

SCIENTIFIC REPORTS



OPEN

Identification of piRNAs and piRNA clusters in the testes of the Mongolian horse

Bei Li¹, Xiaolong He², Yiping Zhao¹, Dongyi Bai¹, Gerelchimeg Bou¹, Xinzhuang Zhang¹, Shaofeng Su¹, Leng Dao¹, Rui Liu¹, Yuejiao Wang¹ & Dugarjaviin Manglai¹

P-element induced wimpy testis-interacting RNAs (piRNAs) are essential for testicular development and spermatogenesis in mammals. Comparative analyses of the molecular mechanisms of spermatogenesis among different organisms are therefore dependent on accurate characterizations of piRNAs. At present, little is known of piRNAs in non-model organisms. Here, we characterize piRNAs in the Mongolian horse, a hardy breed that reproduces under extreme circumstances. A thorough understanding of spermatogenesis and reproduction in this breed may provide insights for the improvement of fecundity and reproductive success in other breeds. We identified 4,936,717 piRNAs and 7,890 piRNA clusters across both testicular developmental stages. Of these, 2,236,377 putative piRNAs were expressed in the mature samples only, and 2,391,271 putative piRNAs were expressed in the immature samples only. Approximately 3,016 piRNA clusters were upregulated in the mature testes as compared to the immature testes, and 4,874 piRNA clusters were downregulated. Functional and pathway analyses indicated that the candidate generating genes of the predicted piRNAs were likely involved in testicular development and spermatogenesis. Our results thus provide information about differential expression patterns in genes associated with testicular development and spermatogenesis in a non-model animal.

P-element induced wimpy testis (PIWI) proteins are a subfamily of Argonaute/PIWI proteins that are mainly expressed in the nuclei and cytoplasm of animal germ cells^{1–4}. PIWI-interacting RNAs (piRNAs) are 26–32-nt PIWI-binding small noncoding RNAs that exhibit significant strand bias⁵. piRNA sequences are generated by repetitive sequences in the genome, which are distributed in clusters called piRNA clusters⁶. The distribution of piRNA clusters on different chromosomes is not uniform and is not proportional to the length of the chromosome⁶.

Cloning of miRNAs in mouse testis identified 381 (~27 nt) putative piRNAs, while only 40 putative miRNAs (~22 nt) were identified in mouse oocytes, suggesting that piRNAs may have a specific role in the germlines of male mammals^{7,8}. Two populations of piRNAs are expressed during the development and differentiation of mouse spermatogenic cells: one population of “classic” piRNAs that silences retrotransposons, and one population of “pachytene” piRNAs, generated from nontransposon intergenic regions primarily located in the pachytene spermatocytes⁹. The functions of this second population of piRNAs remain unknown. Pachytene piRNA characterization is therefore crucial for investigations of PIWI protein activity during mammalian spermatogenesis. In mice, pachytene piRNAs are involved in the elimination of large amounts of mRNA from the elongating spermatids¹⁰. In the germ cells of male mice, pre-pachytene piRNAs interact with the PIWI proteins MIWI2 and MILI1^{11–14}. These proteins, in combination with transposons, retrotransposons, and other mobile genetic elements, ensure the normal development and differentiation of spermatogenic cells by inhibiting the activity of transposable elements at the epigenetic and post-transcriptional levels¹⁵. Deletion of either MIWI2 or MILI1 from the germ cells of fetal mice resulted in a significantly lower level of retrotransposon de novo DNA methylation compared to wildtype mice, indicating that the PIWI/piRNA pathway contributes to the horizontal transfer of silent transposable elements¹⁶.

¹College of animal science, Inner Mongolia Key Laboratory of Equine Genetics, Breeding and Reproduction, Scientific Observing and Experimental Station of Equine Genetics, Breeding and Reproduction, Ministry of Agriculture and Rural Affairs, Equine Research Center, Inner Mongolia Agricultural University, Hohhot, 010018, China. ²Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences, Hohhot, 010031, P. R. China. Correspondence and requests for materials should be addressed to D.M. (email: dmanglai@163.com)

Studies of piRNAs in horses and other non-model organisms are limited, thus hindering our understanding of the gene expression profiles and molecular mechanisms relating to deformation and maturation during spermatogenesis in such species. We selected the Mongolian horse for piRNA characterization, as this breed is particularly ancient, possibly expressing a phenotype ancestral to other Chinese, Japanese, and even Northern European horse breeds^{17–19}. In addition, this breed has high endurance and is unusually hardy compared to other horses. Mongolian horses are capable of thriving in a harsh, cold, arid climate with poor grazing opportunities; horses reproduce in extreme conditions with little shelter and little provender¹⁸. A thorough understanding of spermatogenesis and reproduction in this breed may provide insights for the improvement of fecundity and reproductive success in other breeds. In particular, knowledge of how piRNA affect spermatogenesis in the Mongolian horse may provide a framework against which to compare other, less hardy horse breeds. Therefore, in this study we aimed to characterize the piRNAs from the testes of the Mongolian horse.

Materials and Methods

Sample collection. All experiments involving animals complied with the Animal Care Guidelines set out in the Declaration of the Institutional Animal Ethics Committee of the Inner Mongolia Agricultural University, Hohhot, Inner Mongolia, China, and were authorized by that committee. All possible care was taken to minimize animal suffering.

We received permission from the owner to geld six healthy male Mongolian horses in Xilingol League, Inner Mongolia, China. The ages of the horses were determined based on a physical examination of their teeth, and on information from the owner. Three colts (samples BS1–3) were between 11 and 13 months old, and three were between three and four years old (samples AS1–3). We surgically collected the testes of all six horses. Removed testes were stored in cryogenic vials with an RNA/DNA sample protector (Takara, Dalian, China). Small samples (~5 g) from the testes of each animal were immediately frozen in liquid nitrogen for quantitative real-time polymerase chain reaction (qPCR) analysis.

