. Further, protein expression of calreticulin, SOD, FABP9 and MAST did not vary significantly between XY^{RIII} and XY^{RIII}qdel testes.

Thus, in sperms from the XY^{RIII}qdel, four proteins FABP9, calreticulin, SPINK2 and SOD were upregulated while a novel truncated variant of SPINK2 protein was downregulated. Three proteins were absent in the XY^{RIII}qdel sperm proteome, one of which was SDF2L1 and other two were identified as novel testis-specific proteins. Surprisingly, all the genes corresponding to the differentially expressed proteins localized to different autosomes (Fig. 5*E*).

Next, we analyzed the expression of the transcripts corresponding to the protein spots in testis. Although the proteins of 1700001L19 Riken cDNA, SDF2L1 and MAST were not detectable in sperms of XY^{RIII}qdel, the corresponding RNAs were present in testis. Expression of the transcripts of SDF2L1, MAST, calreticulin and *Spink2* variant 2 were upregulated in XY^{RIII}qdel mice testis. In contrast transcripts of spot A (1700001L19

Riken cDNA) and *Spink2* variant 3 did not differ significantly between the two (Fig. 5*C*). We also checked the copy number of *Pirmy* in DNA isolated from XY^{RIII}qdel mice by Real-time PCR; a significant reduction in copy number of *Pirmy* was observed in the XY^{RIII}qdel genome (Fig. 5*D*).

Figure 5



E

Serial number	Protein spot	MS-tags and N-terminal sequences	Protein identity & Gene ID	Protein expression in XY^{RIII} qdel	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)	SPINK2 truncated variant **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)	Hypothetical protein (Riken cDNA 1700001L19) **69315**	Absent	13
7	B	(218.1635) **EGLFLQ** (1192.3747)	**SDF2L1**	Absent	16
8	C	(214.1065) **VSQSNDLSSL** (156.0965)	Hypothetical protein (Riken cDNA 1700026L06) **69987**	Absent	2

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*) 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. (C) the expression profile of few of the corresponding genes in testis by Real-time PCR analysis. (D) This panel shows the copy number difference in the *Pirmy* gene between XY^{RIII} and XY^{RIII}qdel. (*E*) 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.

UTRs of deregulated autosomal genes show homology to *Pirmy*

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. As *Pirmy* localizes to the region of deletion and its splice variants appear to be long noncoding RNAs (lncRNAs), we hypothesized that *Pirmy* could be regulating the autosomal gene expression in testis. Therefore, BLAST analysis of the *Pirmy* splice variants was carried out against the UTRs of the deregulated genes. This revealed short stretches of homology ranging in size from 10-16 nucleotides, in either +/+ or

+/- orientations. Majority of the sequence homologies observed were in the 3' UTRs whereas *Fabp9* and *Spink2* variant 3 showed homology to *Pirmy* in their 5' UTRs. There were as many as 7 *Pirmy* hits in the 3' UTR of the hypothetical protein spot A (Fig. 6). Of the 10 proteins identified in the lab nine showed homology to *Pirmy* splice variants in their UTRs. Homology between Y-derived *Pirmy* transcripts and UTRs of deregulated autosomal genes provides evidence for interactions between genes on the Y chromosome and autosomes in mouse testis.

Furthermore, we performed BLAST analysis of *Pirmy* 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 on the splice-isoforms of *Pirmy* (Fig. 7*A*). We identified 372 unique homologous stretches in the *Pirmy* splice variants. Of these 302 (81.19%) localized to the exons and 70 (18.82%) to exon-exon junctions (Fig. 7*B*). The genes bearing homology to *Pirmy* in their UTRs, localized to different autosomes and were expressed in multiple tissues including testis/epididymis (Table S1). Further, the fact that *Pirmy* splice variants had homology only to the UTRs and not to the coding exons of those genes, strongly suggests that they might have a role in regulation of these genes.

