

Supplemental Information

An Ancient Transcription Factor

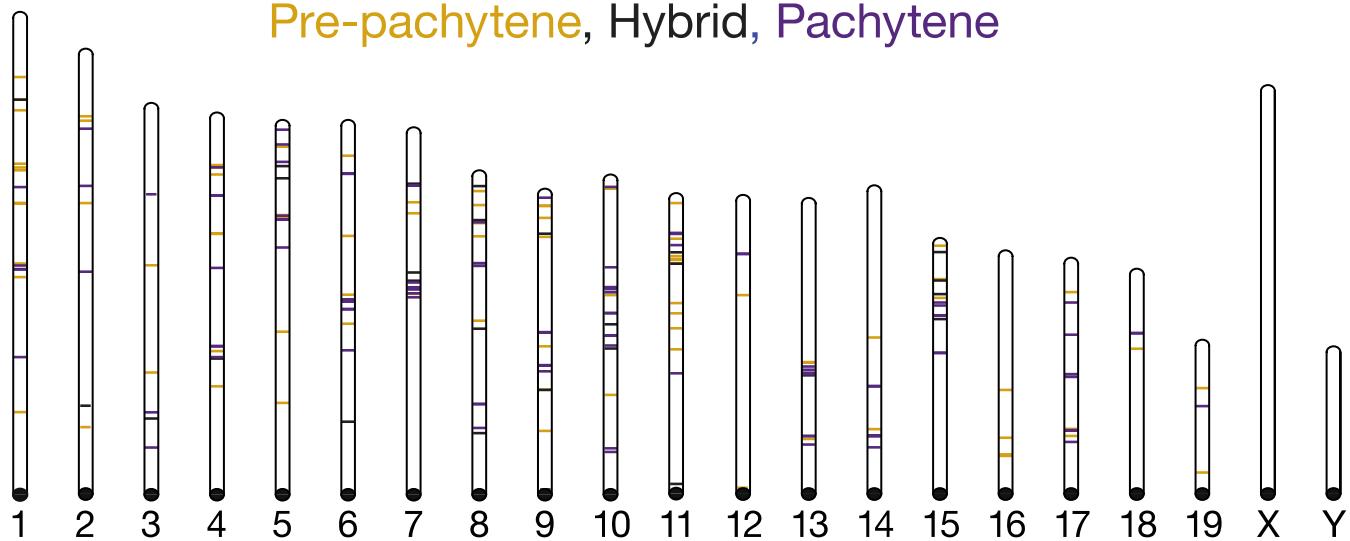
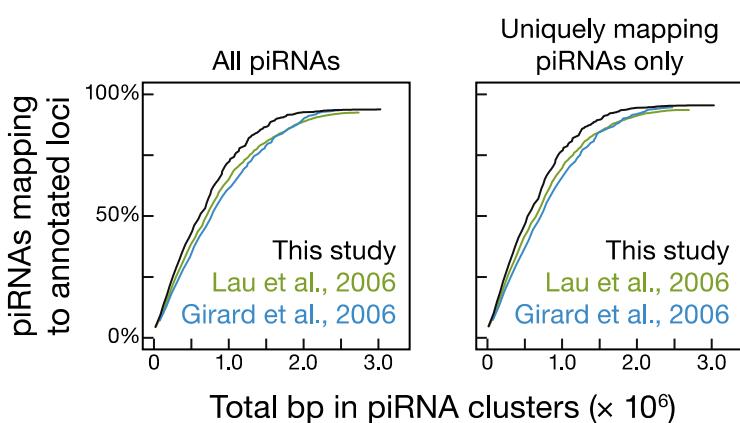
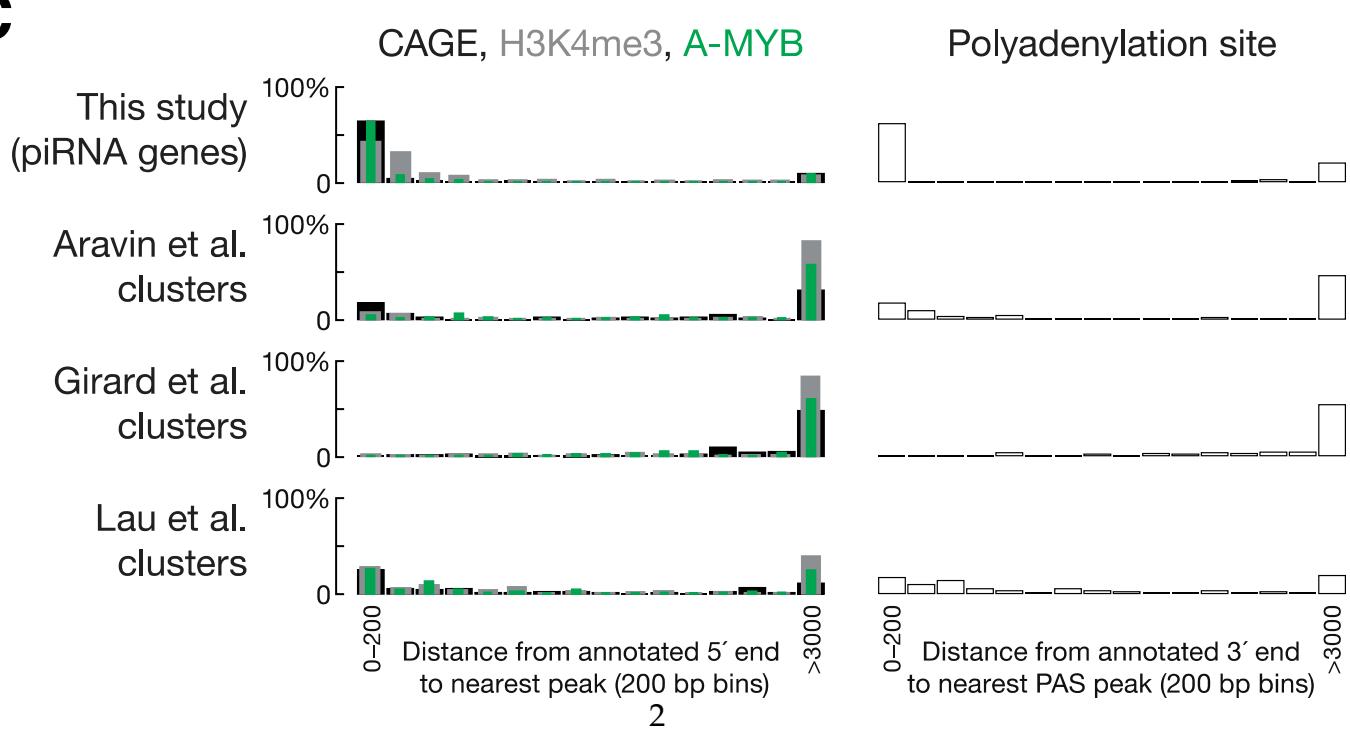
Initiates the Burst of piRNA Production

during Early Meiosis in Mouse Testes

Xin Zhiguo Li, Christian K. Roy, Xianjun Dong, Ewelina Bolcun-Filas, Jie Wang, Bo W. Han, Jia Xu, Melissa J. Moore, John C. Schimenti, Zhiping Weng, and Phillip D. Zamore

A

Pre-pachytene, Hybrid, Pachytene

**B****C**

**Figure S1. The Major piRNA-Producing Genes of the Postpartum Mouse Testis, Related to
Figure 1**

- (A) Positions of the 214 major piRNA-producing genes on the 19 autosomes of mice. We detected no loci on the X or Y chromosomes. Throughout the Supplemental Figures, gold indicates pre-pachytene and purple indicates pachytene piRNA loci.
- (B) Cumulative distributions for all piRNAs and for uniquely mapping piRNAs comparing the piRNA loci defined by our methods and by previous approaches (Girard et al., 2006; Lau et al., 2006).
- (C) Histogram of distances (in 200 bp bins) from the annotated 5' or 3' end of a piRNA gene (this study) or cluster to the nearest tag cluster peak for transcript 5' (CAGE-seq) or 3' (PAS-seq) ends or peak summits for H3K4me3 or A-MYB binding.