Anecdotal evidence suggests that male Mongolian horses are incapable of reproduction before 18 months^{20,21}, and domesticated Mongolian horses are typical bred starting at age three (indicating sexual maturity)^{20,21}. However, to confirm the sexual maturity of the colts from which the testes were taken, all testes were examined histologically.

Construction of small RNA libraries and sequencing. We isolated total RNA from 100 mg of testicular tissue from each animal using TRIzol reagents (Invitrogen, Carlsbad, CA, USA). We tested each sample of total RNA for degradation and contamination with 1% agarose gels. We measured the purity and concentration of the total RNA samples with a NanoPhotometer spectrophotometer (Implen, Munich, Germany) and a Qubit RNA assay kit in a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA), respectively. We measured the integrity of each total RNA sample with an Agilent RNA Nano 6000 assay kit on an Agilent Bioanalyzer 2100 system (Agilent Technologies, Palo Alto, CA, USA). We rejected any total RNA samples with a ratio of optical density (OD) at 260 nm to OD at 280 nm ($OD_{260}/280$) < 1.7; a ratio of OD at 260 nm to OD at 230 nm ($OD_{260}/230$) > 2.0; concentration < 300 ng/ μ L; or integrity < 7.

We used 3 μ g total RNA from each sample as a template for library preparation with the NEBNext Multiplex Small RNA Library Prep Set for Illumina (NEB, Ipswich, MA, USA), following the manufacturer's instructions. We built six libraries of small RNAs, one per sample. We used index codes to link each sequence to one of the six samples. In brief, the NEB 3' SR adaptor was ligated to the 3' ends of all miRNAs, siRNAs, and piRNAs. Following this ligation reaction, the SR RT primer was hybridized to an overabundance of 3' SR adaptors (ensuring that some 3' SR adaptors remained unligated), converting any remaining single-stranded DNA adaptors into double-stranded DNA molecules, and avoiding the formation of adaptor dimers. The 5' SR adaptors were then ligated to the 5' ends of the miRNAs, siRNAs, and piRNAs; as double-stranded DNA is not a substrate of T4 RNA Ligase 1, these were not ligated. We used RNase H- reverse transcriptase (NEB, Ipswich, MA, USA) to synthesize first-strand cDNA, and performed PCR using LongAmp Taq 2X Master Mix, SR Illumina primers, and index (X) primers. The PCR volume contained 2 μ L of 10 μ M SR Illumina primer, 2 μ L of 10 μ M index (X) primer, 2 μ L of template DNA, and 25 μ L of LongAmp Taq 2X Master Mix, made up to 50 μ L with nuclease-free water. The cycling program was as follows: initial denaturation at 94 °C for 30 seconds; 30 cycles of 94 °C for 10 seconds, 50 °C for 30 seconds, and 65 °C for 50 seconds; and a final extension at 65 °C for 10 minutes. PCR products were separated on 8% polyacrylamide gels (100 V for 80 min). DNA fragments between 140 bp and 160 bp (the lengths of the small noncoding RNAs plus the 3' and 5' adaptors) were retrieved and dissolved in 8 μ L elution buffer. cDNA library quality was evaluated using an Agilent Bioanalyzer 2100 System with High Sensitivity DNA Chips (Agilent Technologies, Palo Alto, CA, USA). The library preparations were sequenced on an Illumina HiSeq2500 platform, and 50 bp single-end reads were generated.

Annotation of small RNAs. To produce clean sequence reads, we removed any reads containing polyA/T/G/C sections; any reads with 5' adaptor contamination; any reads without 3' adaptors; any reads without insert tags; and any reads of inferior quality (the quality values $Q < 20$ of the base number accounts for more than 30% of the total read reads.) from raw fastq. We confirmed that each small RNA mapped to a single annotation by performing annotations in a particular order, and removing small RNAs once they had been mapped. The order in which we performed annotations was known miRNAs, rRNAs, tRNAs, snRNAs, snoRNAs, repeats, genes, and novel miRNAs (Supplementary Table 1). The piRNA and piRNA cluster were analyzed based on data “repeat” and “other”.

Identification of piRNAs and piRNA-generating genes. Determination of the best method for piRNAs in non-model organisms, such as the horse, is a difficult and unsolved problem. Reference piRNA sequences

are available for only six model species (human, mouse, rat, *Drosophila*, zebrafish, and duckbill platypus; piRNA-Bank; <http://pirnabank.ibab.ac.in>)²². Here, we used k-mer methods to identify piRNA sequences²³. After aligning the repetitive sequences of identifying all piRNAs with the reference sequence, and then aligning the piRNA sequence without aligning the above repetitive sequence with the gene sequence of the reference genome, the piRNA-generating gene was obtained from the sam file based on alignment. The alignment software and parameters were bowtie ($-v0-k1$)²⁴.

We analyzed the Gene Ontology (GO) of the all piRNA-generating genes using a GSeq-based Wallenius non-central hyper-geometric distribution²⁵, which takes into account gene length bias. We used KOBAS²⁶ to identify Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways that were significantly enriched in the candidate piRNA generating genes.

Identification and functional annotation of piRNA clusters. The bam alignment results of identifying all piRNAs on the reference genome sequence were obtained based on bowtie-v 0-k 1 alignment, and the coverage of piRNA on the reference sequence was obtained by using samtools depth. Only the coverage ≥ 2 piRNA was retained for subsequent analysis. The minimum length of piRNA cluster length is 200, and the threshold distance of interval length is 10000. We then determined the lengths of all piRNAs clusters as described above^{5,15}. We then extended the range of each piRNA cluster to consider the sequences 2000 bp upstream and downstream using Perl script⁵. In this way, we detected neighboring gene piRNA clusters.

