Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline

Toshiaki Watanabe, Ee-chun Cheng, Mei Zhong, and Haifan Lin

Table of Contents

SUPPLEMENTAL RESULTS

Down-regulation of several spermiogenic genes in *Piwil1*$^{-/-}$ testes is due to the absence of cells highly expressing these genes

Previous studies proposed that mRNAs of some spermiogenic genes (*Ace*, *Fhl5/Act*, *Nucb1/Mtest82*, *Odf1/Rt7* and *Tnp1*) were stabilized by PIWIL1 (Deng and Lin 2002; Vourekas et al. 2012), as they were down-regulated in *Piwil1*$^{-/-}$ testes. In our RNA-seq analysis in late spermatocytes, three (*Fhl5, Nucb1,* and *Tnp1*) of them were decreased, though not significantly (Figure 1F and Supplemental Figure 1E). However, because these mRNAs were expressed initially in the round spermatid stage (Supplemental Figure 1F), the detection of these mRNAs in both *Piwil1*$^{+/-}$ and *Piwil1*$^{-/-}$ preparations was likely due to contamination of round spermatids in FACS-purified late spermatocytes, which constitute ~5% of cells in the isolated late spermatocyte fraction from 23 dpp testes (Supplemental Figure 1G). Since the main phase of the expression of these mRNAs is after the stage of spermiogenic arrest in *Piwil1*$^{-/-}$ mice (Supplemental Figure 1F), the decrease of these mRNAs in *Piwil1*$^{-/-}$ mice may be due to the lack of round spermatids that strongly express these mRNAs as contaminants in *Piwil1*$^{-/-}$ preparation. Additionally, piRNAs were not matched preferentially to the down-regulated mRNAs (Figure 2).

Mapping of piRNAs to mRNAs and lncRNAs and counting of matched piRNAs

Since piRNAs that regulate experimentally validated target mRNAs (*Stambp*, *Tdrd1* and *Prelid1*) were mapped with mismatches to their targets (Supplemental Figure 2A, B), some mismatches between piRNAs and target RNAs are likely tolerated for PIWIL1-piRNA-mediated degradation. We allowed four mismatches for the mapping, because all three of experimentally validated target RNAs matched to piRNAs under this condition (Supplemental Figure 2A, B)

For efficient cleavage of target RNAs by PIWIL1 slicer activity *in vitro*, base pairing of nucleotides 2-22 of piRNAs has been reported to be required (Reuter et al. 2011). Preferential mapping of piRNAs to the up-regulated mRNAs was also observed when using the nucleotides 2-22 perfect match condition (Supplemental Figure 2A). However, this condition seems to be unnecessary *in vivo*, as piRNAs were not mapped to one (*Prelid1*) of the validated target mRNAs under this condition (Supplemental Figure 2B). Therefore, we used full-length piRNAs for mapping.

It has been suggested that miRNAs with multiple target RNAs down-regulate each target RNA to a lesser extent when compared with miRNAs with fewer targets (Arvey et al. 2010). This dilution effect can be explained by sequestration of miRNAs by their targets. While we attempted to consider this dilution effect in counting the number of piRNA matches by

regarding a match by a piRNA with multiple target RNAs as having less effect on each target (Supplemental Figure 2C), this did not appreciably improve the correlation between piRNA mapping and changes in RNA transcript level (Supplemental Figures 2D, E). Although several explanations can be offered for this, PIWIL1-piRNAs might degrade target RNAs catalytically rather than stoichiometrically. Therefore, we did not incorporate a dilution effect in subsequent analyses.

Slicer activity contributes to target RNA degradation

LINE1 retrotransposons are repressed in spermatcoytes and round spermatids by PIWIL1 in a slicer activity dependent manner (Reuter et al. 2011). To examine the dependence of piRNA-induced mRNA and lncRNA degradation on the slicer activity of PIWIL1, we examined the expression of several PIWIL1-regulated mRNAs and lncRNAs in *Piwil1* slicer mutant (*Piwil1*$^{-/ADH}$) mice (Reuter et al. 2011). In late spermatocytes from *Piwil1*$^{-/ADH}$ mice, all RNAs examined, including *Tdrd1* mRNA, were increased compared with the heterozygous control (*Piwil1*$^{+/-}$, Figure 5D), suggesting that slicer activity contributes to target RNA degradation. In addition, PIWIL1 slicer activity may partly contribute to production of PIWIL1-associated piRNAs via the secondary biogenesis (PIWIL1:PIWIL1 homotipic

ping-pong), as some L1-derived PIWIL1-associated piRNAs showed characteristics of secondary piRNAs (10 nt overlap and enrichment 10^{th} A) in *Piwil2* CKO (*Piwil2*$^{fl/fl}$, *Stra8-Cre*), as well as control (*Piwil2*$^{+/fl}$, *Stra8-Cre*) (Supplemental Figures 6A, B) (Brennecke et al. 2007; Gunawardane et al. 2007; Reuter et al. 2011; Di Giacomo et al. 2013).

Target RNAs are not enriched in RNAs immunoprecipitated with PIWIL1

In *Arabidopsis*, it has been shown that slicer-defective Argonaute proteins stabilize AGO-miRNA-target RNA ternary complexes, and immunoprecipitation of slicer-defective Argonaute proteins enriches target RNAs more efficiently than that of wild-type Argonaute proteins (Carbonell et al. 2012). Therefore, we attempted to pull-down PIWIL1 target RNAs from *Piwil1* slicer mutant mice. However, the potential target RNAs that were up-regulated in *Piwil1*$^{-/-}$ mice were not particularly enriched in the immunoprecipitated RNAs, even though enrichment of piRNA intermediates was observed (Supplemental Figures 3A-D). Similar results were obtained when wildtype PIWIL1 was immunoprecipitated from *Piwil1*$^{+/-}$ mice. Although there are several possible reasons for this no enrichment, PIWIL1 and slicer-defective PIWIL1 may not form stable complexes with target RNAs. In addition, slicer-independent mechanisms (e.g., recruitment of RNA degradation machinery by PIWIL1) may cause degradation of target

RNAs associated with slicer-defective PIWIL1, since the extent of target RNA increase in *Piwil1*$^{-/ADH}$ mice was milder than that in *Piwil1*$^{-/-}$ mice (Figure 5D).

SUPPLEMENTAL METHODS

Flow cytometry and sorting of spermatogenic cells

To obtain late spermatocytes from *Piwil1*$^{+/-}$ vs. *Piwil1*$^{-/-}$ mice, testes were harvested. To prepare single-cell suspensions, decapsulated testes were treated with a digestion solution (0.5 mg/ml collagenase type IV, 0.05% trypsin-EDTA, and 25 μg/ml DNaseI in DMEM) at 35^{o}C for 10 min. After the addition of 10% volume of fetal bovine serum (FBS) to stop the enzymatic reaction, cell suspensions were filtered through a 70-μm cell strainer (BD Falcon) and washed once. The cells were then incubated with 10% FBS, 5 μM Vybrant Dyecycle Ruby Stain (Invitrogen), and 25 μg/ml DNaseI in DMEM at 35^{o}C for 15 min. The cells were washed once with FACS solution (25 μg/ml DNaseI and 0.5% FBS in DMEM). The DAPI nuclear dye was used at 1 μg/ml to exclude dead cells. In the *Piwil1* mutant allele, the *Piwil1* ORF is replaced with a GFP in order to express GFP from the mid-pachytene spermatocyte to round spermatid stages under the control of the endogenous *Piwil1* promoter (Deng and Lin 2002). On the basis of *Piwil1*-GFP expression level and DNA content, late spermatocytes (GFP+, 4N) were sorted using BD FACSAria II (Supplemental Figures 1C, K). The purity of FACS-isolated cells was estimated by immunocytochemistry using anti-SCP3 and anti-MVH antibodies (Supplemental Figure 1D). To obtain late spermatocytes from *Mov10l1* CKO and control mice, cells were purified according to a published method with some modifications (Bastos et al. 2005). In brief,

single cell suspensions were incubated with 0.5% FBS, 5 μg/ml Hoechst 33342, 1 μg/ml PI, and 25 μg/ml DNaseI in DMEM at 35°C for 20 min. After washing, the cells were sorted based on DNA content and the efficiency of Hoechst 33342 staining (Figure 6B), which varies with cell type. The isolated cells were stained using anti-SCP3 and anti-MVH antibodies as well as DAPI.

Generation of *Stambp-ps1* mutant mice

Gene trap ES cells (IST11216A6 and IST12340E11) were obtained from Texas A&M Institute for Genomic Medicine. The ES cells were from C57BL/6 and contribute to the black coat color. We accordingly injected the ES cells into blastocysts from C57BL/6 albino mice. Chimeric mice were crossed with C57BL/6 albino mice. Agouti or black mice were selected and genotyped to obtain heterozygous mice. The heterozygous mice were crossed with each other to produce homozygous mice for analysis. Genotyping was performed by PCR using genomic DNA from tail tips. To isolate DNA, the tail tip (2 mm) was treated with 100 μl of alkaline solution [0.2 mM EDTA (pH 8) and 25 mM NaOH] at 100°C for 15 min in a 1.5-ml tube. Subsequently, an equal volume of neutralization solution (40 mM Tris-HCl) was added. After mixing well using a Vortex, the tube was centrifuged for 1 min at 12,000 × *g*, and 1 μl of the supernatant was added to the 15-μl reaction mixture for PCR. The primer sequences used are listed below.

Generation of mCherry Transgenic mice

A DNA fragment of *mCherry* was amplified by PCR from the plasmid H2B-*mCherry* (Addgene plasmid 20972) and digested with MfeI and XhoI. Digested DNA was ligated to EcoRI and XhoI sites in the pCAGEN plasmid (Addgene plasmid 11160) to construct pCAGEN-*mCherry*. A DNA fragment of the PEST sequence was amplified from the plasmid pd2EGFP-1 (Clontech), digested with SalI and NotI, and ligated to XhoI and NotI sites in the pCAGEN-*mCherry* to construct pCAGEN-*mCherry*-PEST. A DNA fragment of a polyadenylation site-containing sequence was amplified from the psiCheck2 plasmid (Promega), digested with NotI and HindIII, and ligated to NotI and HindIII sites in the pCAGEN-*mCherry*-PEST to construct pCAGEN-*mCherry*-PEST-polyA. A synthetic DNA fragment containing restriction sites for insertion of retrotransposon sequences and flanking *Loxp* sites was digested with NotI and SalI, and ligated to the NotI and XhoI sites in the pCAGEN-*mCherry*-PEST-polyA to construct pCAGEN-*mCherry*-PEST-*Loxp*-*LoxP*-polyA. DNA fragments of B1 (B1_Mm registered in Repbase; position 147-1) and B2 (B2 registered in Repbase; position 1-159) were synthesized and ligated to EcoRI and XhoI sites in the pCAGEN-*mCherry*-PEST-*Loxp*-*Loxp*-polyA to construct pCAGEN-*mCherry*-PEST-*Loxp*-B1-*LoxP*-polyA and pCAGEN-*mCherry*-PEST-*Loxp*-B2-*LoxP*-polyA. The plasmid DNA was then digested with

HindIII, SpeI and BamHI to remove the vector backbone, and the purified DNA was injected into the pronuclei of 1-cell embryos from B6;SJL F2 mice. Transgenic mice were bred with CD-1 mice. To select a mouse line with a single copy of the transgene, genomic DNA was separated on 0.8% agarose gel after digestion by restriction enzymes, then transferred to the HybondN+ membrane (GE), and hybridized with random primed probes. Genotyping was performed by PCR using a set of primers spanning *Loxp* sites. The oligo DNA sequences used are listed below.

Insertion of polyadenylation signal sequence into 3´ UTR of *Prelid1* using CRISPR technique

An annealed oligo DNA homologous to a part of *Prelid1* 3´ UTR sequence was ligated to BbsI site of the pX330 plasmid (Addgene plasmid 42230). A DNA fragment of Cas9 mRNA with T7 promoter sequence and a DNA fragment of chimeric guide RNA with T7 promoter were amplified by PCR from the pX330. Cas9 mRNA and chimeric guide RNA were transcribed from the PCR products using mMESSAGE mMACHINE T7 ULTRA Transcription kit (Ambion) and MEGAshortscript T7 kit (Ambion), respectively. The reaction was performed according to the manufacture's protocol except for 16-hr incubation of transcription reaction of the chimeric guide RNA. Transcribed RNAs were purified using MEGAclear kit (Ambion)

according to the manufacture's protocol, and then eluted using water. An oligo DNA for homologous recombination (50 nt homologous sequence − polyadenylation signal sequence − 50 nt homologous sequence), Cas9 mRNA, and chimeric guide RNA were mixed in injection buffer [10 mM Tris-HCl (PH 7.4), 10mM NaCl and 0.25 mM EDTA (PH 8.0)] at the final concentration of 30 ng/ul (homologous recombination oligo), 30 ng/ul (Cas9 mRNA), and 15 ng/ul (chimeric guide RNA) and then injected into the cytoplasm of 1-cell embryos of B6;SJL F2 mice. Out of 7 mice born, one mouse had the expected mutation. The mouse was bred with CD-1 mice. Genotyping was performed by PCR using a set of primers spanning the insertion site. The oligo DNA sequences used are listed below.