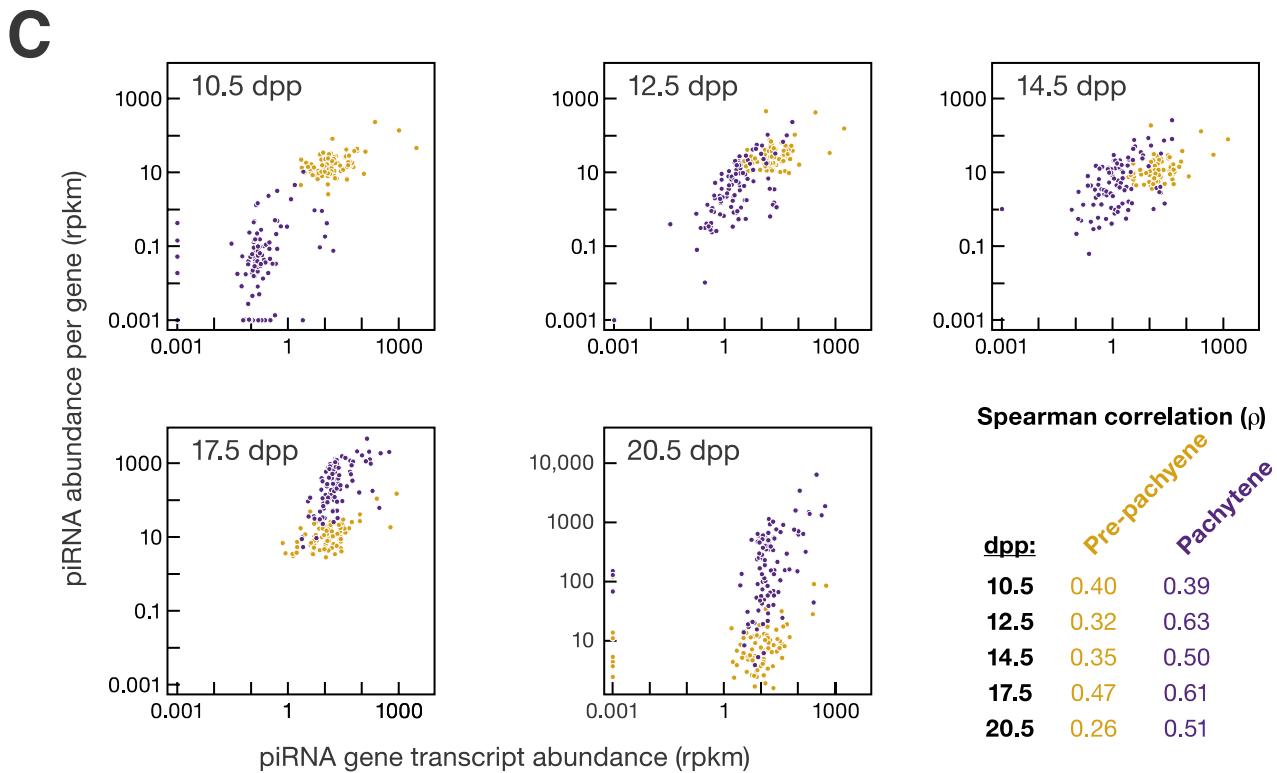
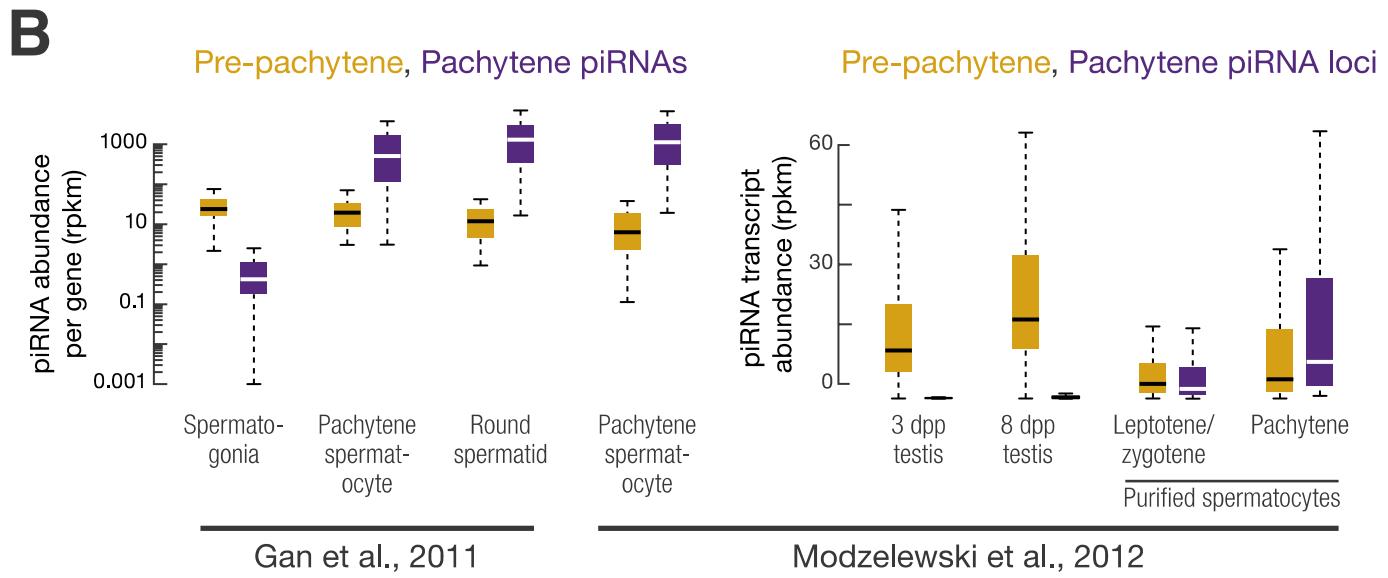
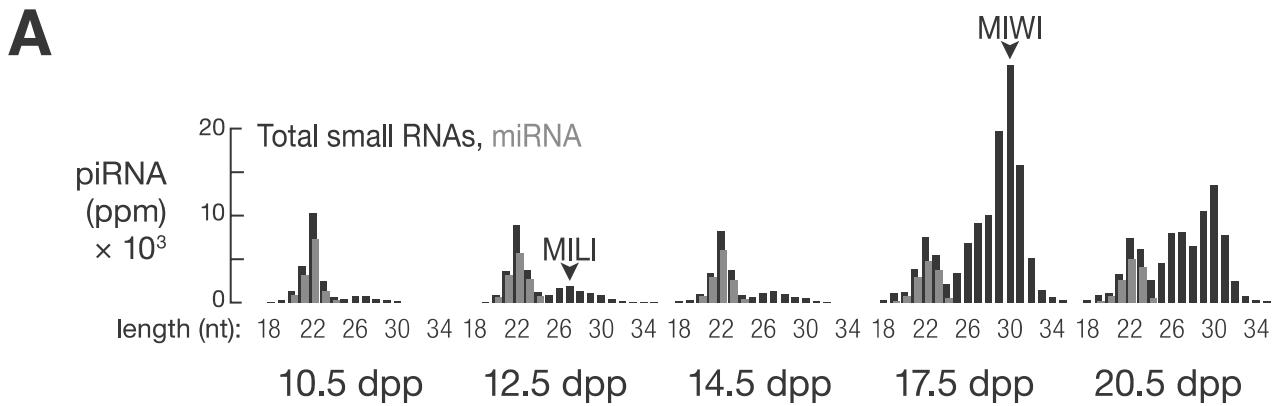


Figure S2. Pre-Pachytene piRNAs Persist in Pachytene Spermatocytes, Related to Figure 2

- (A) As shown previously by others using lower temporal resolution, the modal length of piRNAs increases as spermatogenesis proceeds to more advanced stages.
- (B) Total piRNA and piRNA precursor transcript abundance per locus by class, from purified spermatogonia, spermatocytes, and round spermatids and 3 dpp and 8 dpp testis (Gan et al., 2011; Modzelewski et al., 2012).
- (C) Correlation between piRNA abundance per locus and piRNA precursor transcription from 10.5 to 20.5 dpp.