We measured the differential expression of each piRNA cluster between the mature and immature testes with DESeq^{27,28} in R v1.18.0²⁹, setting padj < 0.05 as cutoffs. DESeq uses a model based on negative binomial distributions to identify differential piRNA cluster expression based on digital piRNA cluster expression data. Three biological replicates were performed for each sample.

Validation of DESeq results. We validated the differential expression of eight randomly-selected piRNAs (uniq_1215060, uniq_4231549, uniq_1214880, uniq_1217459, uniq_7619, uniq_1214806, uniq_1447220, and uniq_1214700) with qPCR. To do this, we first extracted total miRNA from each of the six samples of previously frozen testicular tissue with a miRNeasy mini kit (Qiagen, Dusseldorf, Germany), following the manufacturer's instructions. We resuspended the total miRNA in nuclease-free water and measured miRNA concentration with a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA). We used $\sim 0.5 \mu\text{g}$ total miRNA as a template for the synthesis of first-strand cDNA with a miScript II RT Kit (Qiagen, Dusseldorf, Germany), following the manufacturer's instructions. We diluted the cDNA to $0.1 \mu\text{g}/\mu\text{L}$ using RNAase-free water, following the instructions in the miRNeasy mini kit (Qiagen, Dusseldorf, Germany). Using the diluted cDNA as a template, we measured the expression of the eight piRNAs on an MX3000P Real-Time PCR System (Agilent Technologies, Palo Alto, CA, USA) with a miScript SYBR Green PCR Kit (Qiagen, Dusseldorf, Germany), following the manufacturer's instructions. We used U6 as an internal reference gene to control for differences among samples. Although it has been suggested that U6 is not a suitable endogenous control³⁰, our preliminary results suggested that U6 was reliable, as the cycle threshold values were uniform across samples, with a single smooth peak. We determined relative piRNA expression with the $2^{-\Delta\Delta\text{Ct}}$ method³¹. We measured significant differences in relative piRNA expression between the mature and immature horses using a one-way ANOVA in SAS v9.0 (SAS Institute Inc., Cary, NC, USA). We considered $P < 0.05$ statistically significant.

Ethics approval and consent to participate. All procedures involving animals were approved and authorized by the Inner Mongolia Agricultural University. All experiments and methods were carried out according to guidelines and regulations of Inner Mongolia Agricultural University.

Results

Confirmation of sexual maturity and immaturity of testes. In the testis from colts BSM1-3, the seminiferous tubules had only single layers of germ cells, and were separated by interstitial cells (Fig. S1). Most of these cells were undifferentiated spermatogonia, although some spermatocytes were observed. No mature sperm were observed. Interstitial cells were present between seminiferous tubules. We thus concluded that horses BSM1-3 were sexually immature.

In the testis from colts ASM1-3, numerous germ cells in multiple layers were observed in the seminiferous tubule lumens. Mature sperm, spermatogonia, and spermatocytes were clearly visible (Fig. S1). We thus concluded that horses ASM1-3 were sexually mature.

Putative piRNAs and candidate piRNA-generating genes. We generated 843,017–1,287,314 raw reads for each of the six testicular samples (BS1: 843,017; BS2: 1,287,314; BS3: 941,342; AS1: 1,141,460; AS2: 1,126,573; and AS3: 1,141,984). We identified 4,936,717 unique putative piRNAs across all six libraries (Supplementary Table 2). Of these, 2,236,377 putative piRNAs were only expressed in the mature samples, and 2,391,271 putative piRNAs were only expressed in the immature samples. The putative piRNAs were 26–32 nt long. We observed a strong preference for uridine (U) at the 5' end, and for adenine (A) at the 10th position (Fig. 1).

We predicted 30,639 generating genes for our 4,936,717 putative piRNAs (Supplementary Table 3). After eliminating repeated genes, unknown genes, and non-coding genes, 6921 known protein-coding genes remained. We classified the protein-coding genes related to spermatogenesis by the four types of proteins encoded: zinc finger proteins (targeted by piRNAs including uniq_2329918, uniq_2329932, and uniq_2329970), microtubule-associated proteins (targeted by piRNAs including uniq_2330366, uniq_2330761, and uniq_2337467), spermatogenesis-associated proteins (targeted by piRNAs including uniq_2338707, uniq_2339388, and uniq_2340251), and sperm-associated antigens (targeted by piRNAs including uniq_2343853, uniq_2343741, and uniq_2347690). We considered the genes encoding sperm-tail PG-rich repeat

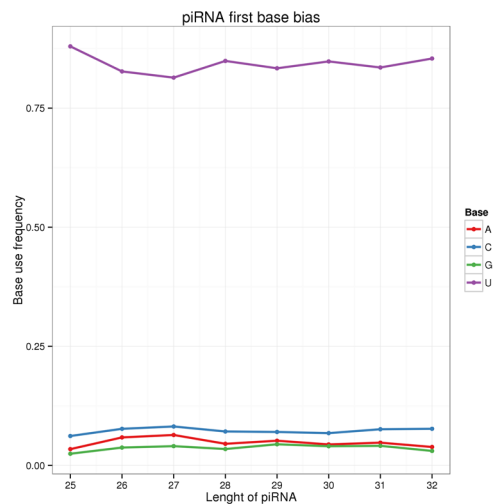


Figure 1. First base preference of piRNAs of different lengths.