Figure 6

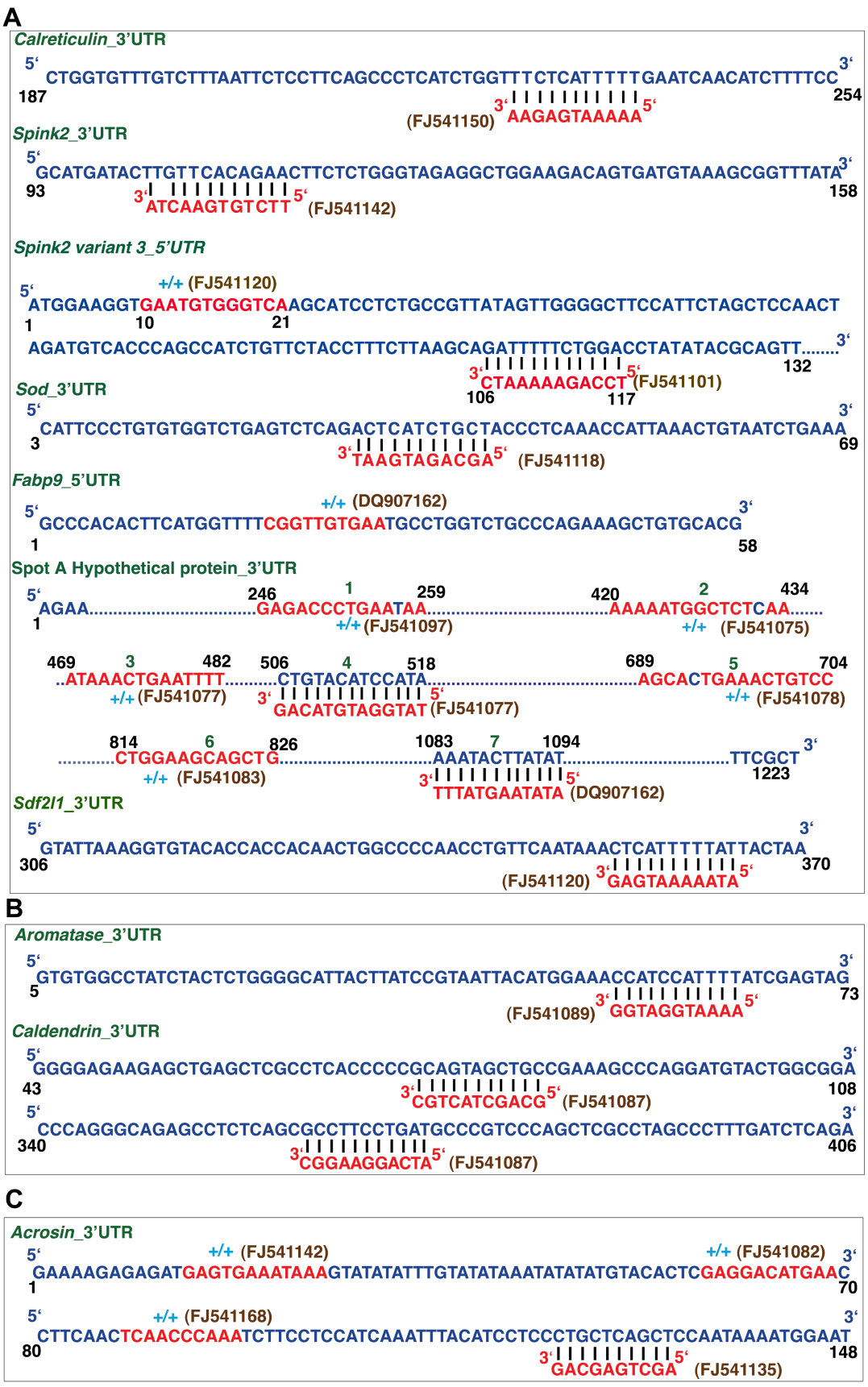


Figure 6. Localization of *Pirmy* homologous sequences to UTRs of deregulated genes. Panel (*A*) shows the UTR regions of the deregulated genes with the sequences homologous to *Pirmy* highlighted in red. Both +/+ and +/- homologies are observed. The splice isoforms of *Pirmy* are indicated in brown 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.*

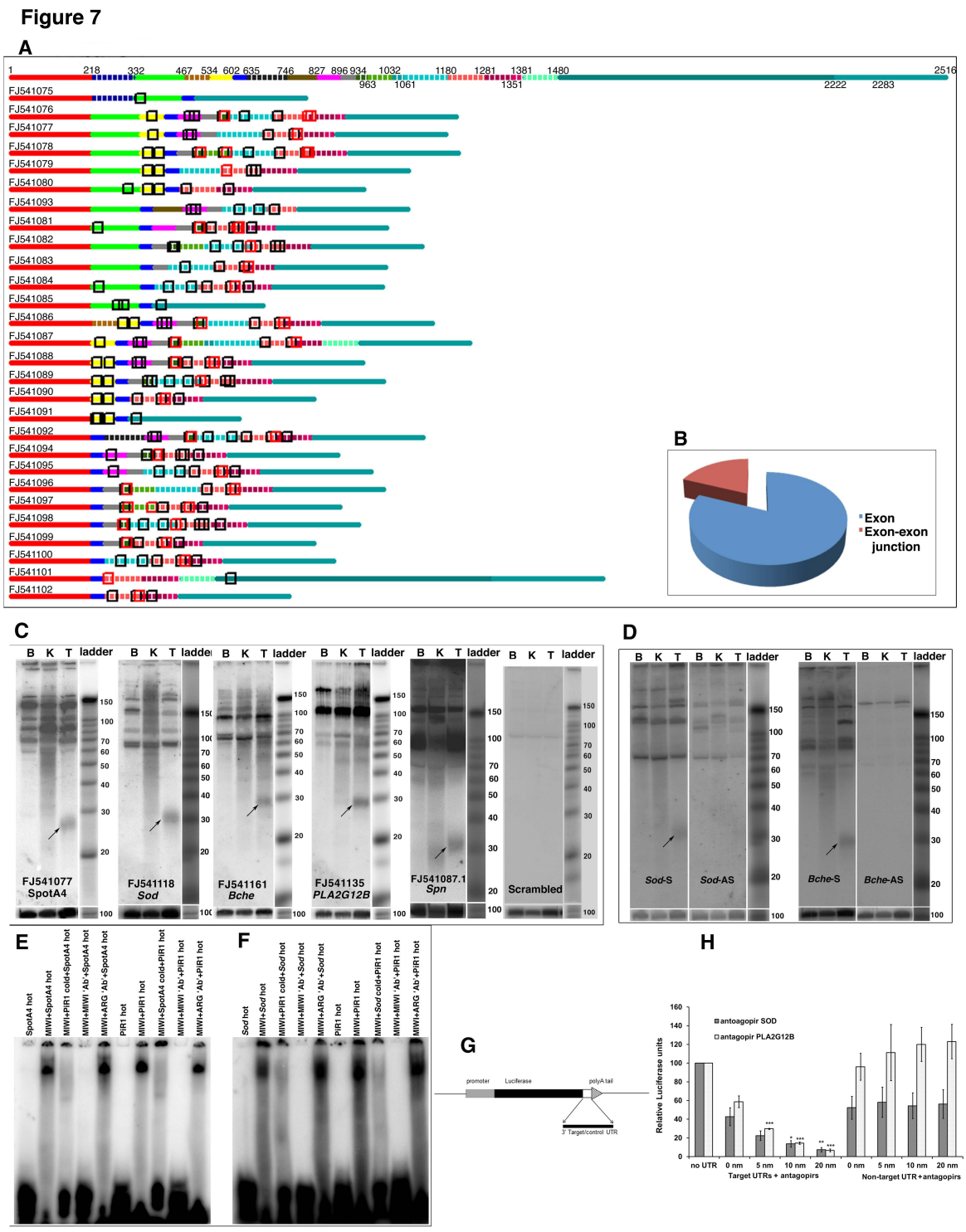


Figure 7. Identification of piRNAs in the splice variants of *Pirmy*. (*A*) Short stretches of homology identified in the UTRs by BLAST analysis against UTR database are represented on the set of 28 *Pirmy* splice variants. The boxes highlighted in black match within exons and boxes highlighted in red 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 UTRs of two genes, *Sod* and *PLA2G12B*, bring about concentration dependent reduction in luciferase expression. Use of these oligos against a non-target UTR (*Cdc2l1*) did not affect luciferase expression.