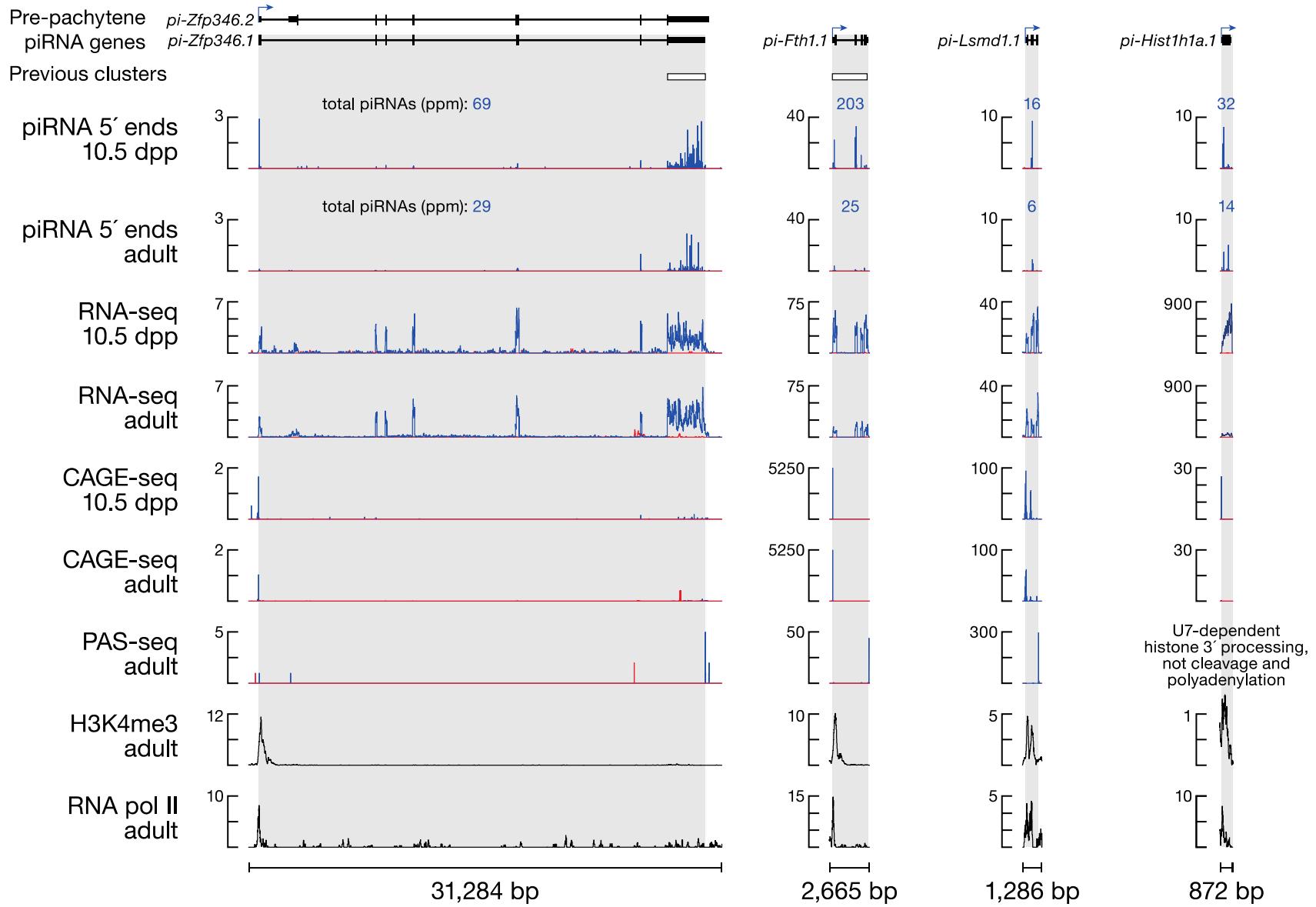


Figure S3. Examples of Pre-Pachytene piRNA Genes, Related to Figure 3

Previous cluster boundaries are from Aravin et al. (2007).

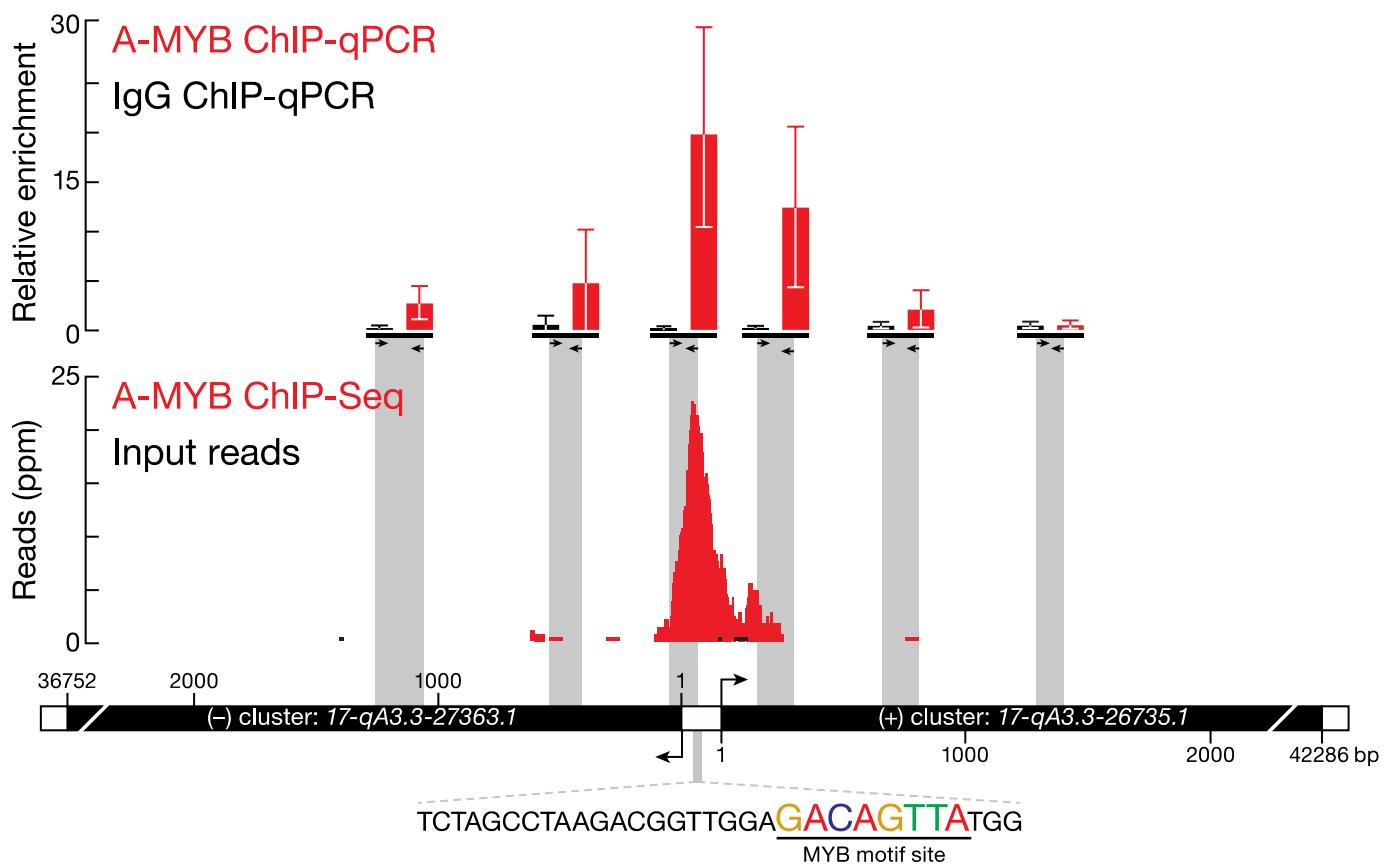
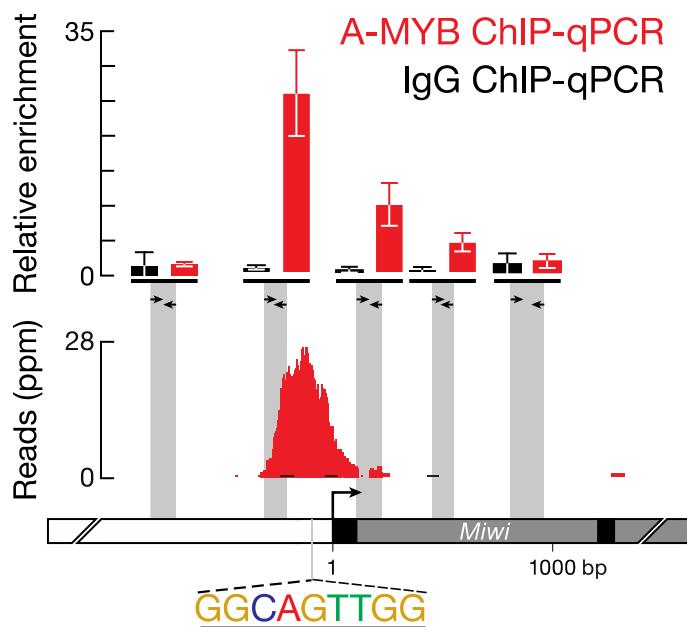
A**B**