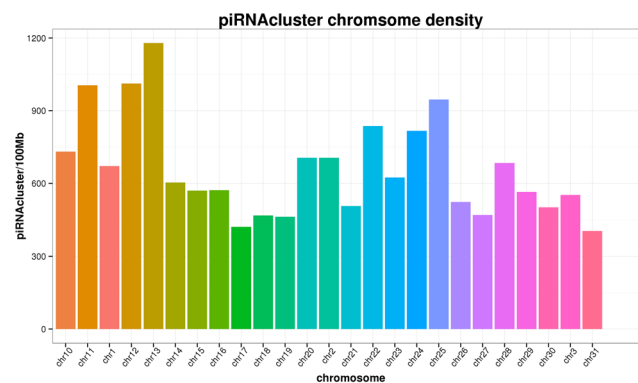


Figure 2. The distribution of piRNA clusters across chromosomes, showing that distributions were not uniform and were not proportional to the length of the chromosome.

containing 2 (*STPG2*; generating genes by uniq_2341752) and meiosis 1-associated protein (*MIAP*; generated by uniq_2350120) spermatogenesis-related genes.

Putative functions of identified piRNA clusters. We identified 7,890 piRNA clusters across all six libraries (Supplementary Table 4). Of these, 199 piRNA clusters were only expressed in the mature testes, and 1148 piRNA clusters were only expressed in the immature testes. The distribution of the piRNA clusters across chromosomes was non-uniform and was not proportional to the length of each chromosome (Fig. 2 and Supplementary Table 5).

Of the piRNA clusters shared by the two sets of samples, DESeq identified 3,016 that were upregulated in the mature testes as compared to the immature testes, and 4,874 that were downregulated in the mature testes as compared to the immature testes (Fig. 3). Several genes neighboring piRNAs clusters were significantly upregulated in the mature horses as compared to the immature horses, including spermatogenesis-associated 6 (*SPATA6*), meiotic double-stranded break formation protein 1 (*MEI1*), histone cluster 1, H2BA family (*HIST1H2BA*), testis-specific serine kinase 1B (*TSSK1B*), and centrosome- and basal body-related protein (*ALMS1*). Our qPCR analysis validated the differential expression of the eight randomly-selected piRNAs in mature and immature horses (Supplementary Table 6), indicating that our DESeq analysis was reliable.

Functional annotation of putative piRNA-generating genes. We identified 7,776 GO terms related to the all candidate piRNA-generating genes; 469 of these were significantly enriched (corrected $P < 0.05$; Fig. 4; Supplementary Table 7). We identified 273 significantly enriched GO terms related to biological processes, including metabolic processes (774 genes), cellular metabolic processes (608 genes), and organic substance metabolic processes (633 genes); 103 significantly enriched GO terms related to cellular components, including intracellular (863 genes), intracellular parts (796 genes), and organelles (774 genes); and 93 significantly enriched GO terms related to molecular function, including binding (1063 genes), protein binding (869 genes), and catalytic activity (435 genes). We identified 11 KEGG pathways significantly enriched in piRNA-generating genes ($P < 0.05$): focal adhesion, phosphatidylinositol signaling system, progesterone-mediated oocyte maturation, glutamatergic

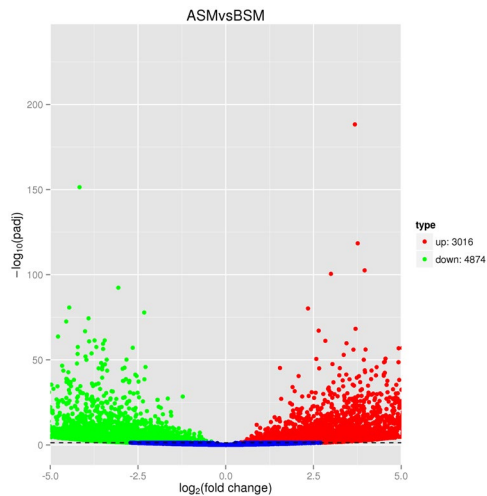


Figure 3. Significantly up- and downregulated piRNA clusters in mature testes as compared to immature testes.

synapse, glycerolipid metabolism, dorsoventral axis formation, endocytosis, alanine, aspartate and glutamate metabolism, oocyte meiosis, adherens junction, and glycerophospholipid metabolism.

Discussion

Consistent with previous results, the piRNAs that we identified in the Mongolian horse had a bias for U at the 5' end, and in some cases the 10th base had a bias for A³²; these bases are characteristic of miRNA sequences^{32,33}. It is possible that these sequence features might be related to the ping-pong model of piRNA generation³³. Piwi proteins are germline-specific argonaute proteins³³ that play vital roles in piRNA biogenesis in *Drosophila* and Zebrafish^{33–35}. Previous studies have identified five argonaute proteins in *Drosophila*: Ago1, Ago2, Ago3, Piwi, and Aubergine^{36,37}. The piRNA sequences associated with Piwi and Aub are similar to the anti-sense strands of retrotransposons, while the piRNA sequences associated with Ago3 are similar to the sense strands³⁸. Interestingly, piRNA sequences associated with Ago3 (sense strands) typically have an A at position 10, and piRNA sequences associated with Piwi and Aub (anti-sense strands) typically have a 5' U³⁹. Previous studies of cleavage activity in Piwi proteins suggest that this structure is typical of piRNAs⁴⁰.