3´ RACE

Total RNA (1 μg) was isolated using TRIZOL (Invitrogen) from FACS-isolated late spermatocytes. For 3´ RACE, the total RNA was reverse-transcribed as previously described (Scotto-Lavino et al. 2006). Then, two rounds of PCR were performed to specifically amplify the 3´ end of *Prelid1* mRNA. The amplified PCR product was sequenced and confirmed. The oligo DNA sequences used are listed below.

Small RNA Northern blot analysis

Total RNAs from 21 dpp testes (20 μg) were loaded onto a 15% denaturing polyacrylamide gel and electroblotted onto a Hybond N+ membrane (GE). Oligo DNA probes were labeled with T4 polynucleotide kinase in the presence of [γ-32P]ATP. Hybridization was performed in 0.2 M $NaHPO_4$ (pH 7.2), 1 mM EDTA (pH 8), 1% BSA and 7% SDS at 50°C overnight. Membranes were washed four times with a 2× SSC/0.1% SDS solution at 50°C for 10 min each. The oligo DNA sequences used are listed below.

Histology and immunostaining analyses

Freshly dissected testes were fixed in 4% paraformaldehyde/PBS at 4°C overnight. Fixed testes were embedded in paraffin and cut into 7-μm thick sections and stained with H&E or PAS/hematoxylin. For immunostaining, fixed testes were embedded in the O.C.T. compound (Sakura) and sectioned. FACS-isolated cells were fixed in 2% paraformaldehyde/PBS on ice for 30 min and then attached to glass slides using cytospin. After washing, the slides were incubated with 1.5% normal goat serum and 0.1% Tween-20 in PBS at room temperature for 30 min. Then, the slides were incubated with the following primary antibodies diluted as indicated at room temperature for 30 min: anti-L1 ORF1 (a gift from Sandra Martin), SCP3 (Santa Cruz, sc-74569) 1/500, MVH (Abcam, ab13840) 1/1000, and STAMBP (sigma, SAB4200144) 1/250. After washing, the slides were incubated with secondary antibodies at room temperature for 30

min.

Western Blot analysis

Late spermatocytes and round spermatids were isolated by FACS using Hoechst 33342 from 26 dpp *Stambp-ps1*$^{+/Gt1}$ and *Stambp-ps1*$^{Gt1/Gt1}$ mice. One million cells (round spermatids) or 200,000 cells (late spermatocytes) were lysed in 200 ul of Laemmli sample buffer, and 10 ul was loaded per lane. After electrophoresis and transfer onto nitrocellulose membrane, the membrane was blocked with 2% skim milk in TBST. The membrane was incubated with the following primary antibodies diluted as indicated at 4^{o}C overnight: anti-L1 ORF1 (a gift from Sandra Martin), GAPDH (sigma, G9545) 1/10,000 and STAMBP (sigma, SAB4200144) 1/250. After washing, the membrane was incubated with the goat anti-rabbit IgG HRP conjugated antibody at room temperature for 30 min.

Northern blot analysis of L1

Total RNAs were isolated from 23-dpp testes of *Piwil1*$^{+/-}$ and *Piwil1*$^{-/-}$ mice, and 10 μg of the total RNAs were loaded onto 1% denaturing agarose gel. The L1 DNA fragment was PCR amplified from genomic DNA and cloned into a PCR2.1-TOPO vector (Invitrogen). PCR primers used for the amplification of the L1 DNA fragment are listed below. The probe for

Northern blotting was synthesized by *in vitro* transcription in the presence of [α-32P]UTP. Hybridization was performed in 0.2 M $NaHPO_4$ (pH 7.2), 1 mM EDTA (pH 8), 1% BSA and 7% SDS at 65°C overnight. A membrane was washed once with 2x SSC/0.1% SDS solution and three times with 0.2x SSC/0.1% SDS solution at 65 °C for 10 min each.

Quantitative real-time PCR analysis

Late spermatocytes were isolated by FACS using either *Piwil1*-EGFP or Hoechst 33342. Total RNA was isolated using TRIZOL (Invitrogen). Total RNAs was treated with DNaseI and reverse-transcribed with the SuperScript III (Invitrogen) or the High-Capacity cDNA Reverse Transcription Kits (ABI) using random primers. q-PCR was performed using the SYBR Premix Ex Taq, Tli RNase H plus (Takara Clontech) or the iQ SYBR Green Supermix (Biorad). The PCR primers used are listed below.

Small RNA-seq analysis of *Stambp* pseudogene mutant mice

Total RNAs were isolated from 23 dpp testes from *Stambp-ps1*$^{+/Gt1}$ and *Stambp-ps1*$^{Gt1/Gt1}$ mice, and 20 μg of the total RNAs were loaded onto 15% denaturing polyacrylamide gels. Small RNAs ranging in size between 15 nt and 40 nt were gel purified and used for library construction using the TruSeq Small RNA Sample Prep kit (Illumina). Small RNA libraries

were sequenced for 50 cycles using the Illumina HiSeq 2000. The extracted small RNAs were annotated as previously described (Watanabe et al. 2011), and miRNAs and degradation products of abundant noncoding RNAs (rRNA, tRNA, snRNA, snoRNA, scRNA, srpRNA and RNA identified by RepeatMasker) were removed before analysis. The remaining small RNAs were mapped to the genome, and small RNAs that were perfectly and uniquely mapped to the *Stambp-ps1* piRNA cluster were analyzed.

Immunoprecipitation of RNA associated PIWIL1

Late spermatocytes were isolated by FACS using Hoechst 33342 from *Piwil1*$^{-/ADH}$, $^{+/-}$, and $^{-/-}$ testes. Approximately one million late spermatocytes were used for each experiment. Cells were lysed in 250 ul of ice-cold lysis buffer [50 mM Tris-Cl (PH 7.4), 2.5 mM MgCl2, 100 mM KCl, 0.1% Nonidet P-40, 1 x Complete proteinase inhibitor (Roche), 50 units/ml SUPERaseIn (Life technologies). The sample was then passed through 26-gauge attached to a sterile plastic syringe 20 times on ice, and rotated for 30 min at 4^{o}C. Cell debris was removed by centrifugation at 21,000 × *g* for 30 min at 4^{o}C. As the input RNA, 10 ul of the lysate of each genotype was taken. In parallel with the lysate preparation, 5 ul of anti-PIWIL1 antibody (Wako) was mixed with 30 ul of Dynabeads Protein G (Life technologies), and rotated for 1 hr at 4^{o}C. After washing, the beads was mixed with the lysate, and rotated for 3 hr at 4 ^{o}C. Wash was performed using

washing buffer [50 mM Tris-Cl (PH 7.4), 2.5 mM MgCl2, 100 mM KCl, 0.1% Nonidet P-40] six times for 10 min each at 4 °C. RNA was purified from the beads using TRIZOL.

RNA-seq libraries were prepared from immunoprecipitated RNAs and input RNAs from three genotype samples (a total of 6 libraries) using ScriptSeq Complete Gold Kit-Low input (Epicentre) according to the manufacture's protocol. AMPure XP beads were used for purification of cDNA and PCR product, and therefore cDNA or PCR products derived from small RNAs (< ~100 nt in length) were removed during these purification steps. The six libraries were sequenced using HiSeq2000.

Enrichment score of each mRNAs and lncRNAs was calculated by dividing RPKM (PIWIL1 IP) by RPKM (input). RPKM values were calculated as described below.

Calculation of reads per kilobase of transcript per million mapped reads (RPKM) values of mRNAs

Read sequences were mapped to the genome using TopHat software (http://tophat.cbcb.umd.edu/), and reads that were uniquely and perfectly mapped to the genome were used for subsequent analysis. For directional RNA-seq, the unique reads extracted after mapping with TopHat were further mapped to mRNAs of Ensembl genes using Bowtie software (http://bowtie-bio.sourceforge.net/index.shtml). RPKM value was calculated for each gene

using (1) A = the number of reads perfectly mapped to either of variants in the sense orientation, (2) B = the average length of the variants, and (3) C = the total number of unique reads extracted after mapping with Tophat (RPKM = $A(10^3/B)(10^6/C)$). For the analysis of PIWIL1 IP and input libraries, sequences mapped to abundant noncoding RNAs (rRNA, tRNA, snRNA, snoRNA, scRNA, srpRNA and RNA) were removed before mapping to the mRNAs. For unstranded RNA-seq, RPKM values were calculated using Cufflinks software (http://bowtie-bio.sourceforge.net/index.shtml). To calculate the RPKM value of each retrotransposon sequence, reads perfectly mapped to the genome were extracted. These were further mapped to retrotransposon sequences downloaded from Repbase (http://girinst.org/repbase/) using Bowtie software allowing up to three mismatches. The RPKM value of each retrotransposon sequence was calculated using (1) A = the number of reads mapped to the retrotransposon sequence in the sense orientation, (2) B = the length of the retrotransposon sequence, and (3) C = the total number of genome-matching reads (RPKM = $A(10^3/B)(10^6/C)$). If one read was mapped to multiple (n) retrotransposon sequences, we considered this read as 1/n read. Up-regulated genes, lncRNAs and retrotransposons were identified by applying combined thresholds based on fold change (>2), expression level (RPKM > 0.5) and one-tailed unpaired Student's t-test ($P < 0.025$).

Identification of lncRNAs

Transcripts were assembled by Cufflinks v1.2.1 (http://cufflinks.cbcb.umd.edu/) using both *Piwil1*$^{+/-}$ and *Piwil1*$^{-/-}$ late spermatocyte RNA-seq reads. Mitochondrial sequence and abundant noncoding RNA sequences (rRNA, tRNA, snRNA, snoRNA, scRNA, srpRNA and RNA identified by Repeatmasker; http://www.repeatmasker.org) were masked using the -M option. The default settings were used and a total of 73149 transcripts were identified. The lncRNA transcripts were identified using the following criteria: (1) >200 nt in length (67665 transcripts), (2) non-overlapping with known protein-coding genes registered in Ensembl, RefSeq, and UCSC databases (42133 transcripts), and (3) RPKM > 0.5 in either *Piwil1*$^{+/-}$ or *Piwil1*$^{-/-}$ late spermatocytes (6014 transcripts).

Mapping of PIWIL1-associated piRNAs to mRNAs and lncRNAs

Ensembl protein-coding genes and late spermatocyte lncRNAs with RPKM > 0.5 in either *Piwil1*$^{+/-}$ or *Piwil1*$^{-/-}$ late spermatocytes were extracted for mapping. Before mapping, simple repeat and low complexity sequences in the mRNAs from the extracted genes and lncRNAs were masked using RepeatMasker with ABBlast as the search engine and slow as the speed/sensitivity. PIWIL1-associated small RNA reads (SRR033657) were mapped to the genome, and perfectly matched sequences were extracted. The extracted small RNAs were

annotated as previously described (Watanabe et al. 2011), and miRNAs and degradation products of abundant noncoding RNAs (rRNA, tRNA, snRNA, snoRNA, scRNA, srpRNA and RNA) were removed. The remaining small RNAs were mapped to the masked mRNAs using SeqMap software (http://www-personal.umich.edu/~jianghui/seqmap/) allowing up to four mismatches including two indels. For each gene or lncRNA, the total number of -associated piRNA reads mapped to the mRNA or the lncRNA in the antisense direction was counted. If a gene or a lncRNA had several variants, we considered the expression levels of each variant. For example, if one gene had two variants with an RPKM ratio of 4:1, then the number of piRNA reads mapped to two variants was weighted by 0.8 and 0.2, respectively. For the analyses shown in Figures 2 and 4 and Supplemental Figures 2 and 6, genes and lncRNAs were classified into five groups based on the extent of up-regulation in *Piwil1*$^{-/-}$ mice (<1, 1−1.25, 1.25−1.5, 1.5−2 and >2), and only those genes and lncRNAs that were significantly changed in expression were used for analysis in 1.5-2 and >2 groups.

Annotation of piRNA-complementary sequences within mRNAs

PIWIL1-associated small RNA reads were mapped to the genome, and perfectly matched sequences were extracted. The extracted small RNAs were annotated as previously described (Watanabe et al. 2011), and miRNAs and degradation products of abundant non-coding RNAs

were removed. The remaining small RNAs were mapped to the up-regulated mRNAs (>2-fold, $P < 0.025$, RPKM > 0.5) using SeqMap (http://www-personal.umich.edu/~jianghui/seqmap/) allowing up to four mismatches including two indels. Using the results of piRNA mapping, we identified the piRNA-mapped (≥5 RPM) regions within the mRNAs. Repeat sequences within the up-regulated mRNAs were identified using RepeatMasker. We then annotated every nucleotide position in the piRNA-mapped regions using the RepeatMasker results. The total number of positions annotated was counted for each annotation (SINE, LTR, LINE, other repeats, and non-repeat). If a gene had several variants, we counted the number of positions in the gene by considering the RPKM values of the variants. For example, if one gene had two variants with an RPKM ratio of 3:2, then the numbers of positions in two variants were weighted by 0.6 and 0.4, respectively.