Identification of ~30nt RNAs from the splice isoforms of *Pirmy*

To check if these short stretches of homologies corresponded to small RNAs, few representative oligonucleotide sequences from *Pirmy* with homology to different UTRs were used as probes on small RNA northern blots. The gene corresponding to hypothetical protein spot A and superoxide dismutase (SOD) were chosen from the deregulated proteins; 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 two 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. Protocols standardized in the lab were used for northern blotting. Fig. S6 shows the location of the LNA (locked nucleic acid) probes used for small RNA northern blots on the corresponding splice variants.

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 protein spot 'A' (1700001L19 Riken cDNA Q9DAR0) and *Sod* are depicted in Fig. 7*E* and *F* respectively. piR1 (23), a known piRNA served as the positive control. Specificity of binding was

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

Pirmy identifies piRNAs from NCBI Archives database

Next line of evidence to the fact these short stretches of homologies are piRNAs came from the NCBI Archives database for piRNAs. BLAST analysis using the 108 *Pirmy* splice isoforms identified a total of 1445 small RNAs in the Sequence Read Archives SRP001701 and SRP000623 with 95% identity. *Pirmy* splice variants identified 352 MIWI2 - MILI-associated reads in SRP000623. Of these 310 had homology to midrepetitive sequences (SINEs, LINEs and LTRs) and 5 had homology to X chromosome. Thus, the 37 small RNAs that have homology only to the Y- transcripts could be piRNAs derived from the Y chromosome.

Antagopirs downregulate reporter gene expression

Complementary oligonucleotides synthesized to piRNA sequences present in UTRs of *Sod* and *PLA2G12B* were designated as antagopirs (Fig. S6). Cloning of these UTRs 3' to the Luciferase gene reduced its expression. Increasing concentrations of antagopirs, i. e. 5 nM, 10 nM and 20 nM caused further concentration dependent reduction in Luciferase expression (Fig. 7*H*). The antagopirs to *Sod* and *PLA2G12B* did not have an effect when the UTR from the *Cdc2l1* gene (non-target UTR) was cloned 3' to the Luciferase gene. Fig. 7*G* is a schematic representation of 3'UTR reporter construct.

Thus, the use of antagopirs to piRNA sequences present in UTRs of genes showed that these piRNAs can modulate gene expression.

Thus, this study 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 largest number of splice variants characterized for any transcript. These alternatively spliced Y-derived transcripts harbor 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 lncRNA. This is, to our knowledge the first report on possible regulation of autosomal genes involved in fertility and spermiogenesis, mediated by Y-encoded small RNAs/piRNAs in any species.

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 resulting in 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 harbor 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) (9, 24). Subsequently more transcripts were identified in mouse testis, using repeat sequences localizing to mouse Y (10, 11).

The *Pirmy* splice isoforms in the present study were discovered serendipitously by cloning and sequencing of the multiple RT-PCR products obtained using primers to the initial and terminal exons. As the primers were restricted to just two of the exons, it is possible that we might discover more isoforms using primers to different exons for RT-PCR amplification. Alternative splicing has been reported in ncRNAs (25), yet, such extensive splicing as observed in our study has not been reported for any of them. Very few polymorphically spliced genes have been described earlier from sex chromosomes, particularly the Y (26, 27). Thus, the 108 splice isoforms observed from RT-PCR amplification of *Pirmy* in this study appears to be by far the maximum number of isoforms characterized from a single lncRNA by alternative splicing.

Based on the mode of generation of mRNA isoforms, large number of variants has been predicted for a few genes. For example, nearly 100 mRNA isoforms have been characterized from the human basonuclin gene. This gene has the potential to generate nearly 90,000 mRNA isoforms using 6 promoters, 4 different poly adenylation sites and multiple alternative splicing (28). Large number of isoforms have also been deduced from a few characterized splice variants of Dscam (Downs syndrome cell adhesion molecule) in Drosophila (29), vertebrate olfactory receptors (30), neurexins (31) and cadherin related neuronal receptors (32).

Elucidation of identical isoforms in mouse brain and testis in this study (Figs. 2, 4) shows that these alternative splicing events are precise and not random because the

same splicing events take place in both the tissues. The consensus splice signal sequences present at the intron-exon junctions in all the isoforms further reaffirm programmed splicing events. The presence of piRNA homologous sequences at the exon-exon junctions of different splice isoforms posits alternative splicing as a mode of generation of more piRNAs from the same transcript (Data sheet).

The presence of *Pirmy* homology in the UTRs of genes corresponding to deregulated proteins in XY^{RIII}qdel mice suggests the regulation of these autosomal genes by sequences present in the deleted regions of the Y. Real-time PCR analysis showed that the transcripts corresponding to the proteins missing from the XY^{RIII}qdel sperm proteome were present in the testis of these mice. The fact that there is comparable expression of MAST, calreticulin and SPINK2 variant2 proteins in XY^{RIII} and XY^{RIII}qdel testes by western blot analysis, despite significantly upregulated expression of these transcripts in testis, could indicate regulation at the level of both transcription and translation (Figs. 5*B*, *C*).