Our results indicated that the putative piRNAs we detected have similar features to previously described piRNAs, and are therefore likely to be true piRNAs. This suggested that primary piRNAs are tremendously heterogeneous group, characterized by a preference for a 5' U, although the mechanistic explanation for this preference is unclear^{39,40}. In PIWI proteins, the middle³⁹ domain provides a structural basis for the enrichment of the 5' U in miRNA sequences³⁹. Indeed, MID domain structures in *Arabidopsis* argonaute proteins are enriched with small RNAs having distinct 5' nucleotide biases for U, A, or cytosine (C)^{39,40}.

We identified 2,236,377 putative piRNAs only expressed in the mature samples, and 2,391,271 putative piRNAs only expressed in the immature samples. This suggested that piRNAs might play important role in spermatogenesis, especially at the early stages. We identified eight novel piRNAs with >500 reads in the immature testes, and eight novel piRNAs with >500 reads in the mature testes. Unfortunately, most of these novel piRNAs have not yet been annotated. This hampers our understanding of the function of these piRNAs in testicular development and spermatogenesis.

However, previous studies of piRNAs in humans indicate that piRNAs are likely to play a significant role in spermatogenesis^{32,34}. For example, a previous study identified 20,121 piRNAs in normal human testis; of the piRNAs with >1000 reads, 12 mapped uniquely within testis developmental related gene 1 (TDRG1), while several others mapped in a sense orientation to an intron of cytochrome P450 family 19 subfamily A member 1 (CYP19A1)^{41,42}. Both of these genes are associated with spermatogenesis: TDRG1 is a developmentally regulated testicular-specific gene, and CYP19A1 catalyzes androgens into estrogens⁴³. Indeed, Flannigan⁴⁴ identified significantly more piRNAs in the testicular tissues of normal men than in men with non-obstructive azoospermia (having Sertoli cells only), indicating that piRNAs are more abundant in sperm and spermatids, and thus probably play an important regulatory role in spermatogonia.

Here, although 4,936,717 putative piRNAs were screened, only 30,639 piRNA generating genes were predicted. Of predicted piRNA-generating genes, 6,969 were known protein coding genes. Two of these predicted piRNA-generating genes encode argonaute proteins: argonaute 1, RISC catalytic component (*AGO1*) and argonaute 3, RISC catalytic component (*AGO3*). Both *AGO1* and *AGO3* were significantly upregulated in the mature horse testes as compared to the immature horse testes, indicating that these piRNAs might be involved in spermatogenesis. Indeed, the *AGO3* protein is required for RNA-mediated gene silencing (RNAi): it binds to short RNAs, such as miRNAs or siRNAs, to repress the translation of complementary mRNAs⁴⁵. The transcriptional gene silencing (TGS) of promoter regions complementary to bound short antigenic RNAs (agrRNAs) also requires *AGO3*. *AGO1* is an important paralog of *AGO3*^{45,46–48}. As typical argonaute proteins, *AGO1* and *AGO2* are also involved in miRNA-mediated gene management and siRNA-mediated mRNA degradation⁴⁵. In contrast,

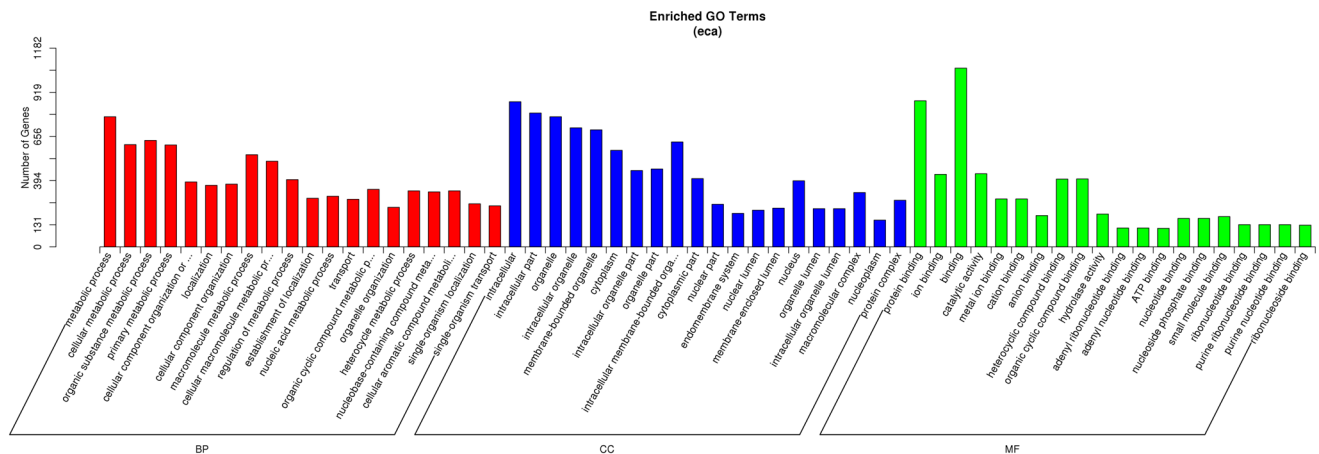


Figure 4. Gene ontology (GO) of the piRNA-generating genes. BP, biological process; CC, cellular component; MF, molecular function.

AGO3, similar to the PIWI proteins and Aubergine, is associated with piRNAs, which primarily originate from transposon-rich clusters and play a pivotal role in transposon silencing⁴⁵.