Identification of piRNA clusters

To identify piRNA clusters, PIWIL1-associated small RNA reads (SRR033657) were annotated as described previously (Watanabe et al. 2011), and then miRNAs and degradation products of abundant non-coding RNAs (rRNA, tRNA, snRNA, snoRNA, scRNA, srpRNA and RNA) were removed. The remaining small RNAs were aligned to the genome, and only small RNAs that were aligned perfectly with the genome 1–10 times were considered for the following analysis. The

genome was scanned using a 10-kb window and the windows that contained more than 50 different small RNA clones were extracted. Any overlapping windows that fulfilled these criteria were combined. If the combined region had >10 different unique hit small RNA clones (25–32 nt), the region was considered to be a potential piRNA cluster. The positions of the most 5′ and 3′ small RNAs were considered the boundaries of the cluster. In cases where the boundary of the next cluster was located within 100 kb, the two clusters were considered to be one cluster. If >50% or >50 reads of small RNAs that constituted a cluster ranged in size between 25 and 32 nt, the cluster was considered to be a piRNA cluster. We used the top 100 piRNA clusters in terms of the number of aligned small RNA clones for further analysis. Genomic positions (mm9) of the top 100 piRNA clusters identified are listed below.

	Chr	Position		Chr	Position		Chr	Position
1	chr17	27402612-27515706	35	chr6	128114295-128200139	68	chr13	52188831-52219397
2	chr9	67529376-67628685	36	chr6	87888146-87953080	69	chr1	93433586-93456094
3	chr2	92365341-92458484	37	chr8	94713344-94726035	70	chr5	135628806-135647029
4	chr12	99601881-99701985	38	chr10	18509526-18558071	71	chr1	57403382-57442187
5	chr7	80888229-81021562	39	chr7	132459870-132516635	72	chr8	28399214-28456713
6	chr6	127716473-127807767	40	chr8	95931451-95954134	73	chr3	20344335-20399151
7	chr18	67182558-67265067	41	chr13	50938900-51029500	74	chr13	49872361-49998145
8	chr9	54040514-54145937	42	chr8	112641159-112661107	75	chr1	93296331-93308897
9	chr15	59069004-59149403	43	chr15	83184102-83202097	76	chr1	127071465-127084341
10	chr7	77012651-77108705	44	chr4	57362505-57381866	77	chr9	88923319-88953101
11	chr5	113737316-113812048	45	chr1	94879184-94899700	78	chr6	83493284-83577012
12	chr5	150582667-150639751	46	chr7	60142989-60175899	79	chr9	51648563-51669434
13	chr7	80232796-80313549	47	chr3	34725541-34785437	80	chr10	94129442-94141191
14	chr14	24889861-24970705	48	chr6	85049957-85081877	81	chr8	119722541-119739642
15	chr4	93941874-94006759	49	chr4	123503582-123582520	82	chr7	48267752-48420722
16	chr17	66492536-66594101	50	chr9	43894937-43966633	83	chr5	137387582-137416227

17	chr2	150964536-151255326	51	chr13	53487806-53501738	84	chr5	115563601-115606574
18	chr10	85989596-86138486	52	chr15	72918584-73023316	85	chr6	92108355-92140406
19	chr15	74442235-74541289	53	chr6	128383265-128409096	86	chr8	38150505-38170276
20	chr10	85187680-85333379	54	chr6	85929891-85958783	87	chr5	102252732-102282938
21	chr10	62105668-62174967	55	chr10	127115992-127138039	88	chr8	113527254-113547913
22	chr11	103278602-103319221	56	chr4	135174069-135193769	89	chr3	28749911-28783264
23	chr10	75229930-75361023	57	chr5	144519219-144542449	90	chr11	95622923-95639433
24	chr5	115257807-115313515	58	chr2	127515470-127531858	91	chr1	162898127-162949924
25	chr15	78463892-78514363	59	chr17	22644164-22662375	92	chr7	39222146-39249650
26	chr13	50244635-50476405	60	chr19	37324049-37346823	93	chr9	122682431-122807450
27	chr13	50597079-50769429	61	chr4	41746073-41821503	94	chr4	42088578-42110768
28	chr14	25149840-25267464	62	chr3	124101056-124129278	95	chr13	24990942-25093629
29	chr14	20436446-20474099	63	chr15	79748722-79802861	96	chr7	48528049-48555782
30	chr14	45544070-45586983	64	chr9	88489783-88530257	97	chr2	37418997-37439369
31	chr14	45369951-45417083	65	chr11	107850632-107862347	98	chr10	60891271-60897842
32	chr10	66134748-66180927	66	chr7	30914865-30934310	99	chr11	5424691-5450231
33	chr4	61874481-61907919	67	chr9	3132889-3223368	100	chr11	108279019-108286433
34	chr6	81841557-81870176						

Identification of pseudogenes in piRNA clusters

To identify pseudogenes that regulate functional cognate genes through the piRNA pathway, we first mapped PIWIL1-associated piRNAs (SRR033657) to mRNAs of Ensembl coding genes. We then selected mRNAs with more than 50 complementary PIWIL1-associated piRNAs. Repeat sequences in the selected mRNAs and the top 100 piRNA clusters listed above were masked using RepeatMasker (http://www.repeatmasker.org). The masked mRNA sequences were aligned to the masked piRNA cluster sequences using Blat (http://hgdownload.cse.ucsc.edu/admin/exe/). Alignments with scores >100 were selected, and the aligned regions in the piRNA clusters were considered to be potential pseudogenes. If a potential pseudogene generated >20 reads of piRNAs that were mapped to its functional cognate mRNA in the antisense orientation, we considered it as a

candidate regulating its functional cognate mRNA. We identified seven candidates including *Stambp* pseudogene (*Stambp-ps1*) by the procedures described above. Of the seven, six (*Stambp*, *Eif2C2/Ago2*, *Fam58b*, *Pphln1*, *Spin1*, and *Unc119b*) were processed pseudogenes, and the remaining (*Gm4884*) was likely to have been generated by duplication of the genomic region encompassing the cognate gene.

Frequency of retrotransposon sequence-containing mRNAs and lncRNAs

Retrotransposon sequences in mRNAs of Ensembl protein-coding genes and late spermatocyte lncRNAs were identified using RepeatMasker. The genes and lncRNAs were classified into the following five groups based on the extent of up-regulation in *Piwil1*$^{-/-}$ mice: <1, 1−1.25, 1.25−1.5, 1.5−2 and >2. Only significantly changed genes and lncRNAs were used for analysis in 1.5−2 and >2 groups. For each group, we counted the number of genes and lncRNAs that contained retrotransposon sequences (all retrotransposons, SINE, LINE, LTR, B1, B2, or B3) in their exons. If a gene or a lncRNA had several variants, we counted the gene by considering the RPKM values of the variants. For example, if one gene had two variants with an RPKM ratio of 3:2, then the two variants were considered to be 0.6 and 0.4 genes, respectively. The R prop.test was used to compare the differences between the two groups.

Retrotransposon sequence density in mRNAs and lncRNAs

The mRNA (lncRNA) sequences with 5' and 3' flanking intergenic sequences were obtained from the UCSC Table Browser. Retrotransposon sequences in the downloaded sequences were identified using RepeatMasker (http://www.repeatmasker.org). At every position analyzed, the number of genes (lncRNAs) annotated as retrotransposon, SINE, B1, and B2 was counted. If a gene or a lncRNA had several variants, we considered the expression levels of the variants. For example, if one gene had three variants with an RPKM ratio of 6:3:1, then three variants were considered to be 0.6, 0.3, and 0.1 genes, respectively. Because genes (lncRNAs) vary in the length of 5' UTR, CDS, and 3' UTR (the length of lncRNA), we also counted the total number of genes (lncRNAs) at every position by considering the expression levels of each variant. Then, we calculated retrotransposon density (occupancy) at every position analyzed, and the average density (occupancy) of each 0.1-kb region.

PCR primer and oligonucleotide sequence

Primers used for q-PCR in *Piwil1* KO mice

Gapdh F	TGTGTCCGTCGTGGATCTGA
Gapdh R	CCTGCTTCACCACCTTCTTGA
4931428L18Rik F	TCCATCATGCTAGCTCCTGA
4931428L18Rik R	AGTATGAACTCTGTGGCTGAG
Bcl7c F	ATTGCTAGAAGCTGAGGCCC
Bcl7c R	CATTTGGACAGATGCGCTTG
Arl16 F	CAACCTTACAGACATTGTGGC
Arl16 R	CCATCATGAACAAGAGCGAATG
Efcab1 F	AATTTGGTGGGAGACGTAGC
Efcab1 R	GCCTCGAAATACTCTGTCCA
Exoc4 F	ATCAGTGCTTCTTGGGCAAA
Exoc4 R	CCAAAGGCTGCCCTTATGAA
Hemt1 F	TAGAACCCTGAGGTTCAACG
Hemt1 R	AGCACTTATGGCATGTAAGCA
Klhl11 F	CGAGTACTTCACGCCCTTAC
Klhl11 R	ATAAGAATCTGTCGGCCAGC
Mmachc F	CTGAGGAGTGTTACAGAGAAGT
Mmachc R	AGTAATAAGCAGCACCTGCC
Morc1 F	ACGGGAGATGTGGAACTTTG
Morc1 R	GTGTAGGCTCTGAATGACCG
Pard6g F	CCACATCTCTAACACCGAGG
Pard6f R	AATGGTCTGCTTCTTCTCGTT
Pogk F	TCGGCCGGGAAACATTGAA
Pogk R	GGCTGGTACAAAGAGGTTCC
Slglece F	CTCAGTGCTAACCATCACCC
Slglece R	TGGAGCATATGACACGTTGA
Smyd4 F	AGGCCCAAACCTTAAAGTCC
Smyd4 R	CTCGTTAGACCCTGTGTGAC
Zfa F	CATCTTATCTGTATGGCCGC
Zfa R	AGAGTCGTGAAAGCAACGAT

Odf1/RT7 F	GCACTTTGACTTGATCAGGTTC
Odf1/RT7 R	CTACAAGCTG TACTGCCTTC
Fhl5/ACT F	GTGTGCAGCCTGCACCAA
Fhl5/ACT R	ACTGGCGGTCTTGAAAGCA
Tnp1 F	GTCAAGAGAGGTGGAAGCAAGA
Tnp1 R	ATTCCGAATTTCGTCACGACTG
Primers for L1 DNA amplification for Northern blotting	
L1 Northern F	AAGTTCCCAACATAGAGTCCTGAG
L1 Northern R	AGTGGGCAGAGTATTCTCTGCAG
Primers for *Stambp-ps1* mouse genotyping	
Genotyping_Stambp-ps1_WT1	GTGTGAGATCACATGACTTGCTGA
Genotyping_Stambp-ps1_WT2	GTTTGTCAGATCACTGTTCCTGTG
Genotyping_Stambp-ps1_mutant1	CCCTAGTCAGTTTACTATTGCACT
Genotyping_Stambp-ps1_mutant2	CTTGCAAAATGGCGTTACTTAAGC
Primers for *Stambp-ps2* mouse genotyping	
Genotyping_Stambp-ps2_WT	GTTTGTCAGATCACTGTTCCTGTG
Genotyping_Stambp-ps2_common	GTGTGAGATCACATGACTTGCTGA
Genotyping_Stambp-ps2_mutant	CCAATAAACCCTCTTGCAGTTGC
Primers for the quantification of the piRNA precursor from the *Stambp-ps1* locus	
piRNA-pre-stambp-ps1 F1	GTGTTCAGTGACTGTCATTGCA
pre-piRNA-stambp-ps1 R1	GACAACTGAAGAACTGTGAGAC
pre-piRNA-stambp-ps1 F2	GCTGTACCTGTTGTATGTCCA
pre-piRNA-stambp-ps1 R2	CTGCAGACTCTATGGTAGAAC
pre-piRNA-stambp-ps1 F3	GAGTTAGTTGCTTAGTGTGATAGGTT
pre-piRNA-stambp-ps1 R3	CCTAAGAGCTACAAACCCTAGTCAGT
pre-piRNA-stambp-ps1 F4	TGGTAACGTCTTCGCTCATCAGT
pre-piRNA-stambp-ps1 R4	AGAAGTGGATCACACTCAGTGT
Primers for the quantification of *Stambp* mRNA and pre-mRNA	
Stambp mRNA F	ACCAGAGTCCATCGCAATCGTCTGT
Stambp mRNA R	TATCGAAGGTCTGTGATCGTCACA
Stambp mRNA Ex6 In6 F	CAGTGCCAATACCGCCAAAG
Stambp mRNA R Ex6 In6 R	CAGGGAAGTTACACCCGCTT
Stambp mRNA F Ex7 In7 F	ACCAGAGTCCATCGCAATCGTCTGT

Stambp mRNA R Ex7 In7 R	GCAGGCATCCTCCTACACTC

Oligos used for small RNA Northern Blotting

Stambp pseudogene piRNA	GGGCCTAACTCAGAAAAGCCTCCAATGAGTA
Chr17 piRNA	GCCACATCACCATATTCATCTAATCAA