Few other autosomal proteins are also deregulated in different strains of Y-deleted mice besides the ones identified in the proteomics screen; caldendrin is upregulated in XY^{RIII}qdel sperm (22) and acrosin is downregulated in B10.BR-Y^{del} sperm. (33). Aromatase is overexpressed in both the Y-deleted strains of mice i.e. XY^{RIII}qdel (our unpublished observation), and B10BR-Y^{del} (34). Therefore, it is not surprising to find *Pirmy* homologous sequences in the UTRs of caldendrin, acrosin and aromatase genes, suggesting Y- mediated regulation for these genes as well.

Ellis and colleagues also observed up or down regulation of genes from the X-chromosome and autosomes in testes of mice with deletions of Y heterochromatin using a microarray approach (35). Homology between UTRs of some of the above

genes and *Pirmy* splice isoforms (Table S2) further strengthens the hypothesis of putative regulation of genes located elsewhere in the genome by Y chromosomal repeats.

The concentration dependent reduction of Luciferase expression by antagopirs in these UTRs corroborates the regulation of these genes by Y-derived piRNAs (Fig. 7*H*). piRNAs regulate translation in early embryos and gonads besides containing transposable elements (36, 37). Previous studies in the lab elucidated another example of Y-ncRNA mediated regulation of an autosomal gene, *CDC2L2* via trans-splicing in human testis (3). The role of mouse Y heterochromatin in the current study therefore reveals a novel pathway for the regulation of autosomal genes by Y chromosome, mediated by piRNAs, in male reproduction.

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. FABP9 upregulated in XY^{RIII}qdel spermatozoa, localizes to perforatorium, the subacrosomal region in spermatozoa with falciform head shapes, mostly rodents (38-40). It is the most abundant protein of the perforatorium, and putatively has a role in shaping the unique sperm head structure of rodents (39, 40). Malformed head structure can affect cell motility and ability of spermatozoa to recognize the conspecific female gamete and other downstream activities associated with fertilization.

Calreticulin overexpressed in XY^{RIII}qdel spermatozoa is a calcium store associated with sperm functions such as hyperactivated motility, capacitation and acrosome reaction (41). The subsequent cascade of events could result in subfertility. Caldendrin yet another protein that is upregulated in XY^{RIII}qdel sperm (22), localizes to acrosome in rats and is considered to be a stimulus dependent regulator of calcium (42). SPINK2

variant3 localizes to the acrosome in mouse spermatozoa (our unpublished observations). The physiological function of SPINK2 is believed to be blocking of deleterious degradation of proteins released by acrosin from spermatozoa during acrosome reaction (43). Being a putative sperm acrosin inhibitor a role in fertility can be envisaged.

MAST a novel protein that is absent in the sperm proteome localizes to both the acrosome and sperm tail, indicating a role in the penetration of the egg and sperm motility (22). Bioinformatic analysis of the transcript (Riken cDNA 1700001L19) of the hypothetical protein from spot A predicts it as a cilia related gene (44). Based on this information putative role of this protein in sperm motility can be envisaged. Sdf2l1 is an ER stress-inducible gene, induced by the unfolded protein response pathway (45). The XY^{RIII}qdel spermatozoa may be more susceptible to stress induced damages as they lack stress response protein, SDF2L1. SOD, a protein upregulated in XY^{RIII}qdel, has been found to be positively associated with sperm concentration and overall motility (46).

The irreversible conversion of androgens to estrogens is catalyzed by aromatase transcribed by *Cyp19* gene (47). Biology behind the skewed sex ratio towards females in the progeny sired by XY^{RIII}qdel males could be explained by the upregulated expression of aromatase in testes of these mice. Acrosin plays a crucial role in acrosome exocytosis and egg zona pellucida penetration (48). Spermatozoa lacking acrosin ($acrosin^{-/-}$) exhibit delayed fertilization as both the processes of acrosome exocytosis and egg penetration are deferred (49, 50).

Functions of Y chromosome have been elucidated using different deletions of the chromosome in the past. Naturally occurring deletions in the euchromatic long arm of

Y chromosome in azoospermic men showed the involvement of this region in human male infertility (51). *Drosophila melanogaster* males with deletions of different regions of the Y chromosome show absence of several sperm axoneme proteins (52). Mice with partial or total deletions of Y heterochromatin show deregulation of testicular gene expression and subfertility/sterility (15, 35). Consolidation of the observations from mouse and human studies therefore shows that Y chromosome regulates autosomal genes expressed in testis using distinct mechanisms in different species.

Comparative sperm proteomics analysis in our study portrays involvement of multiple autosomal genes in subfertility. The regulation of autosomal gene expression appears to be relaxed in sperms of Yq-deleted mice. This reflects a connection between the Y chromosome and autosomes. In fact, as suggested by Piergentili, Y chromosome could be a major modulator of gene expression (4). Our results seem to provide explanation for some of the earlier classical observations of mice with different Y chromosomal deletions exhibiting subfertility/sterility along with sperm morphological abnormalities, fewer motile sperms, sex ratio skewed toward females etc. Similar phenotypes are also observed in cross-species male sterile hybrids of *Drosophila* and mouse (4, 53-59). Y-chromosome has also been implicated in the male sterility phenotype of these interspecies hybrids (56, 58, 60-62). Thus, the phenotypes observed in cross species male sterile hybrids and the Y-deletion mutants are comparable. Introduction of Y-chromosomes into different genetic backgrounds of *Drosophila* resulted in deregulated expression of hundreds of genes localizing to the X-chromosome and autosomes (63, 64). It has also been proposed that incompatibility between the Y-chromosomes and different autosomes could result in the hybrid

dysgenesis of sperm related phenotypes observed in *Drosophila* (61). Zouros and colleagues also suggested the presence of epistatic networks in interspecies hybrids, based on the fact that homospecific combination of alleles at a given set of loci could sustain normal development, but heterospecific combinations could not (60, 65, 66). This early hypothesis seems to be amply supported by our *in silico* and proteome analysis (Figs. 5-7). Further, our results elucidate the Y - derived piRNAs as the genetic basis of epistatic interactions between Y chromosome and autosomes in mouse. Our results also suggest for the first time, the mechanism of piRNA mediated regulation of autosomal genes involved in spermiogenesis and male fertility.