Two of the candidate generating genes encoded proteins in the PIWI clade: piwi-like RNA-mediated gene silencing 1 (*PIWIL1*), and piwi-like RNA-mediated gene silencing 2 (*PIWIL2*). Both *PIWIL1* and *PIWIL2* were upregulated in the mature testes as compared to the immature testes. *PIWIL1* is expressed in the testis, oocytes, and early embryos of cattle⁴⁹, and *PIWIL1* has been shown to inhibit the activity and movement of transposons during spermatogenesis by forming a complex with piRNAs⁵⁰. *PIWIL1* is critical for the maintenance of germline integrity⁵⁰.

piRNAs may also be associated with other meiotic processes, such as the regulation of translation^{51,52}. During piRNA biosynthesis, *PIWIL2* plays a key role in the ping-pong amplification cycle, acting as a slicer-competent piRNA endoribonuclease that cleaves primary piRNAs; cleaved piRNAs are then loaded onto the slicer-incompetent *PIWIL4*⁵³. *PIWIL2* is expressed in male as well as female germ cells⁵⁴, indicating that this gene may function during oogenesis as well as spermatogenesis^{15,49,52,55–58}.

We also identified several Tudor domain-containing genes (*TDRDs*) as candidate generating genes. Five *TDRD* genes (*TDRD1*, *TDRD5*, *TDRD9*, *TDRD12*, and *TDRD15*) were upregulated in the mature testes, as compared to the immature testes. During piRNA metabolism, *TDRD* proteins are associated with piRNA biogenesis⁵⁹. In addition, *TDRD* proteins are critical for spermatogenesis because, like *PIWIL1*, they inhibit the activity and movement of transposons during spermatogenesis by forming piRNA and Piwi protein complexes, and thus maintain germline integrity⁵⁹. The piRNA/PIWI multiprotein complexes are involved in secondary piRNA metabolic processes⁶⁰, acting via the PIWI-EXD1-Tdr12 (PET) complex during the *PIWIL4* piRNA loading that is triggered by *PIWIL2* slicing⁵⁸.

Some piRNA-generating genes with spermatogenesis were detected, such as zinc finger protein genes, microtubule associated protein genes, spermatogenesis-associated protein gene, and sperm-associated antigens. These genes had a strong relationship with spermatogenesis^{61–64}. Their specific functions should be determined in future research.

Therefore, all candidate piRNA-generating genes identified here were associated with the regulation of piRNA generation. Primary piRNA transcripts are generated from the transposon regulatory regions of heterochromatin³³. These primary piRNA transcripts, associated with both Piwi proteins and Aub, are antisense and complement the transposon transcripts³³. Piwi and Aub cleave the target transposon transcripts at 10–11 nt from the 5' end of the antisense piRNA, generating Ago3-associated sense piRNAs³⁸. Ago3 recognizes the piRNA cluster transcripts, and generates more Piwi/Aub-associated antisense strand piRNAs^{38,65}. The regulation of piRNAs and their generating genes in the Mongolian horse is a target of our future research.

Of the 7,890 piRNA clusters we identified, 3,016 were upregulated in mature testes as compared to immature testes, and 4,874 were downregulated in mature testes as compared to immature testes. piRNA transcriptomes might be strongly adaptive because piRNA clusters incorporate exogenous DNA to provide the substrate for new antisense piRNAs⁶⁶. In propagating transposons, piRNA promote adaptive immunity to selfish DNA⁶⁷. However, mRNAs may be reverse transcribed and reintegrated into existing piRNA clusters, generating pseudogenes that encode piRNAs; these piRNAs may regulate genes⁶⁸. Certain loci may also encode gene-regulating piRNAs on one strand, and genes encoding functional proteins on the other strand⁶⁹.

We were able to annotate some of the genes neighboring the identified piRNA clusters. In particular, *SPATA6* is required in late spermatogenesis for the formation of the link connecting the head and flagellum of the sperm^{66,70}, while *MEI1* is required for the formation of meiotic spindles in female germline cells, and may perform a similar function in males^{71,72}.

The GO terms significantly enriched among the candidate piRNA-generating genes included several that are closely related to spermatogenesis: regulation of microtubule-based processes, cell cycle, cellular biosynthetic processes and RNA biosynthetic processes, cell development, sex differentiation, regulation of cell proliferation, positive regulation of mononuclear cell proliferation, and epigenetic regulation of cell growth.

Our functional assessments of the candidate generating genes of the piRNA clusters indicated that the piRNAs and piRNA clusters identified in the Mongolian horses were strongly associated with spermatogenesis. However, the details of these relationships warrant further study. Indeed, we are currently conducting functional verification studies of several piRNAs and their generating genes.

Data Availability

All raw and processed non-coding RNA profiles have been submitted to the NCBI as a GEO dataset (GEO: GSE100852).