Primers used for q-PCR in *Piwil1* slicer mutant mice

Tdrd1F1	GGAACTTTGAGATTCTCAGTCTG
Tdrd1R1	TCTCACTGTGACCATCCTGT
Tdrd1F2	CGTGCTGTGCTAAGTACACA
Tdrd1R2	AGAGGCAGGGTTTCGATG
Tdrd1F3	CAGAAGAGATTCATGTATGAGCT
Tdrd1R3	CCTCATAGTCTTTCCTAGTCGTC
Tdrd1F4	CTAGTCTGTAGCCATTCAGAC
Tdrd1R4	ACGGTTAGGACAATCTCTGATCA
Gm15389F	GAGGCAACTACTACAGATGGAGA
Gm15389R	TGGCTGGATGGCAGTGCATCA
CUFF27295F	TTCAGATGGCAACCTAAGGCAG
CUFF27295R	CCTACCAGTCATAGGTACATG
CUFF22389F	GCTGAAGTGAAGTGTCTGTCAG
CUFF22389R	GAAGACGAACCCTTCAGATGGA
CUFF13988F	CTACAGAGGTCCTCCAAGACTG
CUFF13988R	CTTCTAGAATGCATTGTCTGTTCC

Primers for the generation of *mCherry* transgenic mice

mCherry_MfeI	ATGCATACAATTGCGCCACCATGGTGAGCAAGGGCGAGGA
mCherry_XhoI	ATGCATACTCGAGCTTGTACAGCTCGTCCATGCC
PEST_SalI	ATGCATAGTCGACAGCCATGGCTTCCCGCCGGA
PEST_NotI	CTAGAGTCGCGGCCGCGCATCTACACATTGA
polyA_NotI_EcoRI_XhoI	ATGCATAGCGGCCGCCTAGAGCGAATTCCTAGACTCGAGCTGGCCGCA ATAAAATATC
polyA_HindIII	ATGCATAGAAGCTTCACACAAAAAACCAACACACAGATG
PCR_Not_lox_Eco_Xho-lox_Sal-S	ACTAAAGGGCGGCCGCATAACTTCGTATAGCATACATTATACGAAGTTA TGAATTCCTAGACTCGAG
PCR_Not_lox_Eco_Xho-lox_Sal-AS	TAGTGTCGACATAACTTCGTATAATGTATGCTATACGAAGTTATCTCGA GTCTAGGAATTC

B1mmAS_EcoRI_F	GCGAATTCTTTTGTTTTTTCGAGACAGGGTTTCTCTGTATAGCCCTGGC TGTCCTGGAACTCACTTTGTAGACCAGGCTGGCCTCGAACT
B1mmAS_XhoI_R	ATGCCTCGAGAGCCGGGCGTGGTGGCGCACGCCTTTAATCCCAGCAC TCGGGAGGCAGAG GCAGGCGGATTTCTGAGTTCGAGGCCAGCC
B2S_EcoRI_F	CGGAATTCGGGCTGGAGAGATGGCTCAGTGGTTAAGAGCACCTGACT GCTCTTCCAGAGGTCCTGAGTTCAATTCCCAGCAACCACATGGTGGCT CAC
B2S_XhoI_R	ATGCCTCGAGATATGTAAGTACACTGTAGCTGTCTTCAGACACACCAG AAGAGGGCATCAGATCTCATTACAGATGGTTGTGAGCCACCATGTGGT TGCTG
Genotyping_mCherry_F	GCATGGACGAGCTGTACAAG
Genotyping_mCherry_R	CACAAAAAACCAACACACAGATGTAATG
Genotyping_mCherry_cont_F	CCTAAGAGCTACAAACCCTAGTCAGT
Genotyping_mCherry_cont_R	GTGTGAGATCACATGACTTGCTGA
mCherry_Southern_F	CGAGTTCATCTACAAGGTGAAGCTG
mCherry_Southern_R	CTTGTACAGCTCGTCCATGCC
Primers for the quantification of *mCherry* mRNA and pre-mRNA	
mCherry mRNA F	CGAGTTCATCTACAAGGTGAAGCTG
mCherry mRNA R	GGTCTTGACCTCAGCGTCGT
mCherry pre-mRNA F	CCTCTGCTAACCATGTTCATGC
mCherry pre-mRNA R	CCCTCCATGTGCACCTTGAAG
Primers and oligos for the generation of the polyA insertion mice at the *Prelid1* 3´ UTR	
Prelid1 sgRNA S	CACC GGGTGTGGTGCAGATGCTGACC
Prelid1 sgRNA R	AAAC GGTCAGCATCTGCACCACACCC
Prelid1 sgRNA T7 F	TTAATACGACTCACTATAGGGTGTGGTGCAGATGCTGACC
Common sgRNA R	AAAAGCACCGACTCGGTGCC
Cas9 T7 F	TAATACGACTCACTATAGGGAGAATGGACTATAAGGACCACGAC
Cas9 R	GCGAGCTCTAGGAATTCTTAC
Prelid1 recombination	AAAGAGCAACCTGCAGCCCTCTCTGCTGTCTGGGGTGGTGGGTCAGG GCCCTGTCCTGGTAATAAAATATCTTTATTTTCATTACATCTGTGTGTTG GTTTTTTGTGTGCAGCATCTGCACCACACCCAGCTCTGTCTGCCGGGC AGAGGGGTGCAGTGACTTGCCCAA
Genotyping Prelid1 F	CCAAGGACCTGGCCAACAAG
Genotyping Prelid1 R	GCTATCTGACTCCAGAACTG
Primers for 3´ RACE analysis in polyA insertion mice	

RT primer	CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGTTTTTTTTTTTTTT
	TTT
Prelid1 RACE 1st F	CTCTTCACCCTGATGCTGCA
Prelid1 RACE 1st R	CCAGTGGCAGAGTGACG
Prelid1 RACE 2nd F	CCAAGGACCTGGCCAACAAG
Prelid1 RACE 2nd R	GAGGACTCGAGCTCAAGC
Primers for the quantification of *Prelid1* mRNA and pre-mRNA	
Prelid1 mRNA F1	CTCTTCACCCTGATGCTGCA
Prelid1 mRNA R1	GAACACTTGGTCCCAGGAAC
Prelid1 mRNA F2	CACCTGGAACATCAACCATG
Prelid1 mRNA R2	CCTGGACAGCTCTGGAGA
Prelid1 mRNA F3	TGTCCTGGTCAGCATCTGCA
Prelid1 mRNA R3	CACCTGCGCACTGCTGTGACCTT
Prelid1 pre-mRNA F	CACCTGGAACATCAACCATG
Prelid1 pre-mRNA R	GCAGAACTCTAGAGATGCAGCT

SUPPLEMENTAL REFERENCES

Arvey A, Larsson E, Sander C, Leslie CS, Marks DS. 2010. Target mRNA abundance dilutes microRNA and siRNA activity. *Molecular systems biology* **6**: 363.

Bao J, Ma HY, Schuster A, Lin YM, Yan W. 2013. Incomplete cre-mediated excision leads to phenotypic differences between Stra8-iCre; Mov10l1(lox/lox) and Stra8-iCre; Mov10l1(lox/Delta) mice. *Genesis* **51**(7): 481-490.

Bastos H, Lassalle B, Chicheportiche A, Riou L, Testart J, Allemand I, Fouchet P. 2005. Flow cytometric characterization of viable meiotic and postmeiotic cells by Hoechst 33342 in mouse spermatogenesis. *Cytometry Part A : the journal of the International Society for Analytical Cytology* **65**(1): 40-49.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**(6): 1089-1103.

Carbonell A, Fahlgren N, Garcia-Ruiz H, Gilbert KB, Montgomery TA, Nguyen T, Cuperus JT, Carrington JC. 2012. Functional analysis of three Arabidopsis ARGONAUTES using slicer-defective mutants. *The Plant cell* **24**(9): 3613-3629.

Deng W, Lin H. 2002. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Developmental cell* **2**(6): 819-830.

Di Giacomo M, Comazzetto S, Saini H, De Fazio S, Carrieri C, Morgan M, Vasiliauskaite L, Benes V, Enright AJ, O'Carroll D. 2013. Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. *Molecular cell* **50**(4): 601-608.

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science* **315**(5818): 1587-1590.

Reuter M, Berninger P, Chuma S, Shah H, Hosokawa M, Funaya C, Antony C, Sachidanandam R, Pillai RS. 2011. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* **480**(7376): 264-267.

Sadate-Ngatchou PI, Payne CJ, Dearth AT, Braun RE. 2008. Cre recombinase activity specific to postnatal, premeiotic male germ cells in transgenic mice. *Genesis* **46**(12): 738-742.

Scotto-Lavino E, Du G, Frohman MA. 2006. 3' end cDNA amplification using classic RACE. *Nature protocols* **1**(6): 2742-2745.

Vourekas A, Zheng Q, Alexiou P, Maragkakis M, Kirino Y, Gregory BD, Mourelatos Z. 2012. Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi

in spermiogenesis. *Nature structural & molecular biology* **19**(8): 773-781.

Watanabe T, Chuma S, Yamamoto Y, Kuramochi-Miyagawa S, Totoki Y, Toyoda A, Hoki Y, Fujiyama A, Shibata T, Sado T et al. 2011. MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Developmental cell* **20**(3): 364-375.

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**(7194): 539-543.

Yoshida S, Takakura A, Ohbo K, Abe K, Wakabayashi J, Yamamoto M, Suda T, Nabeshima Y. 2004. Neurogenin3 delineates the earliest stages of spermatogenesis in the mouse testis. *Developmental biology* **269**(2): 447-458.

Zheng K, Wang PJ. 2012. Blockade of Pachytene piRNA Biogenesis Reveals a Novel Requirement for Maintaining Post-Meiotic Germline Genome Integrity. *PLoS genetics* **8**(11): e1003038.

LEGENDS FOR SUPPLEMENTAL FIGURES

Supplemental Figure 1. RNA-seq analysis in *Piwil1*$^{-/-}$ and *Mov10l1* CKO mouse

(A) Length distribution of total small RNAs from adult testis. Small RNA reads (DRX001130) were annotated as previously described (Watanabe et al. 2011), and degradation products of abundant noncoding RNAs (rRNA, tRNA, snRNA, snoRNA, scRNA, srpRNA and RNA) were removed before analysis.

(B) Significant depletion of pachytene piRNAs in *Piwil1*$^{-/-}$ mice. Total RNAs from FACS-isolated late spermatocytes were 5′-labeled using [γ-^{32}P]ATP.

(C) FACS analysis of testicular cells from 23 dpp *Piwil1*$^{+/-}$ mice. In the *Piwil1* mutant allele, the genomic region corresponding to almost the entire *Piwil1* ORF was replaced with a EGFP ORF in order to express EGFP protein from the mid-pachytene spermatocyte stage to the round spermatid stage under the control of the endogenous *Piwil1* promoter (Deng and Lin 2002). The x and y-axis represent DNA content and *Piwil1*-EGFP expression, respectively. The gates for the isolation of late spermatocytes (from mid-pachytene to disakinesis) and early round spermatids are indicated.

(D) Immunocytochemistry of FACS-isolated cells from 23 dpp *Piwil1*$^{+/-}$ mice. The isolated cells were stained with anti-SCP3 antibody (late spermatocytes) or anti-MVH antibody (early round spermatids). The purity of isolated cells was estimated to be ~95% for late spermatocytes and

~98% for early round spermatids.

(E) RNA-seq results of 3 down-regulated mRNAs (*Odf1*, *Fhl5* and *Tnp1*) in *Piwil1*$^{-/-}$ mice.

(F) q-PCR analysis of *Odf1*, *Fhl5* and *Tnp1* mRNAs in late spermatocytes (FACS-isolated from 19, 23 and 28 dpp testes), early round spermatids (round spermatids FACS-isolated from 23 dpp testes) and round spermatids (round spermatids FACS-isolated from 28 dpp testes). Error bars represent the SE (n = 3). N. D. - not detected

(G) FACS analysis of testicular cells from 19 and 28 dpp *Piwil1*$^{+/-}$ mice see (C). The rates of round spermatid contamination in the late spermatocyte population are shown below.

(H) RT-qPCR analysis of 14 mRNAs randomly selected from the up-regulated mRNAs in *Piwil1*$^{-/-}$ mice. Apart from the samples used to perform RNA-seq experiments, late spermatocytes were newly isolated for qPCR analysis. RNA-seq results (top) and q-PCR results (bottom) are shown. Asterisks (*) indicate statistical significance ($P < 0.05$, one-tailed Student's t-test). Error bars represent the SE (n = 3).

(I) Almost complete depletion of pachytene piRNAs in *Mov10l1* CKO mice. Total RNAs from FACS-isolated late spermatocytes were 5'-labeled using [γ-^{32}P]ATP.

(J) Hematoxylin/eosin-stained sections of control and *Mov10l1* CKO testes. Most of the seminiferous tubules in *Mov10l1* CKO testis lack elongated spermatids. However, in some *Mov10l1* CKO seminiferous tubules (arrows), elongated spermatids and sperm are observed.

This observation may result from the presence of cells that did not undergo the excision of the *Mov10l1* gene, as the recombination efficiency of Stra8-Cre has been reported to be ~96% (Sadate-Ngatchou et al. 2008).