In brief, the XY^{RIII}qdel mutant strain of mouse, where there is a partial deletion of long arm of the Y chromosome, exhibit sperm morphological and motility related aberrations and subfertility. A comparative sperm proteomic profiling of the XY^{RIII} and XY^{RIII}qdel mice captured few differentially expressed proteins that could partially account for the aberrant sperm phenotype. Surprisingly, genes corresponding to the deregulated proteins localized to autosomes and not to the deleted region of the Y chromosome. Earlier we demonstrated an event of trans-splicing between a Y-ncRNA and a protein coding autosomal mRNA in human testis for putative translational regulation. A search for the Y-autosome connections in mouse led to the identification of novel ncRNAs from mouse Y long arm that subsequently was shown to regulate the genes expressed in testis via piRNAs. Thus, adopting a top-down approach we have established a novel mode of regulation by the Y chromosome of autosomal genes expressed in mouse testis and the biology behind the aberrant sperm phenotype in Y-deleted mice.

Finally, evolutionary impact of novel genetic interactions or regulatory mechanisms such as those reported in this study could be significant. The generation of piRNAs from species-specific repeats on mouse Y-chromosome that apparently regulate autosomal gene expression in testis raises more questions in the field of speciation and evolution. Do mutations in the Y chromosomal repeats collapse the poise of the species? Are species-specific repeats on the Y chromosome the fulcrum on which rests the fine balance between species identity and evolution?

METHODS

Animals and Reagents

The XY^{RIII} strain (wild type) and the XY^{RIII}qdel strain (Y-deletion mutant) of mice used in the study was a gift from Prof. Paul S Burgoyne, MRC, UK. This study could not have been done without the gift of these mice. All the animals used in the experiments were bred and reared in the in-house animal facility of our institute (CCMB), in accordance with the guidelines from Indian Science Academy under CPCSEA (Committee for the Purpose of Control and Supervision of Experimental Animals). This study was approved by the Institutional Animal Ethics Committee (IAEC 65/28/2006). The recombinant construct in pAAV-IRES-hrGFP from which MIWI protein was isolated was a gift from Arvind Kumar, CCMB, Hyderabad.

Electrophoretic mobility shift assay RNA oligonucleotides corresponding to GAAGCAGAUGAGUAUAUG from *Sod* and UCAUUGGACAUAAACUGAAUUUUCCA from the gene for hypothetical protein spot A were end labeled with γ-^{32}P ATP and column purified using G-25 Sephadex (Sigma-Aldrich) and quantitated on a scintillation counter. EMSA reactions were set up in a total volume of 25 µl using binding buffer (20 mM HEPES, 3 mM MgCl2, 40 mM KCl, 5% Glycerol, 2 mM DTT and RNase inhibitor

(4U)), with MIWI protein (5 μg per reaction). MIWI was over-expressed from a recombinant construct in pAAV-IRES-hrGFP vector and purified using the FLAG tag. Competitors i.e., unlabeled oligonucelotides (30 x concentration of hot oligo), MIWI antibody (90 ng) and Argonaute 3 antibody (100 ng) were added to the reaction, incubated for 1 h on ice, before addition of radio-labeled oligonucleotide (7000-10,000 cpm) and the entire mix was incubated on ice for another 30 min. EMSA was done on 5% native PAGE and image captured using FUJI phosphor Imager (FUJIFILM FLA-3000). A known piRNA, piR1, was used as the positive control, Argonaute 3 antibody served as the antibody control.

Identification of cDNA using M34 (DQ907163):

An amplified mouse testis cDNA library (Mouse testis MATCHMAKER cDNA Library – Clonetech) was screened with the male specific ES cell EST (CA533654) with sequence homology to the genomic clone DQ907163 (Fig. 1*E*) as per manufacturer's protocol. Mouse testis cDNA Library was screened according to standard protocol at a stringency of 2 x SSC, 0.1% SDS at 65^0C for 10 min. A total of 2 x 10^5 colonies were screened to obtain 18 clones after tertiary screening. Male-specificity of the positive clones from the library was determined using Southern blots containing mouse male and female DNA. All 18 clones gave the same sequence. The sequence was submitted to the NCBI database (DQ907162)**.**

Identification of Splice Variants: Total RNA (1 μg) isolated from brain and testes tissues each of XY^{RIII} strain of mouse were reverse transcribed with the (SuperscriptII, Reverse Transcriptase enzyme, Invitrogen), using oligo (dT) primers and random hexamers. The RT-mixes were concentration equalized using GAPDH primers. DQ907162 was amplified after two rounds of PCR. For the first round, forward

(GTGTGACAGGGTGGGGAATC) and reverse primers (TTCCTGAAGATAGCACTTGTG), and the following conditions were used - initial denaturation 95^{0}C, 1min, cycle denaturation 95^{0}C, 1 min, annealing 62^{0}C, 30 sec and extension 72^{0}C for 2min (35cycles), final extension 72^{0}C for 7 min. The second round of amplification was done using nested primers GAGGACCGTATTCATGGAAGAG (forward) and GCAAATGGCTCACATCAGTGG (reverse) using initial denaturation at 95^{0}C for 1 min, cycle denaturation at 95^{0}C, 1 min, annealing at 66^{0}C for 30 sec and extension at 72^{0}C for 2 min (38 cycles) and final extension at 72^{0}C for 7 min. Annealing temperatures up to 70^{0}C yielded multiple products. The multiple products obtained after two rounds of PCR from both testis and brain (Fig. 2*B*) were cloned into pCR TOPO vector using Invitrogen TOPO TA cloning kit. Approximately 1000 clones were sequenced on 3730 DNA Analyser (ABI Prism) using sequencing kit BigDye Terminator V3.1 Cycle Sequencing kit.