Figure S4. ChIP-qPCR Confirms ChIP-Seq Data, Related to Figure 4

(A) A-MYB binds to the common promoter of divergently transcribed pachytene piRNA loci 17-qA3.3-27363.1 and 17-qA3.3-26735.1. The abundance of DNA fragments at the amplified region relative to a control region (mean \pm standard deviation; $n = 3$) was measured by qPCR (top). The A-MYB ChIP-seq (red) and input (black) data for this pair of genes is presented as in Figure 4B.

(B) ChIP-seq and qPCR were as in (A), but for the promoter region of *Miwi* (*Piwi1*). Also shown is the RefSeq gene model. Exons, black; introns, gray.

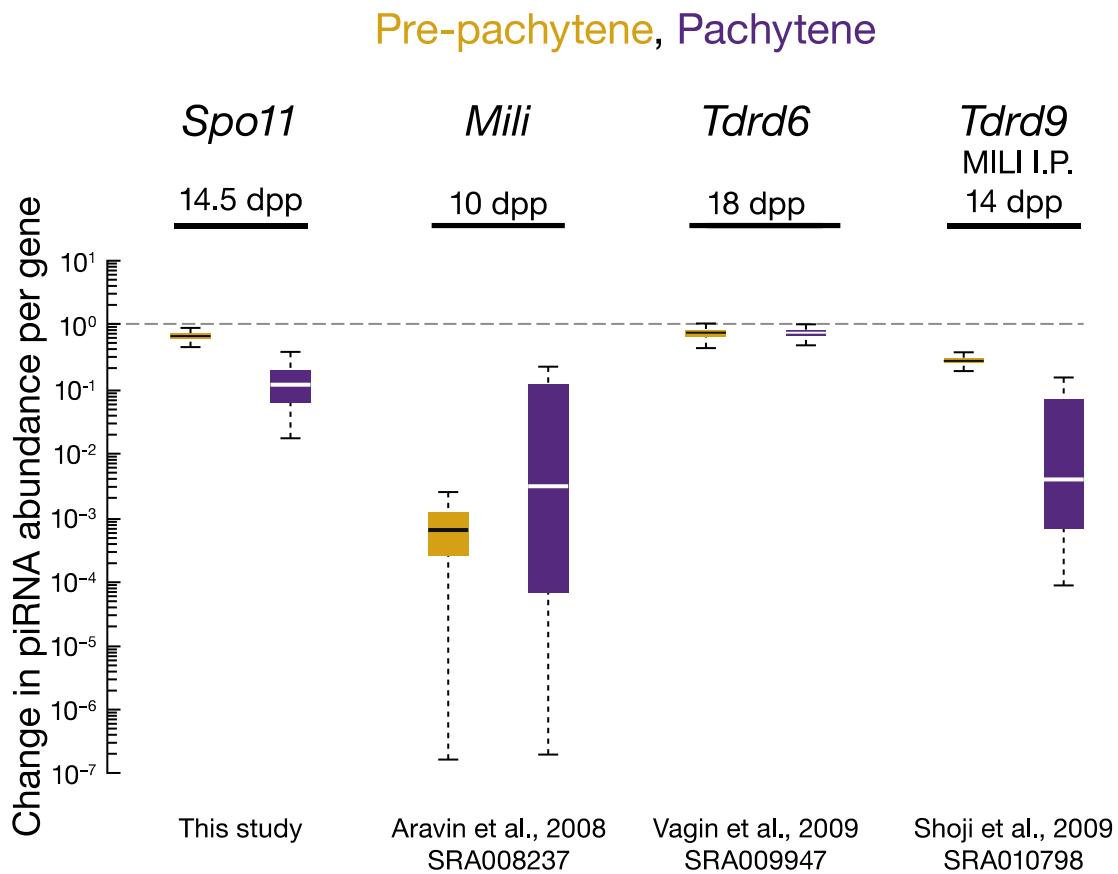


Figure S5. Change in piRNA Expression in *Spo11*, *Mili*, *Tdrd6*, and *Tdrd9* Mutants, Related to Figure 5

Change in piRNA abundance per locus (rpkm) for *Spo11* (14.5 dpp), *Mili* (*Piwil2*; 10.5 dpp), *Tdrd6* (18 dpp), and *Tdrd9* (14 dpp) mutants compared to heterozygous controls.