(K) Early round spermatid arrest in *Mov10l1* CKO (PAS/Hematoxylin). In stage 1−2 seminiferous tubules of the mutant, elongated spermatids were lost. In stage 6−7 tubules of the control mice, the PAS-positive acrosomal cap (arrow) of round spermatids covered one third of the nucleus. In the mutant, the caps are missing and PAS-positive granules (arrows), which are usually observed in step 2−3 round spermatids, are observed. In stage 8−9 tubules from the mutant, dying cells are observed. This phenotype was consistent with the phenotype observed in *Mov10l1* CKO mice by *Ngn3-Cre* (Zheng and Wang 2012), which expresses *Cre* in spermatogonia (Yoshida et al. 2004), but it was different from the zygotene-arrest phenotype reported by the study using *Stra8-Cre* to conditionally mutate *Mov10l1* (Bao et al. 2013). Although we do not know the exact reason for the difference, one possibility is the difference of mouse background. P, pachytene spermatocyte; RS, round spermatid; ES, elongated spermatid

(L) Expression levels of 6014 late spermatocyte lncRNAs identified by Cufflinks (rows) across tissues and cells (columns).

(M) Expression levels of testis-enriched late spermatocyte lncRNAs during spermatogenesis. Many lncRNAs are expressed from 15−18 dpp, when the first wave of spermatogenesis reaches

the mid-pachytene−diplotene spermatocyte stage.

(N) Relatively low expression level of lncRNAs (red) compared with mRNAs (blue) in late spermatocytes. Proportions of mRNAs (lncRNAs) that are expressed at the indicated levels are shown. RNA-seq results in *Piwil1*$^{+/-}$ late spermatocytes were used for this analysis.

Supplemental Figure 2. piRNAs were preferentially mapped to the up-regulated mRNAs

(A) The results of piRNA mapping to mRNAs by allowing 0−4 mismatches including up to two indels or by using perfect match of nucleotides 2-22 condition. Protein-coding genes are classified into five groups based on the extent of up-regulation in *Piwil1*$^{-/-}$ mice.

(B) Percentage (number) of mRNAs having no antisense piRNA hits in the six mapping conditions used in (A) (top). Number of PIWIL1-bound piRNAs mapped to the three validated target mRNAs in the six mapping conditions (below). Reads per million (RPM) of PIWIL1-bound piRNA reads are shown.

(C) Taking dilution effects into consideration in the procedure for counting the number of antisense piRNA matches. (1) Map piRNAs to RNA-seq reads allowing up to four mismatches. (2) Calculate the total number (read per million, RPM) of antisense piRNA matches (RPM^{target}) for each piRNA. (3) Map piRNAs to known mRNAs allowing up to four mismatches. (4) Count the total number of antisense piRNA matches for each mRNA. If a piRNA has a small number

of targets (RPM^{target} is smaller than or equal to x), count the piRNA match as 1. If a piRNA has many targets (RPM^{target} is larger than x), consider that this piRNA has a smaller effect on each target mRNA than piRNAs with fewer targets and count this piRNA match as x/ RPM^{target}.

(D) piRNAs are preferentially matched with the up-regulated mRNAs under any threshold condition. The results obtained by using several threshold conditions (x = 0.1, 1, 10, and 100) are shown. *D* indicates the difference of cumulative fractions at the same number (RPM) of PIWIL1-bound piRNA hits between the up-regulated (>2, $P < 0.025$) and the down-regulated (<1) mRNAs when the cumulative fraction of the down-regulated mRNAs is 0.8.

(E) Comparison of the results using several threshold conditions; the difference of cumulative fractions of mRNAs [X-axis in (D)] at the same number (RPM) of PIWIL1-bound piRNA hits [Y-axis in (D)] between the up-regulated (>2, $P < 0.025$) and the down-regulated (<1) mRNAs is shown. *D* is also indicated in this figure.

(F) Average number (RPM) of antisense piRNA matches in the down-regulated and the up-regulated mRNAs under several threshold conditions. The ratios of the average number in the up-regulated mRNAs to that in the down-regulated mRNAs are shown. There is no big differences between the ratios with (x = 0.1, 1, 10, 100, or 1000) and without ($x = \infty$) the dilution effect.

Supplemental Figure 3. Target RNAs are not enriched in RNAs immunoprecipitated with PIWIL1

(A) RNAs immunoprecipitated with anti-PIWIL1 antibody or normal IgG from *Piwil1*$^{-/ADH}$, $^{+/-}$, and $^{-/-}$ late spermatocyte lysates. Isolated RNA was dephosphorylated, 5'-end labeled and loaded onto denaturing 14% acrylamide gel. piRNA is indicated by arrow.

(B) lncRNAs regulated by PIWIL1 are not enriched in RNAs immunoprecipitated with anti-PIWIL1 antibody from *Piwil1*$^{-/ADH}$, $^{+/-}$, and $^{-/-}$ late spermatocyte lysates. Top 200 lncRNAs enriched in the immunoprecipitated RNA from animal of each genotype (n = 1 for each genotype) are shown in red on a scatter plot of RNA-seq results. Blue circle represents a lncRNA population up-regulated in both *Piwil1* and *Mov10l1* CKO, which is enriched in lncRNAs targeted by PIWIL1-piRNAs. Orange circle represents a lncRNA population up-regulated only in *Mov10l1* CKO, which is enriched in lncRNAs derived from piRNA clusters.

(C) lncRNAs derived from piRNA clusters are enriched in the immunoprecipitated RNAs from *Piwil1*$^{-/ADH}$ and $^{+/-}$ late spermarocyte lysates, but mRNAs regulated by PIWIL1 are not enriched. Shown are boxplots analyzing enrichment of target mRNAs, non-target mRNAs, and piRNA cluster-derived lncRNAs in the immunoprecipitated RNA.

(D) Three validated target mRNAs are not particularly enriched in the immunoprecipitated RNAs from *Piwil1*$^{-/ADH}$ and $^{+/-}$ late spermarocyte lysates.

Supplemental Figure 4. *Stambp-ps1*$^{Gt1/Gt1}$ mice are fertile and have no discernible phenotype in spermatogtenesis

(A, B) Hematoxylin/eosin staining of testis and cauda epididymis sections from wild type and *Stambp-ps1*$^{Gt1/Gt1}$ mice.

(C) Average weight of testes from 8 week old wild type and *Stambp-ps1*$^{Gt1/Gt1}$ mice. Error bars represent the SE (n = 3).

(D) The number of apoptotic cells in the testis sections from 2 month old *Stambp-ps1*$^{+/Gt1}$ and *Stambp-ps1*$^{Gt1/Gt1}$ mice. Y-axis represents the number of TUNEL positive cells per field of view.

(E) Expression levels of six genes predicted to be regulated by their pseudogenes. RNA-seq results in late spermatocytes from *Piwil1* KO and *Mov10l1* CKO are shown. Asterisks (*) indicate statistical significance ($P < 0.05$, Student's t-test). Error bars represent SE (n = 3).

Supplemental Figure 5. *Piwil2* and *Piwil1* mRNAs contain SINE sequences that are potentially targeted by piRNAs

(A) The distribution of PIWIL1-associated piRNAs mapped to *Piwil2* and *Piwil1* mRNA in the antisense orientation. CDS is represented by the hatched region in mRNA. SINEs are shown in gray boxes below mRNAs.

(B) *Tdrd1*, *Piwil2*, and *Piwil1* mRNAs are up-regulated in piRNA pathway mutants. The results

of RNA-seq analysis in late spermatocytes from *Piwil1* KO (top) and *Mov10l1* CKO (bottom) mice are shown. Asterisks (*) indicate statistical significance ($P < 0.025$, Student's t-test). Error bars represent the SE (n = 3).

Supplemental Figure 6. PIWIL1-associated secondary piRNAs are present in *Piwil2* CKO testes

(A) The distance between 5' ends of sense and antisense piRNAs derived from L1Md_A sequence. Published PIWIL1-associated piRNA sequences from *Piwil2* CKO and control mouse testes (Di Giacomo et al. 2013) were mapped to L1Md_A sequence (M13002) allowing 2 mismatches.

(B) The nucleotide composition of the 1st and 10th nucleotides of PIWIL1-associated piRNAs.

Supplemental Figure 7. PIWIL1 and pachytene piRNAs target B1 and B2 sequences in 3´ UTR

(A) High density of B1 (top) and B2 (bottom) sequences in the 3' UTR of the up-regulated genes in *Piwil1*$^{-/-}$ late spermatocytes. The regions examined are divided into 0.1-kb bins (window size = 100 bases, step size = 100 bases).

(B) Southern blot analysis of mCherry transgenic mice (*mCherry-B1*$^{flox/0}$ and *mCherry-B2*$^{flox/0}$ mice). Genomic DNA was digested with either ApaI and EcoRV (left) or EcoRI and EcoRV

(right). The transgenes contain ApaI and EcoRI sites, but not EcoRV site. The following mouse lines that were used for subsequent experiments are underlined: line #2 (*mCherry-B1*$^{flox/0}$ mice) and line #4 (*mCherry-B2*$^{flox/0}$ mice). Line #5 (*mCherry-B2*$^{flox/0}$ mice) has two copies of the transgene.

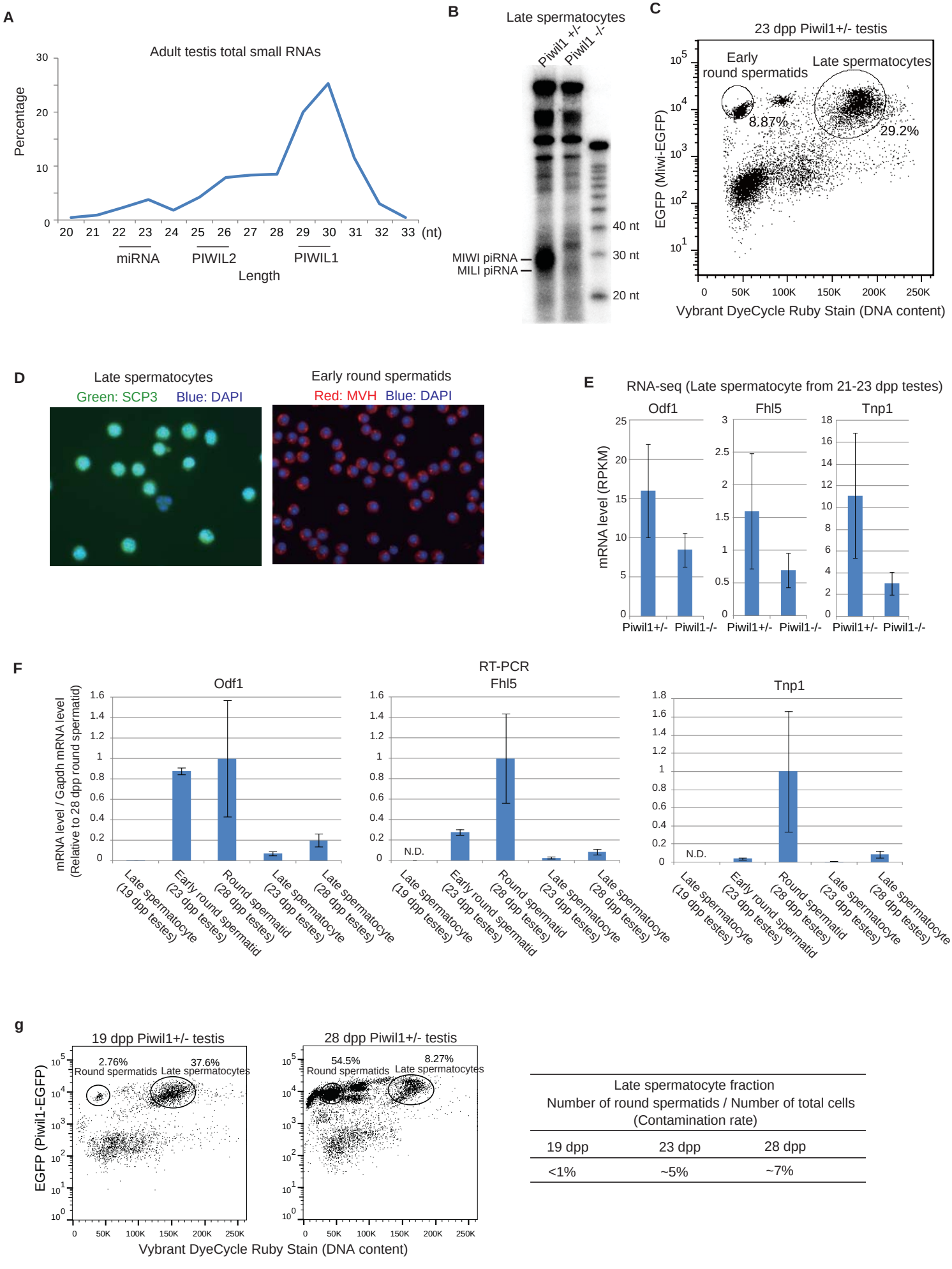
Supplemental Figure 8. Inactive retrotransposon-derived pachytene piRNAs are involved in mRNA regulation

(A) Annotation (exon, repeat, and intergenic/intron) of 25–32 nt small RNAs from E16.5 (GSM509275) and adult (DRX001130) testes. The top 10 most frequently observed retrotransposon classes are shown in the table below each pie chart. The retrotransposons indicated in red represent active retrotransposons. Reads per million (RPM) of 25–32 nt small RNA reads are shown. Small RNA sequences were annotated as previously described (Watanabe et al. 2008). The miRNAs and degradation products of abundant non-coding RNAs (rRNA, tRNA, snRNA, snoRNA, scRNA, srpRNA and RNA) were removed prior to analysis.