Small RNA Isolation

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

Northern Blotting

20-50 µg of small RNA from each tissue was resolved on a 12% Urea PAGE gel in 0.5 x TBE running buffer and transferred onto Hybond N^+ membrane. Decade marker (Ambion) was labeled and loaded according to the manufacturer's instructions. 10-25 µM of each LNA- oligonucleotide probe (Exiqon), was end labeled for use as probes (hybridization buffer - 5 x SSC, 5 x Denhardt's and 1% SDS). Blots were hybridized at 37^0C and washed from 37^0C to 65^0C in 2 x SSC, 0.2% SDS depending on the intensity of the signal. U6 was used as the loading control.

Real-time PCR analysis

Copy number estimation of *Pirmy*

Genomic DNA was isolated from testes of XY^{RIII} and XY^{RIII}qdel mice using phenol-chloroform method and quantified using a Nanodrop (NANODROP 2000, ThermoScientific). Quantitative Real-time PCR (LightCycler 480, Roche) was performed using SYBR green master mix (Qiagen) with 2 ng of genomic DNA and a primer concentration of 200 nM per reaction. The primers used were as follows: *Pirmy:* forward 5'-GTG CGG TTG TGA AGG TGT TC– 3', reverse 5'-CCT CCA CCT TCC ATT CAC CC-3'; *Gapdh:* forward 5'-ACG GGA AGC TCA CTG GCA TGG -3', reverse 5'-CAA CAG CGA CAC CCA CTC CTC–3'. PCR conditions included an initial denaturation for 5 min at 95°C followed by 45 cycles of denaturation at 95°C for 10 secs, annealing at 58°C for 20 sec and elongation at 72°C for 30 sec. The amplification of specific product was confirmed by melting curve profile (cooling the sample to 65°C for 1 min and heating slowly with an increase in temperature of 5°C at each step till 95°C, with continuous measurement of fluorescence). The relative fold change in *Pirmy* was analyzed based on Livak method ($2^{-\Delta\Delta Ct}$).

Real-time PCR analysis of differentially expressed genes

The total RNA was extracted from testes using Trizol (Ambion). One µg of RNA was reverse transcribed to cDNA using Verso cDNA synthesis kit (ThermoScientific). qPCR was performed using SYBR green master mix (Qiagen) and analysed in Roche Light Cycler LC480. The primer sequences corresponding to mouse cDNA used were as follows: forward 5'-CGA GGG CCA GAC AGG GAT TG–3' and reverse 5'-CCC ATA GAC AGA GGA CAT CAG- 3' for Riken cDNA 1700001L19; forward 5'-ACT TCC CGT CGC CGC TAT C-3' and reverse 5'-TGA CCG ACA GGA ACA CAG AGG-3' for *Sdf2l1*; forward 5'-CAG CAT CGA GCA GAA GTA TAA GC-3' and reverse 5'-TGG GTG GAG TTA TTG CAG TAG-3' for *Mast*; forward 5'-GGA AAC CAC GTC AAA TTG -3' and reverse 5'-GGT GAT GAG GAA ATT GTC-3' for Calreticulin; forward 5'-GGC TAC TTG ACC ACT GC-3' and reverse 5'-TTT GAG AAT CGG AAG AGT C-3' for *Spink2* Variant 2; forward 5'-TTC CGA ACA CCA GAC TG-3' and reverse 5'-ATG GCT ACC GTC CTC C-3' for *Spink2* Variant 3; forward 5'-GCA AGC GGT GAA CCA GTT-3' and reverse 5'-GCC ACC ATG TTT CTT AGA-3' for *Sod*; forward 5'-TGA AGT CGC AGG AGA CAA CCT-3' and reverse 5'-ATG GCC TTC CGT GTT CCT A-3' for *Gapdh*. PCR conditions included an initial denaturation for 5 min at 95°C followed by 45 cycles of denaturation at 95°C for 10 secs, annealing at 58°C for 20 sec and elongation at 72°C for 30 sec. The amplification of specific product was confirmed by melting curve profile (cooling the sample to 65°C for 1 min and heating slowly with an increase in temperature of 5°C at each step till 95°C, with continuous measurement of fluorescence). The relative fold change in expression was estimated based on Livak method ($2^{-\Delta\Delta Ct}$).

Luciferase assay

The 5'UTR sequences from the human clone RP11-130C19 and the Gapdh genes were amplified from mouse cDNA and cloned into the Luciferase vector. Either Luciferase

gene alone or luciferase along with each UTR was cloned into pcDNA3.1 expression vector for assaying the effect of antagopirs (Figure 7H) on Luciferase expression. Co-transfection experiments were done using the GC-1spg cell line (ATCC CRL-2053) and lipofectamine 2000 (Invitrogen) using protocols specified by the manufacturer. Cells were seeded in 48-well plates, 24 hrs prior to transfection to obtain approximately 80% confluency. Each well was transfected with 50 ng of pcDNA3.1 plasmid containing either Luciferase gene alone or along with the cloned UTRs, 50 ng of the β-gal plasmid and varying concentrations of oligonucleotides complementary to the piRNA along with 0.5µl of lipofectamine 2000 in antibiotic and serum free DMEM (GIBCO). The complementary oligonucleotide to the piRNA has been designated as antagopirs. The antagopirs (*Sod* - 5'GAAGCAGAUGAGUAUAUG3'; *PLA2G12B* - 5'CCAAACUGUUGGAAGAAGGAAU3') were procured as RNA oligonucleotides from Eurofins Genomics India Pvt. Ltd, Bangalore, India. Different concentrations of antagopirs (0 nM, 10 nM, 20 nM and 40 nM/ well) were tested in the assay for their effect on the UTRs. Five hrs post transfection, the medium was replaced with complete growth medium. The cell extracts were prepared 24 hrs post transfection using Reporter Lysis Buffer (Promega) and assayed for Luciferase activity in EnSpire 2300 multimode plate reader (Perkin Elmer). The Luciferase activity was normalized using β galactosidase. Three independent sets of experiments were done in triplicates.