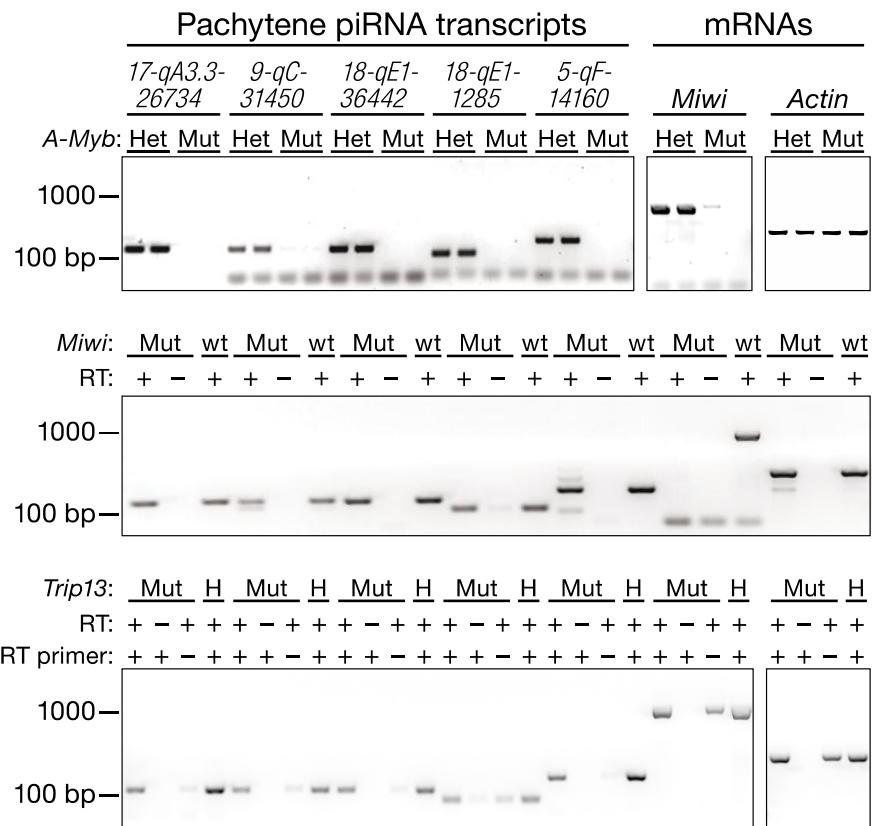
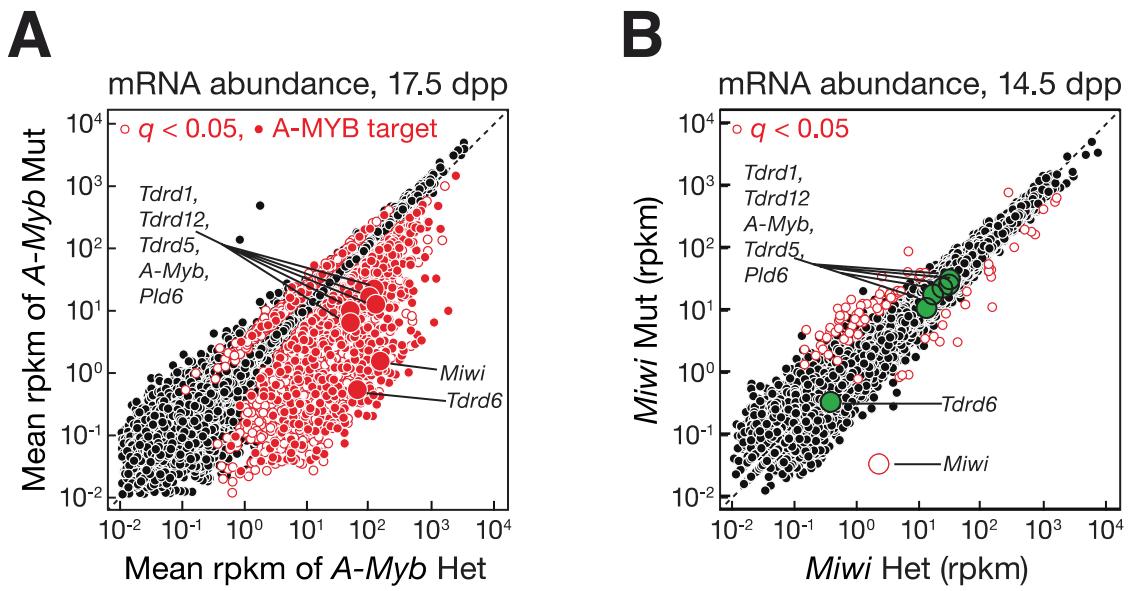


Figure S6. Pachytene piRNA Precursor Abundance in *A-Myb*, *Miwi*, and *Trip13* Mutants, Related to Figure 6

Transcripts were detected in total RNA from adult testes by RT-PCR (using random primers) for five pachytene piRNA loci as well as *Miwi* and *Actin*. Mut, mutant; Het or H, heterozygote; wt, wild type.



C

	<i>A-Myb</i> Het/Mut		<i>Miwi</i> Het/Mut	
	14.5 dpp	17.5 dpp	14.5 dpp	17.5 dpp
<i>Ddx39</i>	2.3 ± 0.2	2.9 ± 0.1	1.1	0.82
<i>Vasa (Ddx4)</i>	2.4 ± 0.2	3 ± 1	1.4	1.0
<i>Ago2 (Eif2c2)</i>	1.7 ± 0.7	1.0 ± 0.4	1.0	0.59
<i>Mael</i>	2.3 ± 0.2	3.5 ± 0.5	1.2	0.84
<i>Mov10l1</i>	2.1 ± 0.3	1.6 ± 0.2	1.1	0.85
<i>A-Myb (Mybl1)</i>	4.1 ± 0.7	5.0 ± 0.2	1.2	0.99
<i>Miwi (Piwil1)</i>	50 ± 8	100 ± 100	66	41
<i>Mili (Piwil2)</i>	2.3 ± 0.2	2.3 ± 0.6	0.85	0.85
<i>MitoPld (Pld6)</i>	3.9 ± 0.6	8.0 ± 2.1	1.2	1.4
<i>Tdrd1</i>	3.4 ± 0.4	5 ± 1	0.99	0.90
<i>Tdrd12</i>	5 ± 2	6 ± 1	1.0	0.72
<i>Tdrd5</i>	8 ± 1	10 ± 3	0.93	0.84
<i>Tdrd6</i>	64 ± 30	100 ± 100	1.1	0.84
<i>Tdrd9</i>	2.4 ± 0.5	2.8 ± 0.5	1.1	0.79

Figure S7. *A-Myb* Mutants, but Not *Miwi* Mutants, Change the Expression of RNA Silencing Pathway Genes, Related to Figure 7

(A) mRNA abundance in 17.5 dpp *A-Myb* versus heterozygous testes. The 2,860 genes with a significant ($q < 0.05$) change in steady-state mRNA abundance are shown as open red circles. Among them, 879 genes also had A-MYB peaks within 500 bp of their transcription start sites. These “A-MYB targets” are marked with filled red circles.

(B) Same as (A) but in 14.5 dpp *Miwi* mutant versus heterozygous testes. The genes encoding proteins implicated in RNA silencing pathways that were labeled in (A) and that showed no change in expression in *Miwi* mutant testes are highlighted as green filled circles. As expected, *Miwi*, showed a significant decrease in mRNA abundance in *Miwi*^{-/-} testes.

(C) The change in mRNA abundance (rpkm) in *A-Myb* and *Miwi* mutant testes versus heterozygous controls for the RNA silencing genes highlighted in Figure 7 and this figure, panels (A) and (B).