References

- Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207 (2006).
- Girard, A. S. R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006).
- Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes & Development* **20**, 1709–1714 (2006).
- Lau, N. C. *et al.* Characterization of the piRNA complex from rat testes. *Science* **313**, 363–367 (2006).
- Rosenkranz, D. Z. H. proTRAC - a software for probabilistic piRNA cluster detection, visualization and analysis. *Bmc Bioinformatics* **13**, 5 (2012).
- David, R. piRNA cluster database: a web resource for piRNA producing loci. *Nucleic Acids Res* **44**, D223–230 (2016).
- Kim, V. N. Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes & Development* **20**, 1993–1997, <https://doi.org/10.1101/gad.1456106> (2006).
- Watanabe, T. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes & Development* **20**, 1732–1743, <https://doi.org/10.1101/gad.1425706> (2006).
- Gou, L. T. *et al.* Ubiquitination-Deficient Mutations in Human Piwi Cause Male Infertility by Impairing Histone-to-Protamine Exchange during Spermiogenesis. *Cell* **169**, 1090–1104 (2017).
- Gou, L. T. *et al.* Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res* **24**, 680–700 (2014).
- Deng, W. Miwi, a Murine Homolog of piwi, Encodes a Cytoplasmic Protein Essential for Spermatogenesis. *Dev Cell* **2**, 819–830 (2002).
- Kotaja, N. The chromatoid body: a germ-cell-specific RNA-processing centre. *Nat Rev Mol Cell Biol* **8**, 85–90 (2007).
- Onohara, Y. F. T., Yasukochi, T., Himeno, M. & Yokota, S. Localization of mouse vasa homolog protein in chromatoid body and related nuage structures of mammalian spermatogenic cells during spermatogenesis. *Histochem Cell Biol* **133**, 627–639 (2010).
- Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**, 246–258 (2011).
- Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. *Science* **316**, 744–747 (2007).
- Aravin, A. A. & Hannon, G. J. Small RNA Silencing Pathways in Germ and Stem Cells. *Cold Spring Harbor Symposia on Quantitative Biology* **73**, 283–290, <https://doi.org/10.1101/sqb.2008.73.058> (2009).
- Bjørnstad, G., Nilsen, N. Ø. & Røed, K. H. Genetic relationship between Mongolian and Norwegian horses? *Anim Genet* **34**, 55–58 (2003).
- Huang, J. Z. Y. *et al.* Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype. *Sci Rep* **4**, 4958 (2014).
- Kefena, E. M. *et al.* Discordances between morphological systematics and molecular taxonomy in the stem line of equids: A review of the case of taxonomy of genus *Equus*. *Livestock Science* **143**, 105–115 (2012).
- Johnson, L., Varner, D. D. & Jr, T. D. Effect of age and season on the establishment of spermatogenesis in the horse. *Journal of Reproduction & Fertility Supplement* **44**, 87 (1991).
- Naden, J., Amann, R. P. & Squires, E. L. Testicular growth, hormone concentrations, seminal characteristics and sexual behaviour in stallions. *Journal of Reproduction & Fertility* **88**, 167 (1990).
- Sai Lakshmi, S. & Agrawal, S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Research* **36**, D173–D177, <https://doi.org/10.1093/nar/gkm696> (2008).
- Zhang, Y., Wang, X. & Kang, L. A k-mer scheme to predict piRNAs and characterize locust piRNAs. *Bioinformatics* **27**, 771–776, <https://doi.org/10.1093/bioinformatics/btr016> (2011).
- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25 (2009).
- Young, M. D., Wakefield M. J. & Smyth, G. K., Oshlack, g. gseq: Gene Ontology testing for RNA-seq datasets. *Research Gate*, 1–21 (2010).
- Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res* **36** (2008).
- Anders, S. Differential expression analysis for sequence count data. *Genome Biology* **11**, R106 (2010).
- Anders, S. H. W. Differential expression of RNA-Seq data at the gene level. *EMBL* (2013).
- Wang, L., Feng, Z., Wang, X., Wang, X. & Zhang, X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138 (2010).
- Xiang, M. *et al.* U6 is not a suitable endogenous control for the quantification of circulating microRNAs. *Biochem Biophys Res Commun* **454**, 210–214 (2014).
- Livak, K. J. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} Method. *Methods* **25**, 402–408 (2001).
- Aravin, A. A. *et al.* The Small RNA Profile during *Drosophila melanogaster* Development. *Developmental Cell* **5**, 337–350 (2003).
- Brennecke, J. *et al.* Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
- Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590 (2007).
- Houwing, S. *et al.* A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**, 69–82 (2007).
- Mi Shijun, C. T. *et al.* Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5' Terminal Nucleotide. *Cell* **133**, 116–127 (2005).
- Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev* **16**, 2733–2742 (2002).
- Klattenhoff, C. & Theurkauf, W. Biogenesis and germline functions of piRNAs. *Development* **135**, 3–9, <https://doi.org/10.1242/dev.006486> (2007).
- Frank, F. S. N. & Nagar, B. Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818–822 (2010).