Sperm Proteome Analysis: Sperm lysates (1 mg of cell weight per 5 µl of lysis buffer – [8 M Urea, CHAPS 4% (w/v), 40 mM Tris, Biolyte (3-10) 0.2% (w/v) and TBP (1µl/100 µl]). The suspension was incubated on ice for one hour to allow buffer to permeabilise and lyse the sample. Further, the sample was briefly sonicated on ice. The lysate was centrifuged for 15 min at 13,000 rpm at 4^0C. The supernatant was collected and was further taken for ultra-centrifugation at 55,000 rpm for one hour at 4^0C. The clear

lysate was collected in fresh tube and PMSF added to a final concentration of 1mM. The protein concentration in the cell lysate was estimated by Bicinchoninic acid assay (Pierce, Rockford IL) following manufacturer's instructions in a micro titer plate. BSA was used as the standard for estimation. The proteins were separated in the first dimension on 4-7 and 5-8 IPG strips (Bio-Rad). The strips were then loaded onto 8-20% gradient PAGE for separation on the second dimension. Protein spots were visualized by Coomassie Blue staining. Spot to spot matches were done to identify differences. Analysis of five sets of gels after normalization with control spots using PDQUEST software version 6.0 (Bio-Rad) identified the differential proteins. Measuring the optical density of these differentially expressed proteins in arbitrary units validated the quantitative differences (Fig. S7). These values were subjected to nonparametric Kruskal-Wallis H test and the levels of confidence determined by the Chi-squared test (75-95% degrees of confidence). Trypsin digested spots were processed to obtain the protein tags by MS analysis on Hybrid Quadrupole TOF mass spectrometer (API QSTAR PULSAR i, PE SCIEX).

DATA ACCESS

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

ACKNOWLEDGEMENTS

We gratefully acknowledge the gift of the XY^{RIII} and XY^{RIII}qdel mice by Prof. Paul S Burgoyne, MRC, Edinburgh, UK. The GC-1spg cell line was gifted by Prof. MRS Rao, JNCASR, Bangalore, India. We thank Dr. Dinesh Kumar and Professor B. K. Thelma for reading the manuscript and giving useful 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, SR 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:

1. Tiepolo L & Zuffardi O (1976) Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Human genetics* 34(2):119-124.
2. Kuroda-Kawaguchi T*, et al.* (2001) The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature genetics* 29(3):279-286.
3. Jehan Z*, et al.* (2007) Novel noncoding RNA from human Y distal heterochromatic block (Yq12) generates testis-specific chimeric CDC2L2. *Genome research* 17(4):433-440.
4. Piergentili R (2010) Multiple roles of the Y chromosome in the biology of Drosophila melanogaster. *The Scientific World Journal* 10:1749-1767.
5. Soh YQ*, et al.* (2014) Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* 159(4):800-813.
6. Moretti C*, et al.* (2017) SLY regulates genes involved in chromatin remodeling and interacts with TBL1XR1 during sperm differentiation. *Cell Death Differ* 24(6):1029-1044.
7. Bellott DW*, et al.* (2014) Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* 508(7497):494-499.
8. Burgoyne PS & Mitchell MJ (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.
9. Conway SJ*, et al.* (1994) Y353/B: a candidate multiple-copy spermiogenesis gene on the mouse Y chromosome. *Mamm Genome* 5(4):203-210.

10. Prado VF, Lee CH, Zahed L, Vekemans M, & Nishioka Y (1992) Molecular characterization of a mouse Y chromosomal repetitive sequence that detects transcripts in the testis. *Cytogenetics and cell genetics* 61(2):87-90.
11. Toure A*, et al.* (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.
12. 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.
13. Toure A*, et al.* (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.
14. Burgoyne PS, Mahadevaiah SK, Sutcliffe MJ, & Palmer SJ (1992) Fertility in mice requires XY pairing and a Y-chromosomal "spermiogenesis" gene mapping to the long arm. *Cell* 71(3):391-398.
15. Cocquet J*, et al.* (2009) The multicopy gene Sly represses the sex chromosomes in the male mouse germline after meiosis. *PLoS Biol* 7(11):e1000244.
16. Reynard LN, Cocquet J, & Burgoyne PS (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.
17. Choi E*, et al.* (2007) Integrative characterization of germ cell-specific genes from mouse spermatocyte UniGene library. *BMC Genomics* 8:256.
18. Schultz N, Hamra FK, & Garbers DL (2003) A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proceedings of the National Academy of Sciences of the United States of America* 100(21):12201-12206.
19. 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.
20. Bajpai A, Sridhar S, Reddy HM, & Jesudasan RA (2007) BRM-Parser: a tool for comprehensive analysis of BLAST and RepeatMasker results. *In silico biology* 7(4-5):399-403.
21. Singh L, Panicker SG, Nagaraj R, & Majumdar KC (1994) Banded krait minor-satellite (Bkm)-associated Y chromosome-specific repetitive DNA in mouse. *Nucleic Acids Res* 22(12):2289-2295.
22. Bhattacharya R, Devi MS, Dhople VM, & Jesudasan RA (2013) A mouse protein that localizes to acrosome and sperm tail is regulated by Y-chromosome. *BMC cell biology* 14(1):50.
23. Girard A, Sachidanandam R, Hannon GJ, & Carmell MA (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442(7099):199-202.
24. Bishop CE & Hatat D (1987) Molecular cloning and sequence analysis of a mouse Y chromosome RNA transcript expressed in the testis. *Nucleic acids research* 15(7):2959-2969.
25. Pang KC*, et al.* (2005) RNAdb--a comprehensive mammalian noncoding RNA database. *Nucleic Acids Res* 33(Database issue):D125-130.