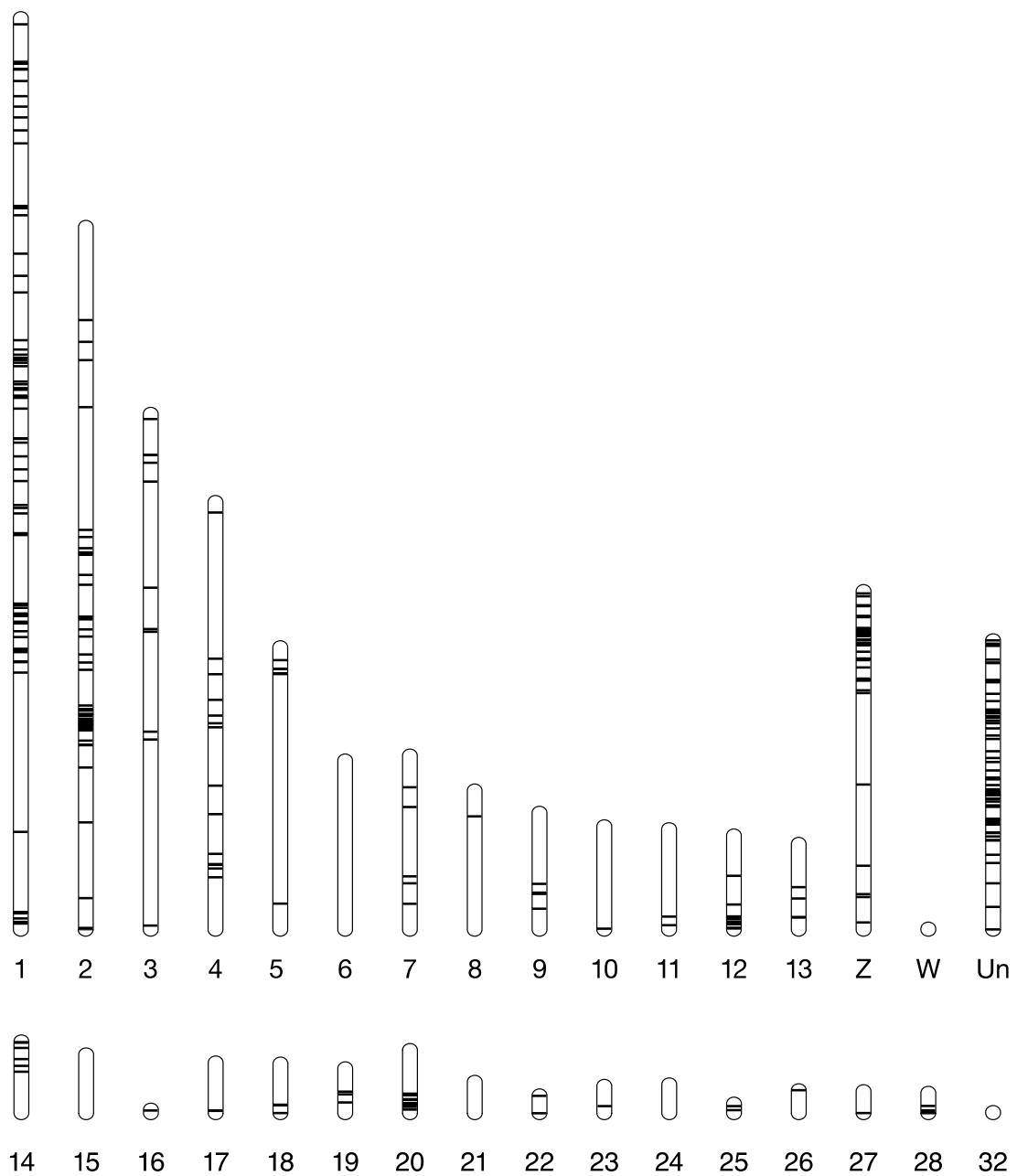


Figure S8. Genomic Locations of piRNA Clusters in the Rooster Testis, Related to Figure 8

Black horizontal lines denote the locations on the rooster (*Gallus gallus*; galGal3) chromosomes of the piRNA clusters identified by small RNA sequencing. The figure shows 324 clusters; clusters on E64 (cluster 370) and E22C19W28_E50C23 (clusters 109 and 563) are not shown.

Table S2. Three Classes of piRNA Genes, Related to Figure 3

		Pre-pachytene	Hybrid	Pachytene	
Median transcript abundance by locus (rpkm)	(1) 10.5 dpp	15	10	0.15	
	(1) 12.5 dpp	19	17	1.9	
	(1) 14.5 dpp	16	16	1.2	
	(1) 17.5 dpp	12	33	17	
	14.5 dpp	(3) <i>A-Myb</i> ^{+/−}	12 ± 2	22 ± 4	
		(3) <i>A-Myb</i> ^{−/−}	13 ± 3	13 ± 4	
		(1) <i>Miwi</i> ^{+/−}	12	14	
		(2) <i>Miwi</i> ^{−/−}	15	13	
	17.5 dpp	(3) <i>A-Myb</i> ^{+/−}	9 ± 1	31 ± 3	
		(3) <i>A-Myb</i> ^{−/−}	13 ± 3	16 ± 2	
		(1) <i>Miwi</i> ^{+/−}	14	22	
		(1) <i>Miwi</i> ^{−/−}	9.8	25	
Median piRNA abundance by locus (rpkm)	(1) 10.5 dpp	16	3.7	0.049	
	(1) 12.5 dpp	25	8.3	3.7	
	(1) 14.5 dpp	11	4.1	4.6	
	(1) 17.5 dpp	11	59	300	
	14.5 dpp	(2) <i>A-Myb</i> ^{+/−}	19	15	
		(3) <i>A-Myb</i> ^{−/−}	10 ± 4	2.9 ± 0.8	
		(2) <i>Miwi</i> ^{+/−}	14	4.7	
		(2) <i>Miwi</i> ^{−/−}	16.45	6.3	
	17.5 dpp	(1) <i>A-Myb</i> ^{+/−}	6.9	23	
		(1) <i>A-Myb</i> ^{−/−}	12	3.7	
		(1) <i>Miwi</i> ^{+/−}	26	17	
		(1) <i>Miwi</i> ^{−/−}	26	12	
Relative abundance in adult testes?		0.36%	2.7%	97%	
Congruent with protein-coding genes?		81/84	26/30	7/100	
Transcription requires <i>A-Myb</i>?		No	Partially	Yes	

The times and extents of pre-pachytene, pachytene, and hybrid piRNA production reflects their distinct patterns of expression, as well their response to loss of A-MYB and MIWI function. The number of replicates is shown in parentheses. Mean is reported where biologically independent replicates were available; when three replicates were available, S.D. is also provided.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Animals

Mice were maintained and used according to the guidelines of the Institutional Animal Care and Use Committee of the University of Massachusetts Medical School. C57BL/6J (Jackson Labs, Bar Harbor, ME, USA; stock number 664); *Mybl1^{repro9}* in a mixed 129X1/SvJ × C57BL/6J background; *Spo11^{tm1Sky}* in a C57BL/6J background; and *Piwi1^{tm1Hf}* in a mixed 129X1/SvJ × C57BL/6J background (“*Miwi*”) mice were genotyped as described (Baudat et al., 2000; Deng and Lin, 2002; Deng and Lin, 2002; Bolcun-Filas et al., 2011). Rooster testes from White Leghorn of the Cornell Special C strain, about 15 months old, were used for small RNA analysis; and testes from the Brown Leghorn strain, about one year old, were used for ChIP analysis.

RNA Sequencing

Small RNA libraries were constructed and sequenced as described (Ghildiyal et al., 2008; Seitz et al., 2008) except that 18–35 nt RNA was isolated and 2S rRNA depletion omitted. Sequencing was performed using either a Genome Analyzer GAI (36 or 76 nt reads) or HiSeq 2000 (50 nt) instrument (Illumina, San Diego, CA, USA). We analyzed published small RNA libraries from purified mouse spermatogonia (SRR069809), spermatocytes (SRR069810, GSE39652), or spermatids (SRR069811; Gan et al., 2011; Modzelewski et al., 2012); from Mili mutant or heterozygous testes at 10 dpp (SRX003089 and SRX003088; Aravin et al., 2008); from Tdrd6 mutant or heterozygous testes at 18 dpp (SRX012165 and SRX012166; Vagin et al., 2009); and MILI IP samples from *Tdrd9* mutant or heterozygous testes at 14 dpp (SRX015795, SRX015796, SRX015797, and SRX015798; Shoji et al., 2009).

Strand-specific RNA-seq libraries (Zhang et al., 2012) using Ribo-Zero Gold (Epicentre Biotechnologies, Madison, WI, USA) were sequenced using the paired-end protocol on a HiSeq 2000.