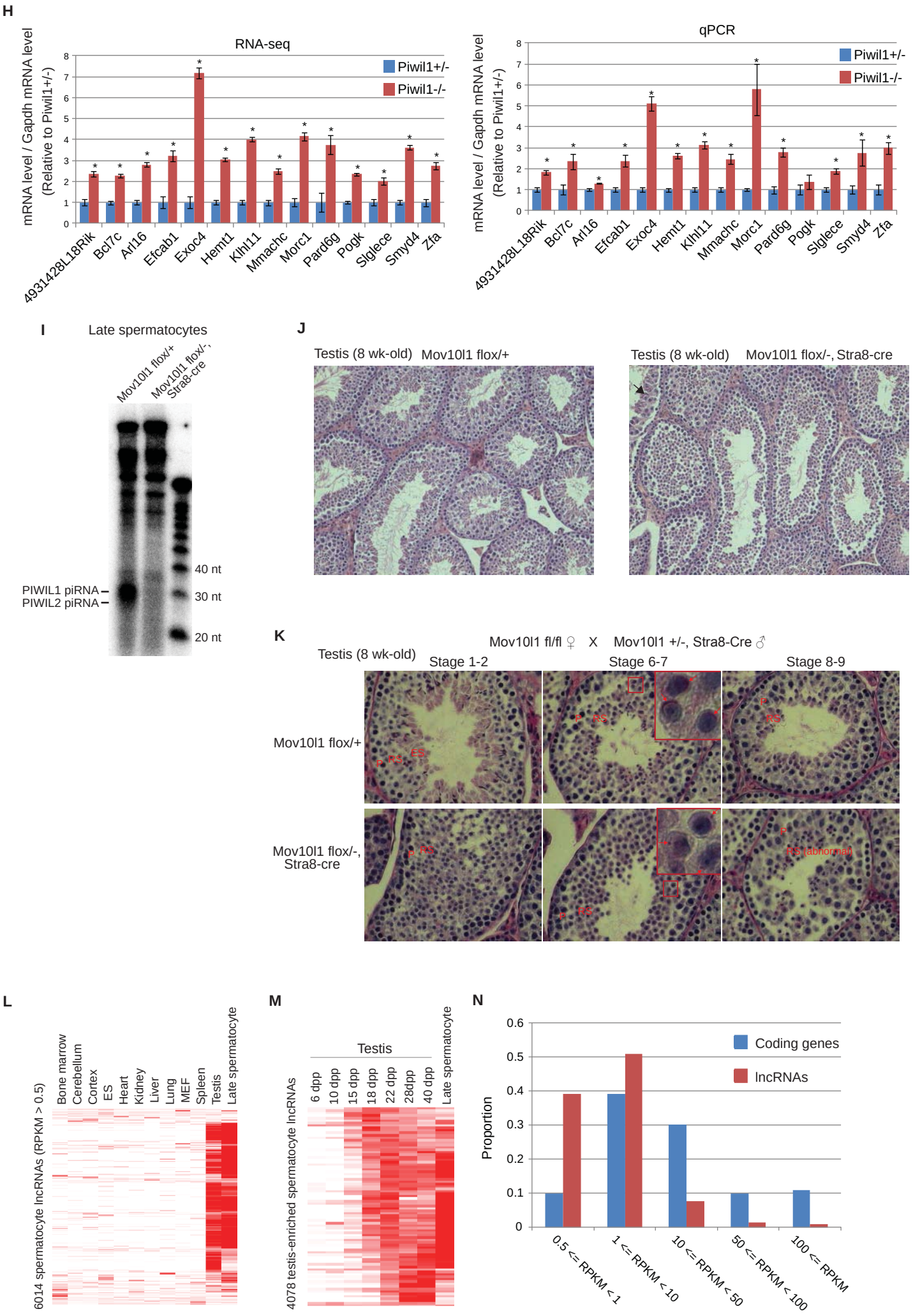
(B) q-PCR analysis of *Prelid1* mRNA and pre-mRNA in *Piwil1*$^{-/-}$ and *Mov10l1* CKO late spermatocytes. The locations of the primers used for the analyses of the mRNA are shown in Figure 7A. Error bars represent the SE (n = 3). *P < 0.05 (Student's t-test).

(C) q-PCR analysis of truncated *Prelid1* mRNA in *Piwil1*$^{-/-}$ late spermatocytes.

Supplemental Figure 1

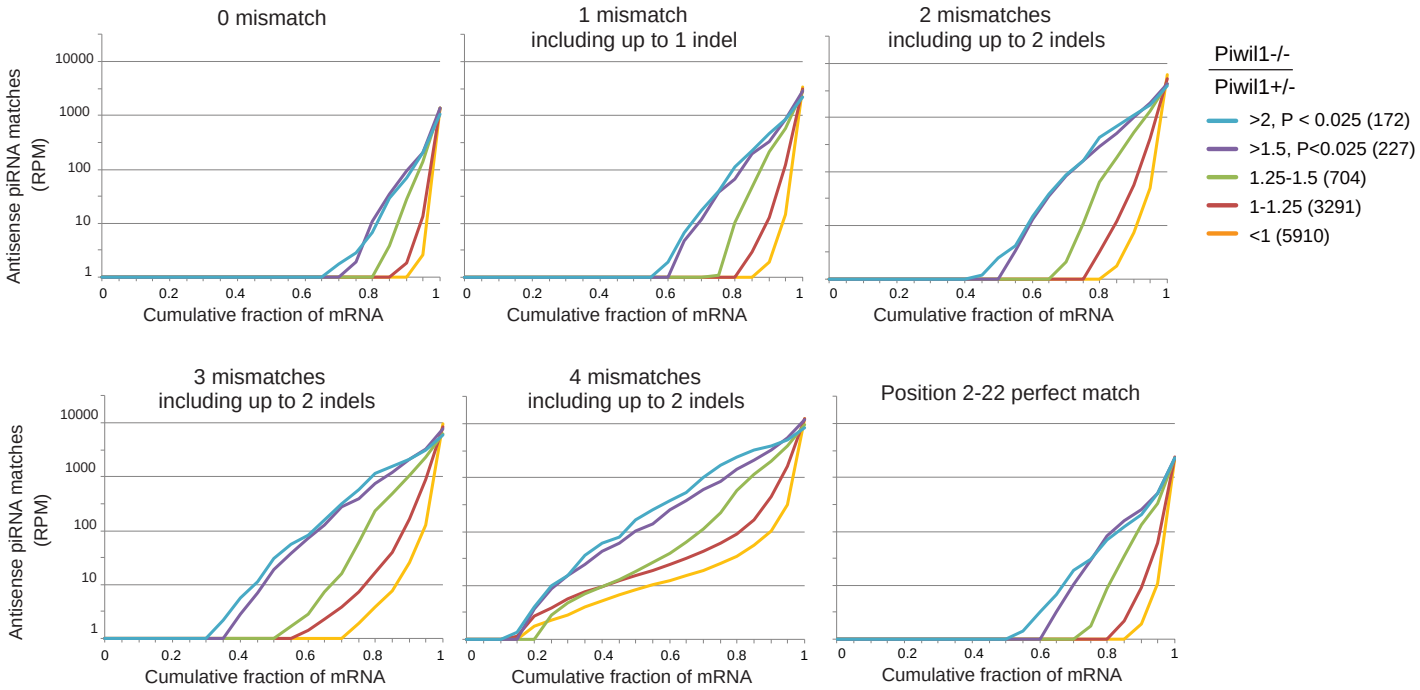


Late spermatocyte fraction Number of round spermatids / Number of total cells (Contamination rate)		
19 dpp	23 dpp	28 dpp
<1%	~5%	~7%


H
RNA-seq
qPCR
mRNA level / Gapdh mRNA level (Relative to Piwil1+/-)
Piwil1+/-
Piwil1-/-
4931428L18Rik
Bcl7c
Arl16
Efcab1
Exoc4
Hemt1
Klhl11
Mmachc
Morc1
Pard6g
Pogk
Slglece
Smyd4
Zfa
I
Late spermatocytes
Mov10l1 flox/+
Mov10l1 flox/-, Stra8-cre
PIWIL1 piRNA
PIWIL2 piRNA
40 nt
30 nt
20 nt
J
Testis (8 wk-old) Mov10l1 flox/+
Testis (8 wk-old) Mov10l1 flox/-, Stra8-cre
K
Mov10l1 fl/fl ♀ X Mov10l1 +/-, Stra8-Cre ♂
Testis (8 wk-old)
Stage 1-2
Stage 6-7
Stage 8-9
Mov10l1 flox/+
Mov10l1 flox/-, Stra8-cre
P
RS
ES
RS (abnormal)
L
6014 spermatocyte lncRNAs (RPKM > 0.5)
Bone marrow
Cerebellum
Cortex
ES
Heart
Kidney
Liver
Lung
MEF
Spleen
Testis
Late spermatocyte
M
Testis
4078 testis-enriched spermatocyte lncRNAs
6 dpp
10 dpp
15 dpp
18 dpp
22 dpp
28dpp
40 dpp
Late spermatocyte
N
Proportion
Coding genes
lncRNAs
0.5 <= RPKM < 1
1 <= RPKM < 10
10 <= RPKM < 50
50 <= RPKM < 100
100 <= RPKM

Supplemental Figure 2

A



B

	Percentage (number) of mRNAs without antisense piRNA matches				
Piwil1-/- / Piwil1+/-	<1 (5910)	1-1.25 (3291)	1.25-1.5 (704)	>1.5, P<0.025 (227)	>2, P < 0.025 (172)
0 mismatch	0.93 (5495)	0.89 (2916)	0.82 (575)	0.73 (227)	0.67 (172)
1 mismatch	0.88 (5213)	0.82 (2696)	0.75 (528)	0.62 (140)	0.58 (100)
2 mismatches	0.84 (4936)	0.75 (2480)	0.68 (477)	0.50 (114)	0.45 (77)
3 mismatches	0.70 (4161)	0.58 (1909)	0.53 (375)	0.37 (83)	0.33 (56)
4 mismatches	0.17 (1034)	0.15 (486)	0.20 (143)	0.15 (35)	0.15 (26)
Position 2-22 perfect match	0.75 (4448)	0.70 (2299)	0.66 (467)	0.55 (125)	0.47 (81)

	Number (RPM) of antisense piRNA matches		
	Stambp	Tdrd1	Prelid1
0 mismatch	0	0	0
1 mismatch	26.92	0	0
2 mismatches	92.29	3.85	0
3 mismatches	271.09	21.15	0
4 mismatches	375.87	201.92	70.18
Position 2-22 perfect match	44.22	0.96	0

C

1. Mapping of each piRNA to RNA-seq reads

2. Calculation of RPM^{target} for each piRNA

RPM^{target} : Total expression level (RPM) of potential target RNAs

RPM^{target} = (a + c + g)

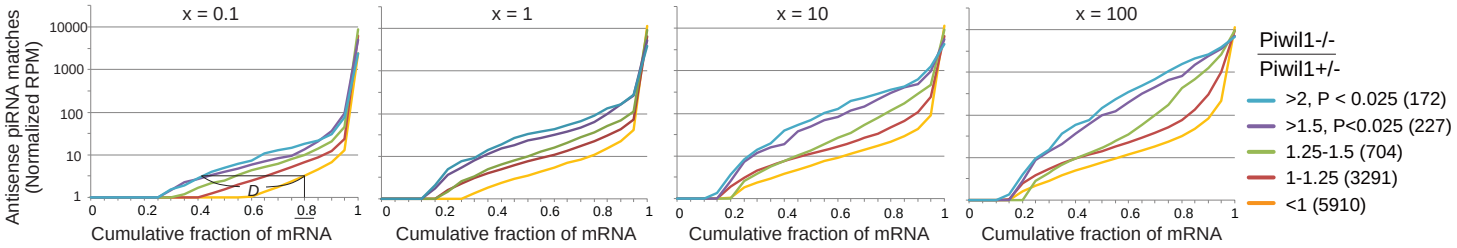
3. Mapping of piRNAs to each mRNA

4. Counting the total number of antisense piRNA matches for each mRNA

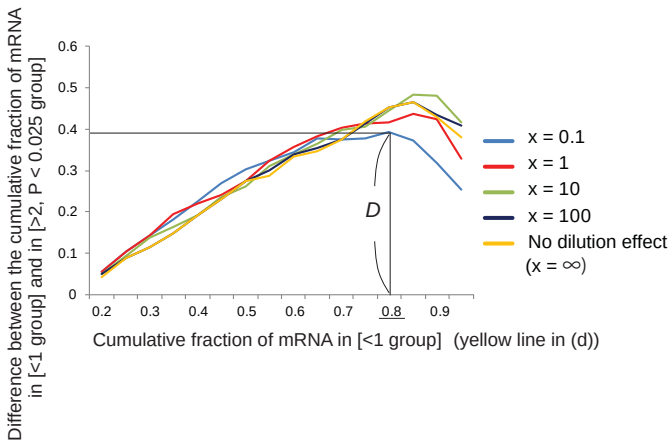
If (RPM^{target} <= x) { count a piRNA hit as 1};