26. Wegmann D, Dupanloup I, & Excoffier L (2008) Width of gene expression profile drives alternative splicing. *PloS one* 3(10):e3587.
27. Ellis PJ, Ferguson L, Clemente EJ, & Affara NA (2007) Bidirectional transcription of a novel chimeric gene mapping to mouse chromosome Yq. *BMC Evol Biol* 7:171.
28. Vanhoutteghem A & Djian P (2007) The human basonuclin 2 gene has the potential to generate nearly 90,000 mRNA isoforms encoding over 2000 different proteins. *Genomics* 89(1):44-58.
29. Schmucker D*, et al.* (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101(6):671-684.
30. Wang F, Nemes A, Mendelsohn M, & Axel R (1998) Odorant receptors govern the formation of a precise topographic map. *Cell* 93(1):47-60.
31. Missler M & Sudhof TC (1998) Neurexins: three genes and 1001 products. *Trends in genetics : TIG* 14(1):20-26.
32. Wu Q & Maniatis T (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 97(6):779-790.
33. Styrna J, Klag J, & Moriwaki K (1991) Influence of partial deletion of the Y chromosome on mouse sperm phenotype. *J Reprod Fertil* 92(1):187-195.
34. 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 Histochem* 106(1):55-64.
35. Ellis PJ*, et al.* (2005) Deletions on mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids. *Hum Mol Genet* 14(18):2705-2715.
36. Aravin AA, Hannon GJ, & Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318(5851):761-764.
37. Deng W & Lin H (2002) miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell* 2(6):819-830.
38. 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.
39. Oko R & Morales CR (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.
40. Pouresmaeili F, Morales CR, & 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.
41. 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.
42. Redecker P, Kreutz MR, Bockmann J, Gundelfinger ED, & Boeckers TM (2003) Brain synaptic junctional proteins at the acrosome of rat testicular germ cells. *J Histochem Cytochem* 51(6):809-819.
43. 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.

44. McClintock TS, Glasser CE, Bose SC, & Bergman DA (2008) Tissue expression patterns identify mouse cilia genes. *Physiol Genomics* 32(2):198-206.
45. Fukuda S*, et al.* (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.
46. Lu JC, Huang YF, & Lu NQ (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.
47. Simpson ER*, et al.* (1994) Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis. *Endocr Rev* 15(3):342-355.
48. Yamagata K*, et al.* (1998) Acrosin accelerates the dispersal of sperm acrosomal proteins during acrosome reaction. *J Biol Chem* 273(17):10470-10474.
49. Adham IM, Nayernia K, & Engel W (1997) Spermatozoa lacking acrosin protein show delayed fertilization. *Mol Reprod Dev* 46(3):370-376.
50. Mao HT & Yang WX (2013) Modes of acrosin functioning during fertilization. *Gene* 526(2):75-79.
51. Vogt PH (1998) Human chromosome deletions in Yq11, AZF candidate genes and male infertility: history and update. *Mol Hum Reprod* 4(8):739-744.
52. Goldstein LS, Hardy RW, & Lindsley DL (1982) Structural genes on the Y chromosome of Drosophila melanogaster. *Proc Natl Acad Sci U S A* 79(23):7405-7409.
53. Heikkinen E & Lumme J (1998) The Y chromosomes of Drosophila lummei and D. novamexicana differ in fertility factors. *Heredity (Edinb)* 81 (Pt 5):505-513.
54. Johnson NA, Hollocher H, Noonburg E, & Wu CI (1993) The effects of interspecific Y chromosome replacements on hybrid sterility within the Drosophila simulans clade. *Genetics* 135(2):443-453.
55. Tao Y, Zeng ZB, Li J, Hartl DL, & Laurie CC (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.
56. 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.
57. Albrechtova J*, et al.* (2012) Sperm-related phenotypes implicated in both maintenance and breakdown of a natural species barrier in the house mouse. *Proceedings. Biological sciences / The Royal Society* 279(1748):4803-4810.
58. Campbell P, Good JM, Dean MD, Tucker PK, & Nachman MW (2012) The contribution of the Y chromosome to hybrid male sterility in house mice. *Genetics* 191(4):1271-1281.
59. White MA, Stubbings M, Dumont BL, & Payseur BA (2012) Genetics and evolution of hybrid male sterility in house mice. *Genetics* 191(3):917-934.
60. 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 (Edinb)* 76 (Pt 6):603-609.

61. 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.
62. Carvalho AB, Vaz SC, & Klaczko LB (1997) Polymorphism for Y-linked suppressors of sex-ratio in two natural populations of Drosophila mediopunctata. *Genetics* 146(3):891-902.
63. David JR, *et al.* (2005) Male sterility at extreme temperatures: a significant but neglected phenomenon for understanding Drosophila climatic adaptations. *J Evol Biol* 18(4):838-846.
64. Lemos B, Araripe LO, & Hartl DL (2008) Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science* 319(5859):91-93.
65. Davis AW, Noonburg EG, & Wu CI (1994) Evidence for complex genic interactions between conspecific chromosomes underlying hybrid female sterility in the Drosophila simulans clade. *Genetics* 137(1):191-199.
66. Wu CI & Palopoli MF (1994) Genetics of postmating reproductive isolation in animals. *Annu Rev Genet* 28:283-308.