Small RNA Analysis

Small RNA sequence analysis was as described (Li et al., 2009) using mouse genome release mm9 and chicken genome release galGal3. Non-coding RNA annotations comprised data from ncRNAscan, the known tRNAs from UCSC, and 18S, 28S and 5.8S rRNAs. miRNA hairpin and mature miRNA annotation was from miRBase Release 19. Mouse and chicken transposons were annotated using Repeat Masker from UCSC. Reads that did not map to the mouse genome (mm9) were mapped to piRNA precursor transcripts to obtain splice junction-mapping small RNAs. Total small RNA libraries from different developmental stages and from mutants were normalized to the sum of all miRNA hairpin-mapping reads. Oxidized samples were calibrated to the corresponding total small RNA library via the abundance of shared, uniquely mapped piRNA species. Table S1 reports the statistics for high-throughput sequencing. For oxidized (i.e., piRNA-enriched) samples, uniquely mapping small RNAs >23 nt were mapped to each assembled piRNA precursor transcript and reported as reads per kilobase pair per million reads mapped to the genome (rpkm) using a pseudo count of 0.001.

RNA-Seq Analysis

RNA-seq reads were aligned to the genome (NCBI 37/mm9) using TopHat 2.0.4 (Trapnell et al., 2009). Reads were mapped uniquely using the '-g 1' switch. We assembled the mouse testis transcriptome (see below). For genes with multiple isoforms, the transcript with the highest average rpkm value among the three replicates of adult testes was selected for further analysis. Fragments with both reads mapped within a transcript, or to piRNA precursor transcripts, were counted using BEDTools (Quinlan and Hall, 2010). The sum of the reads aligning to the top quartile of expressed transcripts per library was used to calibrate the samples. The number of reads per transcript was normalized by length, divided by the library-specific calibration factor, and reported as rpkm with a pseudo count of 0.001. Table S1 presents the statistics for the RNA-seq data. Sequences mapping to five genes (Table S1) that overlapped with or were embedded within a piRNA gene were excluded when calculating piRNA precursor transcript abundance.

PAS-Seq Library Construction and Analysis

PAS-seq libraries (Table S1) were prepared essentially as described (Shepard et al., 2011) and sequenced using a HiSeq 2000 (100 nt read length). We first removed adaptors and performed quality control using Flexbar 2.2 (<http://sourceforge.net/projects/theflexibleadap>) with the parameters “-at 3 -ao 10 --min-readlength 30 --max-uncalled 70 --phred-pre-trim 10.” For reads beginning with GGG including (NGG, NNG or GNG) and ending with three or more adenosines, we removed the first three nucleotides and mapped the remaining sequence with and without the tailing adenosines to the mouse genome using TopHat 2.0.4. We retained only those reads that could be mapped to the genome without the trailing adenosine residues. Genome-mapping reads containing trailing adenosines were regarded as potentially originating from internal priming and thus discarded. The 3' end of the mapped, retained read was reported as the site of cleavage and polyadenylation.

CAGE Library Construction and Analysis

CAGE (cap analysis of gene expression; Table S1) was as described (Yang et al., 2011) and sequenced using a HiSeq 2000 (100 nt reads). After removing adaptor sequences and checking read quality using Flexbar 2.2 with the parameters of “-at 3 -ao 10 --min-readlength 20 --max-uncalled 70 --phred-pre-trim 10”, we retained only reads beginning with NG or GG (the last two nucleotides on the 5' adaptor). We then removed the first two nucleotides and mapped the sequences to the mouse genome using TopHat 2.0.4. All unique 5' ends of the mapped positions were considered as CAGE-tag starting sites and grouped into tag clusters using a distance-based method in which the maximal distance between two neighboring tags was required to be <20 bp. The peak position of a tag cluster was then reported as the transcription start site.

Transcriptome Assembly and Annotation

De novo transcriptome assembly from three biological replicates of strand-specific RNA-seq data from adult testes was performed using Trinity (r2012-06-08) with default parameters (Grabherr et al., 2011). The assembled RNA sequences were aligned to the mouse genome (mm9) with BLAT

(Kent, 2002), and the alignments with more than 95% of sequence length mapped and fewer than 1% mismatches retained.

We extracted novel junctions from Trinity (i.e., reads with [0-9]+M[0-9]+N[0-9]+M pattern in the CIGAR string of SAM output), and re-mapped all RNA-seq reads to these junctions, rescuing 1,402,444 reads in three replicates. Rescued reads were combined with TopHat alignments (supplied with "-max-multihits 100" to assembly through repetitive regions) and used as input for reference-based assembly.

We used Cufflinks v2.0.2 (Trapnell et al., 2010) with parameters of "-u -j 0.2 --min-frags-per-transfrag 40" to assemble transcripts. To join small transcript fragments caused by insufficient read coverage or embedded repetitive elements, two different gap-joining distance cutoffs were used for the assembly of genes ("--overlap-radius 100") and piRNA loci ("--overlap-radius 250"). We used Cuffcompare v2.0.2 (Trapnell et al., 2010) to annotate the 49,840 Cufflinks-assembled transcripts using parameters optimized for genic conditions ("--overlap-radius 100").

piRNA Precursor Transcript Annotation

We combined transcripts from the two Cufflink assemblies with those from the Trinity assembly, producing 136,069 unique transcripts. Those transcripts with 100 ppm or 100 rpkm unique mapping piRNAs at any time point (10.5, 12.5, 14.5, 17.5, 20.5 dpp and adult oxidized small RNA from testis) were selected for manual annotation.

To refine the termini of the piRNA-producing transcripts, we supplemented the RNA-seq data with high-throughput sequencing of 5' ends of RNAs bearing (5')ppp(5') cap structures (CAGE) and of the 3' ends of transcripts flanking the poly(A) tail (PAS-seq). To provide independent confirmation of the 5' ends of each piRNA-producing transcript, we used chromatin immunoprecipitation (ChIP-seq) of RNA polymerase II (pol II) and histone H3 bearing trimethylated lysine-4 (H3K4me3). Refinement of transcriptional starts required both a CAGE and a H3K4me3 peak to support the 5' end of the transcript. When no H3K4me3 peak corroborated alternative transcription start sites proposed by the CAGE data, the alternative transcripts were merged with the fully substantiated transcript.

piRNA Gene Nomenclature

When piRNA-producing genes overlap an annotated protein coding gene, we refer to them using the name of the overlapping gene preceded by ‘pi-’; when they do not, their names refer to their genomic location followed by a number indicating the piRNA abundance in ppm at 6 weeks post-partum. The last digit of a piRNA gene name specifies the rank order of expression among isoforms, determined by the highest abundance of transcripts (rpkm) observed for that gene among the six developmental stages of testis.

Grouping piRNA Precursor Transcripts

For the most abundant transcript in each locus, the abundance (rpkm) of piRNAs at each stage was expressed as a fraction of the maximum abundance reached during the developmental time course. These data were then analyzed by hierarchical clustering according to Euclidean distance and complete linkage using Cluster 3.0. Clustering results were visualized using Java Tree View 1.1.3.

Analysis of Differential Gene Expression

We determined differential gene expression using DESeq R (Anders and Huber, 2010). For each annotated mRNA, reads from each library were aligned to the most abundant assembled transcript. Transcripts with $q < 0.05$ were considered to be differentially expressed. Table S3 lists the genes that were differentially expressed in *A-Myb* at 14.5 dpp. Three biologically independent replicates were used for *A-Myb* homozygotes and heterozygotes at 14.5 and at 17.5 dpp.

Motif Discovery

For divergently transcribed piRNA gene pairs, the promoter region was defined as the region between the transcription start sites defined by CAGE peaks. Sequence motifs in these putative promoter regions were detected ab initio using MEME (Bailey and Elkan, 1994; Bailey et al., 2009) in TCM mod (any number of repetitions per sequence) and compared to existing JASPAR and TRANSFAC libraries via TOMTOM (Gupta et al., 2007). FIMO was used to detect motif sites within the putative promoters (default $p < 10^{-4}$; Grant et al., 2011).