If (RPM^{target} > x) { count a piRNA hit as x / RPM^{target}};

D



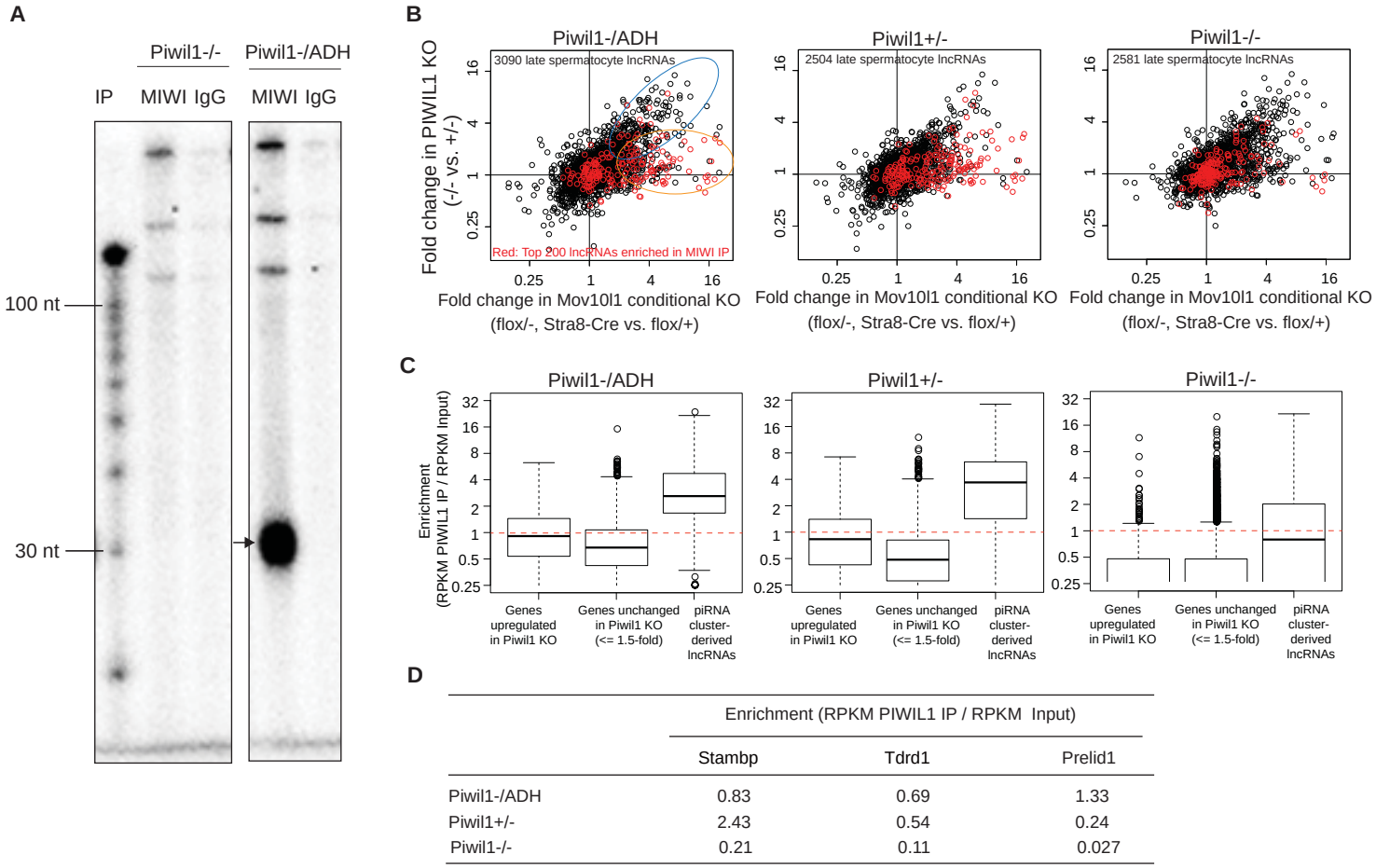
E



F

	Average number of antisense piRNA matches per mRNA (Normalized RPM)		
	[<1 group] = A	[>2, P < 0.025 group] = B	Ratio (B/A)
x = 0.1	4.51	48.07	10.66
x = 1	14.89	120.18	8.07
x = 10	30.16	279.48	9.27
x = 100	75.22	823.16	10.94
x = 1000	108.14	1145.96	10.6
No dilution effect (x = ∞)	108.22	1145.99	10.59

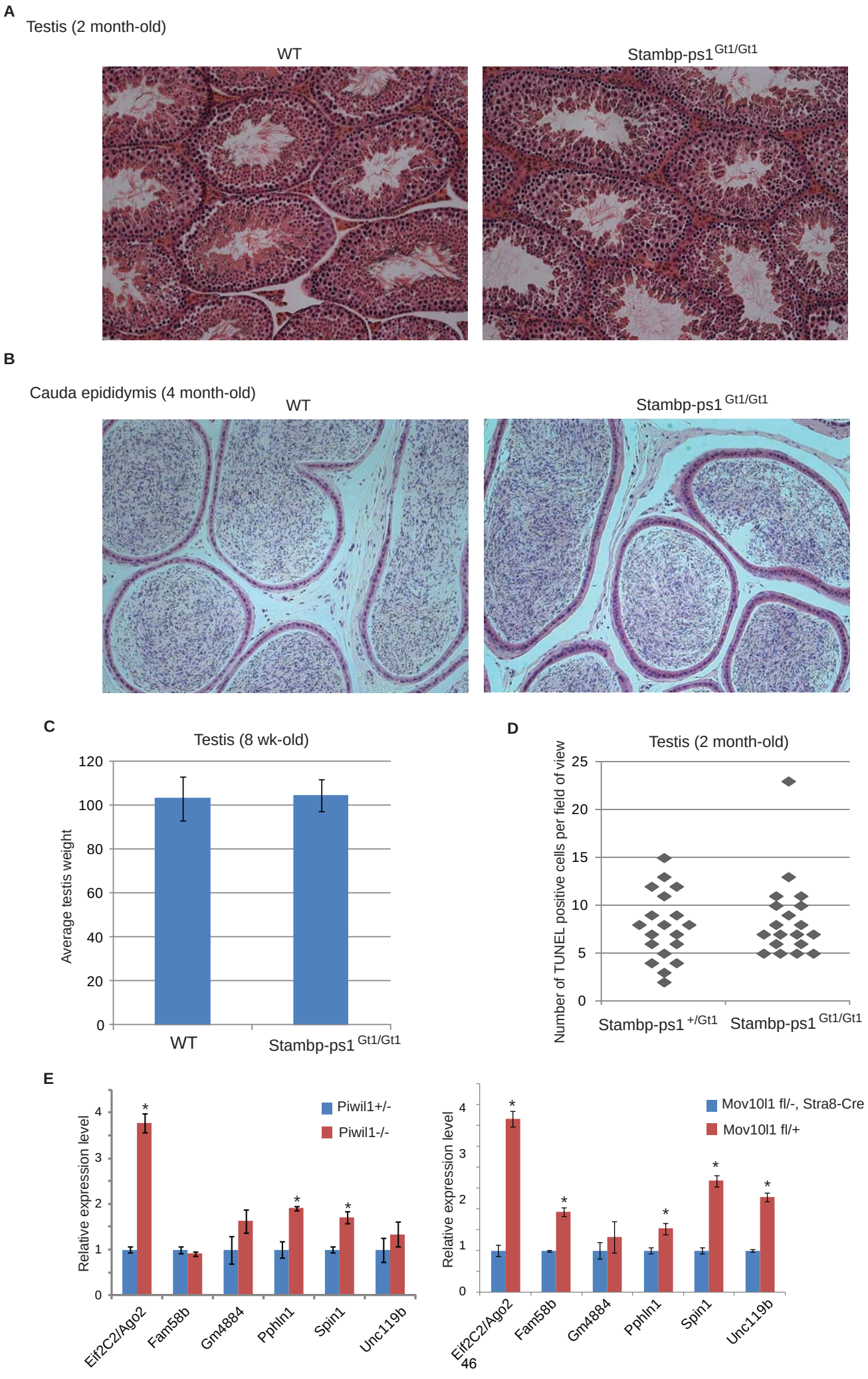
Supplementary Figure 3



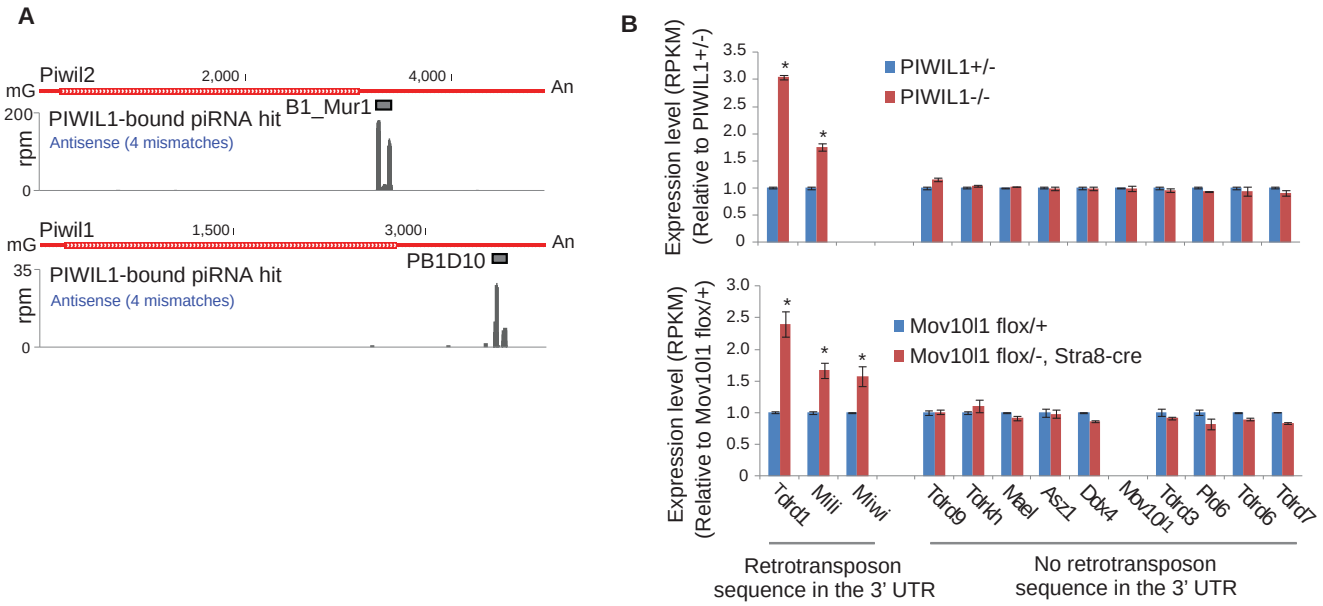
D

	Enrichment (RPKM PIWIL1 IP / RPKM Input)		
	Stambp	Tdrd1	Prelid1
Piwil1-/ADH	0.83	0.69	1.33
Piwil1+/-	2.43	0.54	0.24
Piwil1-/-	0.21	0.11	0.027