40. Chu, C. Y. Small RNAs: regulators and guardians of the genome. *Journal of Cellular Physiology* **213**, 412–419 (2010).
41. Lazaros, L. *et al.* The association of aromatase (CYP19) gene variants with sperm concentration and motility. *Asian J Androl* **13**, 292–297 (2011).
42. Yang, Q. *et al.* MicroRNA and piRNA Profiles in Normal Human Testis Detected by Next Generation Sequencing. *PLoS ONE* **8**(6), e66809 (2013).
43. Schmid, R. *et al.* The splicing landscape is globally reprogrammed during male meiosis. *Nucleic Acids Res* **41**, 10170–10184 (2013).
44. Flannigan, R. *et al.* Large scale miRNA and piRNA sequencing analysis of testis biopsies from fertile an infertile men reveals differences between miRNA and piRNA expression during spermatogenesis cycle. *Fertility and Sterility* **108**, e312 (2017).
45. Meister, G. *et al.* Identification of novel argonaute-associated proteins. *Curr Biol* **15**, 2149–2155 (2005).
46. Janowski, B. A. *et al.* Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nature Structural & Molecular Biology* **13**(9), 787–792 (2006).
47. Wu, L., Fan, J. & Belasco, J. G. Importance of Translation and Nonnucleolytic Ago Proteins for On-Target RNA Interference. *Current Biology* **18**(17), 1327–1332 (2008).
48. Stry, M. V. *et al.* Enhanced Susceptibility of Ago1/3 Double-Null Mice to Influenza A Virus Infection. *Journal of Virology* **86**(8), 4151–4157 (2012).
49. Russell, S. J. *et al.* Identification of PIWIL1 Isoforms and Their Expression in Bovine Testes, Oocytes, and Early Embryos. *Biol Reprod* **94**, 75 (2016).
50. Xu, L. *et al.* Piwil1 mediates meiosis during spermatogenesis in chicken. *Anim Reprod Sci* **166** (2016).
51. Chang, G. *et al.* DNA methylation and NF-Y regulate Piwil1 expression during chicken spermatogenesis. *Anim Reprod Sci* **162**, 95–103 (2015).
52. Kuramochi-Miyagawa, S. *et al.* Two mouse piwi-related genes: miwi and mili. *Mech Dev* **108**, 121–133 (2001).
53. Thomson, T. & Lin, H. The Biogenesis and Function of PIWI Proteins and piRNAs: Progress and Prospect. *Annual Review of Cell and Developmental Biology* **25**, 355–376, <https://doi.org/10.1146/annurev.cellbio.24.110707.175327> (2009).
54. Gomes Fernandes, M. *et al.* Human-specific subcellular compartmentalization of P-element induced wimpy testis-like (PIWIL) granules during germ cell development and spermatogenesis. *Hum Reprod* **33**, 258–269 (2018).
55. Lee, J. H., Engel, W. & Nayernia, K. Stem cell protein Piwil2 modulates expression of murine spermatogonial stem cell expressed genes. *Mol Reprod Dev* **73**, 173–179 (2006).
56. Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**, 785–799 (2008).
57. Unhavaithaya, Y. *et al.* H. MILI, a PIWI-interacting RNA-binding protein, is required for germ line stem cell self-renewal and appears to positively regulate translation. *J Biol Chem* **284**, 6507–6519 (2009).
58. Yang, Z. *et al.* PIWI Slicing and EXD1 Drive Biogenesis of Nuclear piRNAs from Cytosolic Targets of the Mouse piRNA Pathway. *Mol Cell* **61**, 138–152 (2016).
59. Sato, K. I. Y., Siomi, H. & Siomi, M. C. Tudor-domain containing proteins act to make the piRNA pathways more robust in *Drosophila*. *Fly (Austin)* **9**, 86–90 (2015).
60. Pandey, R. R. *et al.* Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice. *Proc Natl Acad Sci USA* **110**, 16492–16497 (2013).
61. Ishizuka, M. *et al.* Abnormal spermatogenesis and male infertility in testicular zinc finger protein Zfp318-knockout mice. *Dev Growth Differ* **58**, 600–608 (2016).
62. Katayama, K. Y. H., Tochigi, Y. & Suzuki, H. The microtubule-associated protein astrin transgenesis rescues spermatogenesis and renal function in hypogonadic (hgn/hgn) rats. *Andrology* **1**, 301–307 (2013).
63. Huo, S., Du, W., Shi, P., Si, Y. & Zhao, S. The role of spermatogenesis-associated protein 6 in testicular germ cell tumors. *Int J Clin Exp Pathol* **8**, 9119–9125 (2015).
64. Wu, H. *et al.* Sperm associated antigen 8 (SPAG8), a novel regulator of activator of CREM in testis during spermatogenesis. *FEBS Lett* **584**, 2807–2815 (2010).
65. Kirino, Y. *et al.* Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nature Cell Biology* **11**, 652–658, <https://doi.org/10.1038/ncb1872> (2009).
66. Kawaoka, S. *et al.* A role for transcription from a piRNA cluster in de novo piRNA production. *Rna* **18**, 265–273, <https://doi.org/10.1261/rna.029777.111> (2011).
67. Vagin, V. V. A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. *Science* **313**, 320–324, <https://doi.org/10.1126/science.1129333> (2006).
68. Pantano, L. *et al.* The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. *Rna* **21**, 1085–1095, <https://doi.org/10.1261/rna.046482.114> (2015).
69. Antoniewski, C., Gebert, D., Ketting, R. F., Zischler, H. & Rosenkranz, D. piRNAs from Pig Testis Provide Evidence for a Conserved Role of the Piwi Pathway in Post-Transcriptional Gene Regulation in Mammals. *Plos One* **10**, e0124860, <https://doi.org/10.1371/journal.pone.0124860> (2015).
70. Kirino, Y. & Mourelatos, Z. Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nature Structural & Molecular Biology* **14**, 347–348, <https://doi.org/10.1038/nsmb1218> (2007).
71. Srayko, M., Buster, D. W., Bazirgan, O. A., McNally, F. J. & Mains, P. E. MEI-1/MEI-2katanin-like microtubule severing activity is required for *Caenorhabditis elegans* meiosis. *Genes Dev* **14**, 1072–1084 (2000).
72. Yang, H. Y., McNally, K. & McNally, F. J. MEI-1/katanin is required for translocation of the meiosis I spindle to the oocyte cortex in *C. elegans*. *Dev Biol* **260**, 245–259 (2003).

Acknowledgements

National Natural Science Foundation of China (31472070, 31360538), People's Republic of China Specific Scientific and Technological Cooperation (2017YFE0108700), the Natural Science Foundation of Inner Mongolia (2015BS0314, 2017ZD06), Horse science engineering technology research center of Inner Mongolia Science and Technology Agency (201603002) Inner Mongolia Key Laboratory Project (20130902) and Special fund for transformation of scientific and technological achievements for breeding of new animal and plant in Inner Mongolia Agricultural university (YZGC2017001) supported this study. Funding bodies had no influence on the design of the study, the collection of the data, the analysis, and the interpretation of the data, or the writing of the manuscript. We would like to thank Dr Liping Jiang and Nei Mongol BioNew Technology Co. Ltd. for providing assistance during the preparation of this manuscript.

Author Contributions

B.L. and X.H. conceived of the study; D.M. and D.B. collected the testicular samples; Y.Z., G.B. and X.Z. performed the RNA extractions; S.S. and L.D. performed the bioinformatics analysis; R.L. and Y.W. performed the qPCR validation; and B.L. wrote the paper. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-41475-9>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019