Chromatin Immunoprecipitation

ChIP was performed as described (Chen et al., 2008) except that testes were macerated on ice and then fixed with 1.5% (w/v) formaldehyde for 20 min. Samples were then further crushed using 20 strokes with a 'B' pestle in a Dounce homogenizer (Kimble-Chase, Vineland, NJ, USA). Chromatin was sheared by sonication and immunoprecipitated using anti-A-MYB (HPA008791; Sigma, St. Louis, MO, USA), anti-RNA polymerase II (N20, sc899, Santa Cruz Biotechnology, Dallas, TX, USA), or anti-H3K4me3 (ab8580; Abcam, Cambridge, MA, USA) antibody; immunoglobulin G (IgG; Sigma, item 2729) served as a control. ChIP-quantitative PCR (qPCR) was performed using the CFX96 Real-Time PCR Detection System with SsoFast EvaGreen Supermix (Bio-Rad, Hercules, CA, USA). Data were analyzed using DART-PCR (Peirson et al., 2003). Relative ChIP enrichment values were normalized to *MyoD1*, a gene not expressed in testes. Table S1 lists ChIP-qPCR primers. ChIP-seq libraries for anti-A-MYB and control input DNA were prepared following the Illumina ChIP-seq protocol and sequenced on a HiSeq 2000 (50 nt reads).

ChIP-Seq Analysis

ChIP-seq reads were aligned to the genome using Bowtie version 0.12.7 (Langmead et al., 2009). Reads were mapped uniquely using the '-M 1 --best --strata' switches and one mismatch was allowed (-v 1). ChIP peaks were identified using MACS version 1.4.1 (Zhang et al., 2008) using default arguments, input as control, and a cutoff *p*-value = 10^{-25} was used. BEDTools was used to assign peaks to the nearest 5' end of genes. Table S1 reports sequencing statistics for ChIP-seq.

RT-PCR

Total RNA was treated with Turbo DNase (Ambion, Austin, TX, USA), and then reverse transcribed using SuperScript III (Invitrogen, Eugene, OR, USA) with random primers (Promega, Madison, WI, USA). The resulting cDNA was analyzed by conventional PCR. Table S1 lists the primers used in Figure S6.

Ping-Pong Analysis

Ping-Pong amplification was analyzed by the 5'-5' overlap between piRNA pairs from opposite genomic strands (Li et al., 2009). Overlap scores for each overlapping pair were the product of the number of reads of each of the piRNAs from opposite strands. The overall score for each overlap extend (1–30) was the sum of all such products for all chromosomes. Heterogeneity at the 3' ends of small RNAs was neglected. Z-score for 10 bp overlap was calculated using the scores of overlaps from 1–9 and 11–30 as background.

Rooster piRNA Cluster Detection

We developed a dynamic programming algorithm to identify the genomic regions with the highest piRNA density. We used oxidized small RNA reads (>23 nt) to detect clusters. We used the conservative assumption that piRNA clusters compose at most 2% of the chicken genome. We first split the genome into 1 kbp non-overlapping windows and computed piRNA abundance for each window. The mean of the top 2% of windows was used as the penalty score for the dynamic programming algorithm. The algorithm computes the cumulative piRNA abundance score as a function of the window index along each chromosome. The score at a window is the sum of the score in the previous window and the piRNA abundance in the current window, minus the penalty score; if the resulting score was negative it was reset to 0. The maximal score points to the largest piRNA cluster. We extracted the largest piRNA cluster, recomputed the scores at the corresponding windows, and searched for the next cluster. The process continued until the scores for all windows were zero. The boundaries of each cluster were further refined by including those base pairs for which piRNA abundance exceeded the mean piRNA abundance of the top 2% windows. We considered only those clusters with abundance >10 ppm for uniquely mapping piRNAs. In Figure 8E, gene models were corrected using data from our unpublished adult rooster testis RNA-seq data.

Supplemental References

- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* 11, R106.
- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J. J., Ju, J., Sheridan, R., Sander, C., Zavolan, M., and Tuschl, T. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442, 203-207.
- Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T., and Hannon, G. J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* 31, 785-799.
- Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G. J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* 316, 744-747.
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37, W202-8.
- Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2, 28-36.
- Baudat, F., Manova, K., Yuen, J. P., Jasin, M., and Keeney, S. (2000). Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Mol Cell* 6, 989-998.
- Bolcun-Filas, E., Bannister, L. A., Barash, A., Schimenti, K. J., Hartford, S. A., Eppig, J. J., Handel, M. A., Shen, L., and Schimenti, J. C. (2011). A-MYB (MYBL1) transcription factor is a master regulator of male meiosis. *Development* 138, 3319-3330.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., Loh, Y. H., Yeo, H. C., Yeo, Z. X., Narang, V., Govindarajan, K. R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W. K., Clarke, N. D., Wei, C. L., and Ng, H. H. (2008). Integration of external

signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106-1117.

Deng, W., and Lin, H. (2002). *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell* 2, 819-830.

Gan, H., Lin, X., Zhang, Z., Zhang, W., Liao, S., Wang, L., and Han, C. (2011). piRNA profiling during specific stages of mouse spermatogenesis. *RNA* 17, 1191-1203.

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L., Zapp, M. L., Weng, Z., and Zamore, P. D. (2008). Endogenous siRNAs Derived from Transposons and mRNAs in *Drosophila* Somatic Cells. *Science* 320, 1077-1081.

Girard, A., Sachidanandam, R., Hannon, G. J., and Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442, 199-202.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29, 644-652.

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017-1018.

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., and Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biol* 8, R24.

Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res* 12, 656-664.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.

Lau, N. C., Seto, A. G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. P., and Kingston, R. E. (2006). Characterization of the piRNA complex from rat testes. *Science* 313, 363-367.

- Li, C., Vagin, V. V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M. D., Syrzycka, M., Honda, B. M., Kittler, E. L., Zapp, M. L., Klattenhoff, C., Schulz, N., Theurkauf, W. E., Weng, Z., and Zamore, P. D. (2009). Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. *Cell* 137, 509-521.
- Modzelewski, A. J., Holmes, R. J., Hilz, S., Grimson, A., and Cohen, P. E. (2012). AGO4 regulates entry into meiosis and influences silencing of sex chromosomes in the male mouse germline. *Dev Cell* 23, 251-264.
- Peirson, S. N., Butler, J. N., and Foster, R. G. (2003). Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Res* 31, e73.
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.
- Seitz, H., Ghildiyal, M., and Zamore, P. D. (2008). Argonaute loading improves the 5' precision of both microRNAs and their miRNA* strands in flies. *Curr Biol* 18, 147-151.
- Shepard, P. J., Choi, E. A., Lu, J., Flanagan, L. A., Hertel, K. J., and Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* 17, 761-772.
- Shoji, M., Tanaka, T., Hosokawa, M., Reuter, M., Stark, A., Kato, Y., Kondoh, G., Okawa, K., Chujo, T., Suzuki, T., Hata, K., Martin, S. L., Noce, T., Kuramochi-Miyagawa, S., Nakano, T., Sasaki, H., Pillai, R. S., Nakatsuji, N., and Chuma, S. (2009). The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev Cell* 17, 775-787.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28, 511-515.
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105-1111.

- Vagin, V. V., Wohlschlegel, J., Qu, J., Jonsson, Z., Huang, X., Chuma, S., Girard, A., Sachidanandam, R., Hannon, G. J., and Aravin, A. A. (2009). Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev* 23, 1749-1762.
- Yang, Z., Bruno, D. P., Martens, C. A., Porcella, S. F., and Moss, B. (2011). Genome-wide analysis of the 5' and 3' ends of vaccinia virus early mRNAs delineates regulatory sequences of annotated and anomalous transcripts. *J Virol* 85, 5897-5909.
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137.
- Zhang, Z., Theurkauf, W. E., Weng, Z., and Zamore, P. D. (2012). Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence* 3, 9.