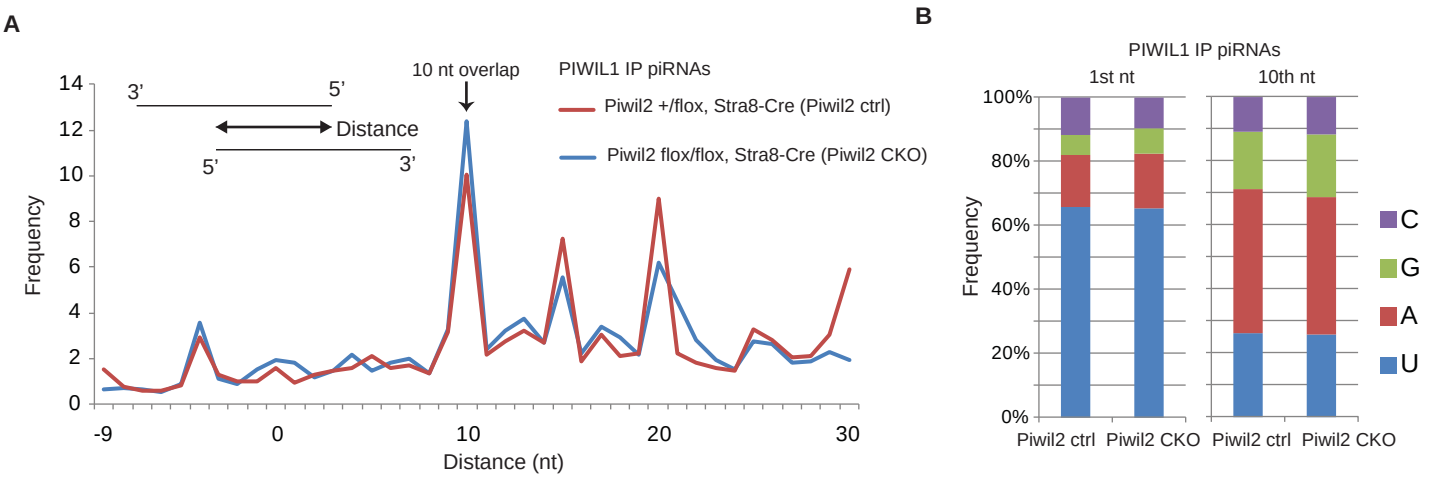
Supplementary Figure 4



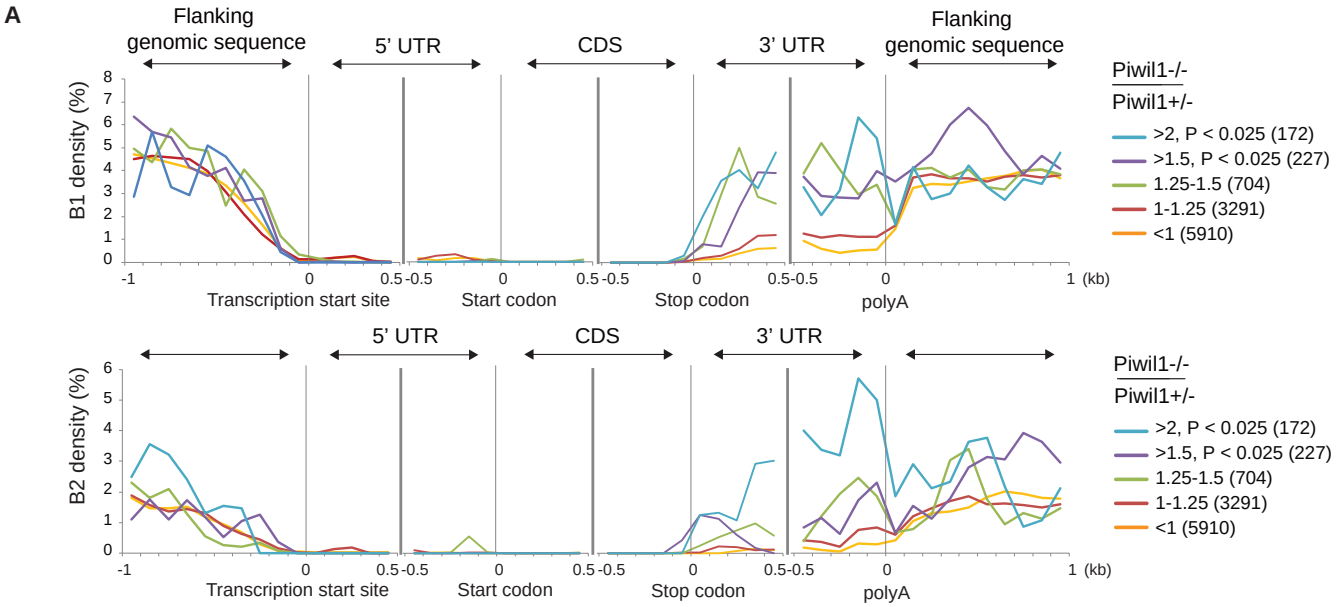
Supplementary Figure 5

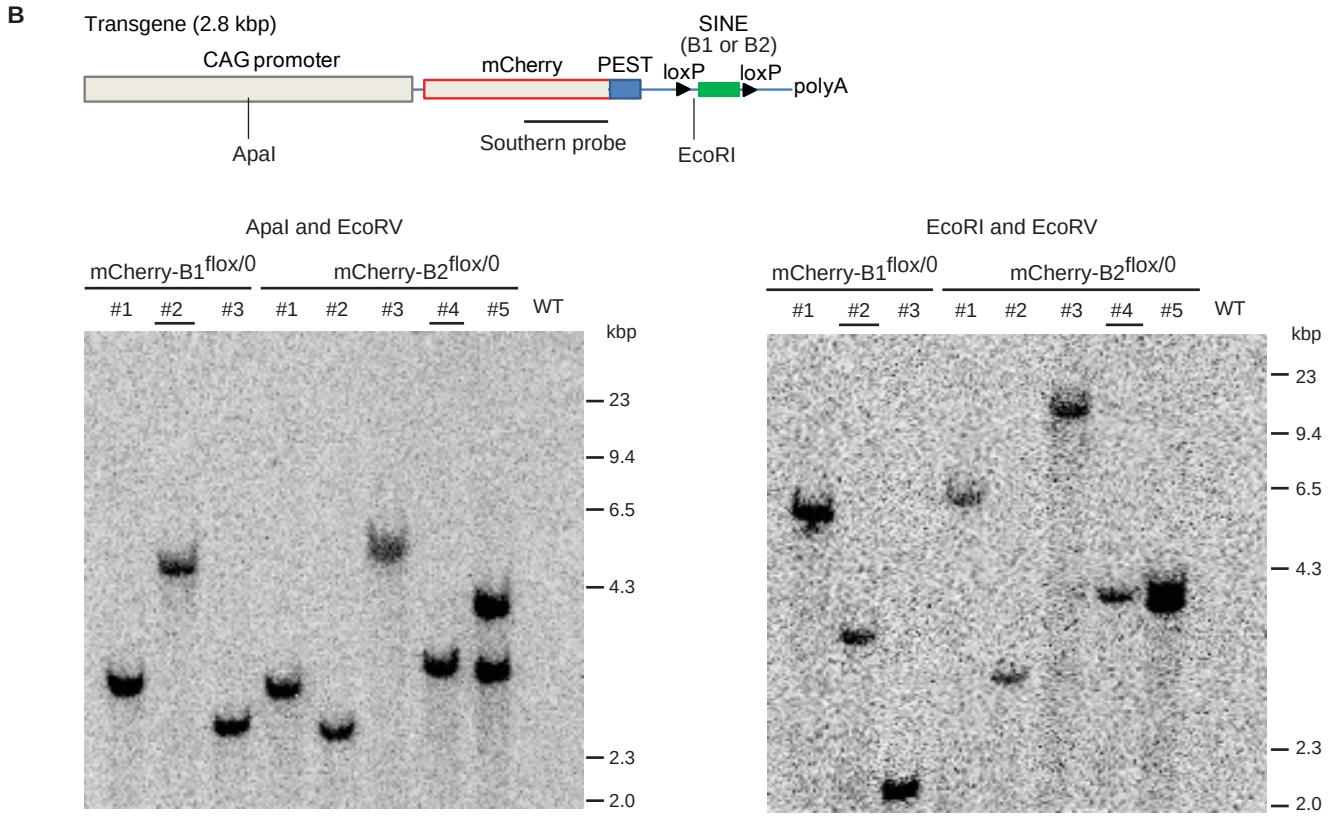


Supplementary Figure 6



Supplementary Figure 7

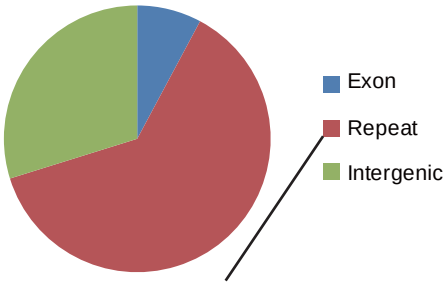




Supplementary Figure 8

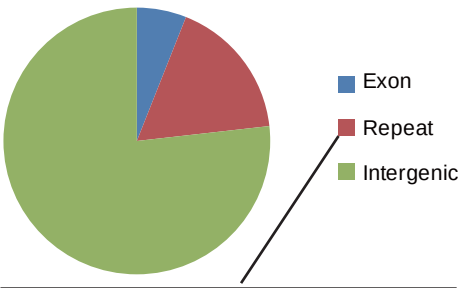
A

Total small RNA (25-32 nt)
E16.5 testis (Gonocyte piRNAs)



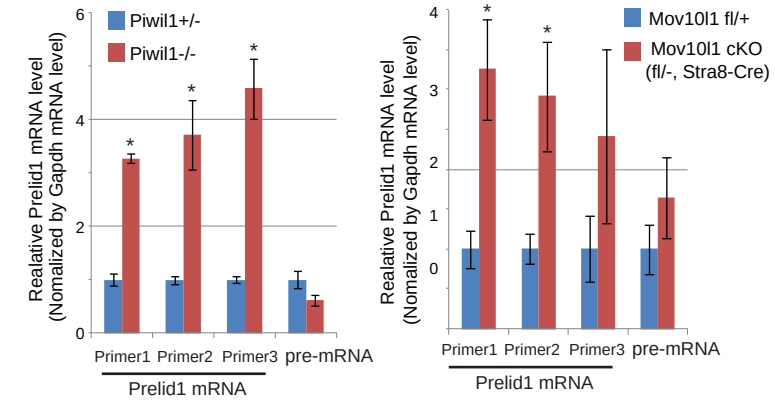
Class	Read number (Per million reads)
L1Md_T	65483.79946
L1Md_A	61212.96236
IAPEz-int	42941.48472
L1Md_F2	37321.76842
MMERGLN-int	20862.34114
MMERVK10C-int	12723.60152
IAPEy-int	10896.57645
RLTR10-int	10256.64525
L1Md_F	9893.682303
RLTR4_MM-int	9799.293253

Total small RNA (25-32 nt)
Adult testis (Pachytene piRNAs)



Class	Read number (Per million reads)
L2c	8788.98885
MIRc	7158.907229
RMER10A	6443.489887
MER20	5212.876924
L2a	4799.531945
L2	4658.214542
L1M2	4299.487064
Charlie1a	3848.430739
Lx6	3353.54583
MTE2a	2896.334188

B



C

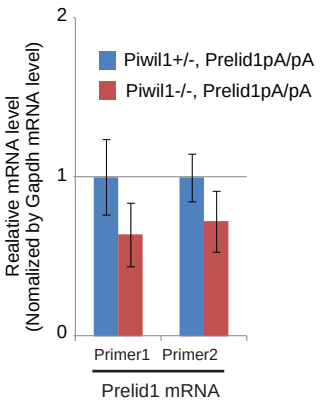


Table S1. Summary of statistics for RNA-seq.

Genotype	Cell type	Library type	Cycles	No. of raw reads	No. of unique reads
Piwil1+/-1	Late spermatocyte (21-23 dpp)	Stranded	100	137536285	88771002
Piwil1+/-2	Late spermatocyte (21-23 dpp)	Stranded	100	116004827	72531925
Piwil1+/-3	Late spermatocyte (21-23 dpp)	Stranded	100	154583767	101051156
Piwil1-/-1	Late spermatocyte (21-23 dpp)	Stranded	100	114820032	73178080
Piwil1-/-2	Late spermatocyte (21-23 dpp)	Stranded	100	156057978	100364905
Piwil1-/-3	Late spermatocyte (21-23 dpp)	Stranded	100	111658852	51696123
Mov10l1 fl/+1	Late spermatocyte (21-23 dpp)	Stranded	50	44558819	25212879
Mov10l1 fl/+2	Late spermatocyte (21-23 dpp)	Stranded	50	41811917	22892500
Mov10l1 fl/+3	Late spermatocyte (21-23 dpp)	Stranded	50	35257924	22200539
Mov10l1 fl/-, Stra8-Cre1	Late spermatocyte (21-23 dpp)	Stranded	50	83767521	50195406
Mov10l1 fl/-, Stra8-Cre2	Late spermatocyte (21-23 dpp)	Stranded	50	98966898	60207082
Mov10l1 fl/-, Stra8-Cre3	Late spermatocyte (21-23 dpp)	Stranded	50	55138896	33298523
Piwil1+/-	Leptotene/Zygotene (15 dpp)	Unstranded	50	31732213	19156521
Piwil1+/-	Mid-pachytene (21-23 dpp)	Unstranded	50	44708085	27265845
Piwil1+/-	Round spermatid (30 dpp)	Unstranded	50	39891467	24843412
C57BL/6	6 dpp testis	Stranded	50	48649965	38494132
C57BL/6	10 dpp testis	Stranded	50	54883013	41711181
C57BL/6	15 dpp testis	Stranded	50	96457472	68113443
C57BL/6	18 dpp testis	Stranded	50	61954056	49308123
C57BL/6	22 dpp testis	Stranded	50	77776131	61440421
C57BL/6	28 dpp testis	Stranded	50	66617747	51433881
C57BL/6	40 dpp testis	Stranded	50	59573247	40069541
Stambp-ps1+/Gt1(smallRNA)	23 dpp testis	Stranded	50	62773871	38756005
Stambp-ps1Gt1/Gt1 (small RNA)	23 dpp testis	Stranded	50	54384974	36741378
Piwil1-/ADH (PIWIL1 IP)	Late spermatocyte	Stranded	50	135976360	23953198
Piwil1+/- (PIWIL1 IP)	Late spermatocyte	Stranded	50	167396536	14813810
Piwil1-/- (PIWIL1 IP)	Late spermatocyte	Stranded	50	174775296	10899716
Piwil1-/ADH (input)	Late spermatocyte	Stranded	50	51563288	11147778
Piwil1+/- (input)	Late spermatocyte	Stranded	50	97287492	16692288
Piwil1-/- (input)	Late spermatocyte	Stranded	50	76029124	15285184