Cell, Volume *138*

Supplemental Data

The *Drosophila* HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters

Carla Klattenhoff, Hualin Xi, Chengjian Li, Soohyun Lee, Jia Xu, Jaspreet S. Khurana, Fan Zhang, Nadine Schultz, Birgit S. Koppetsch, Anetta Nowosielska, Herve Seitz, Phillip D. Zamore, Zhiping Weng, and William E. Theurkauf

Supplemental Experimental Procedures

Antibody Production

The full-length coding sequence was amplified from a *rhino* cDNA (DGRC clone RE36324) using primers with attached Gateway (Invitrogen) recombination sites. The resulting PCR product was used to make a Rhi-DONR construct which in turn was used to subclone into the 6X-His-tagged Gateway vector pDest17 (Invitrogen), yielding a 6X-His tagged Rhino fusion protein. The fusion protein was purified over a Probond Ni matrix (Invitrogen) under denaturing conditions, isolated by SDS-PAGE and used to immunize 2 Guinea pigs (Pocono Rabbit Farm and Laboratory, Inc.) using standard protocols for antibody production. Anti-Rhi antibody was affinity purified on fusion protein coupled to CNBr activated Sepharose 4B (Pharmacia) as described elsewhere (Harlow and Lane, 1999).

Rabbit polyclonal antisera directed against the N-terminal 14-16 amino acids of Piwi, Aub and Ago3 (Brennecke et al., 2007) were raised by the Pocono Rabbit Farm and Laboratory, Inc. using their proprietary Quick Draw 49 Day protocol. Antisera were affinity purified over epoxy-activated Sepharose 6B (Pharmacia) columns coupled to their respective free peptides as described (Harlow and Lane, 1999).

To generate anti-HOAP antibody, a full length HOAP cDNA was cloned into the pQE31 6X his-tagged expression vector (Qiagen) and expressed in E. coli BL21(DE3) cells. The resulting fusion protein was purified on a Probond Ni matrix under denaturing conditions (Invitrogen). The purified protein was injected into rabbits and serum was produced by standard protocol at Pocono Rabbit Farm and Laboratory, Inc.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation was performed using a modification of the procedure of Austin et al. (1999). 60 whole ovaries were dissected in 1X modified Robb's medium, (Theurkauf, 1994), transferred to 1.5 ml microfuge tubes, brought up to 1 ml of modified Robb's medium and fixed at room temperature in 1.8% formaldehyde for 10 minutes. Crosslinking was stopped by the addition of glycine to 125mM, final concentration, and further incubating for 5 minutes. Fixed ovaries were washed twice with ice-cold Tris buffered saline (TBS) and twice with FA lysis buffer (50mM

HEPES/KOH ph 7.5, 140mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate and 0.1% SDS) with freshly prepared proteinase inhibitors. Egg chambers were finally brought up in 500ul of FA lysis buffer and sonicated four times for 12 seconds each, with 2 minutes on ice between sonications. This generated DNA fragment ranging from 100 to 500 bp. Then samples were micro-centrifuged at 14,000 rpm for 10 min. at 4^{o} C, and supernatants were removed and diluted twice in FA lysis buffer. 500ul of lysates were pre-incubated with Dynabeads (Invitrogen) for 30 minutes at 4^{o} C to minimize background. Clearing beads were removed and replace with beads conjugated to antibody and the mixture was incubated overnight at 4C. Beads were then washed as previously described (Salma et al 2004). To reverse crosslinking, beads were suspended in 200 ul of PK buffer (200mM Tris-HCl, pH 7.5, 25 mM EDTA, 300mM NaCl, 2% SDS), 1 ul of 5M NaCl was added, and the mixture incubated at 65^{o} C for 4h . 200ml of water and 10ml of 20mg/ml proteinase K (Roche cat. # 3115887) were added and samples were incubated for 1 hr at 60^{o} C. The beads were then removed and DNA was phenol/chloroform extracted, ethanol precipitated, and resuspended in dH_2O. qPCR was performed in triplicate using primers specific to the indicted clusters or protein coding genes (Supplementary Table S6).

Strand specific RT-PCR

The total RNA was prepared from whole ovaries using RNeasy Mini Kit (Qiagen#74104). Plus or minus strand primers to the indicated regions of the piRNA clusters (Supplementary Table S6) were mixed an RP49-specific control primer and first-strand cDNAs were synthesized using SuperScript^{TM} III reverse transcriptase (Invitrogen#18080-093) following the manufacture's protocol. The resulting cDNAs were used as templates for quantitative real-time PCR using the primers indicated (Supplemental Table X). qPCR reactions were perfored using a 7500 Fast Real-time PCR system (Applied Biosystems) and SYBR Green I (Qiagen#204145). The expression level of strand specific RNA was measured relative to the RP49 internal control. Background obtained with no RT primer reactions was subtracted. No reverse transcriptase control reactions did not produce significant signal, indicating that DNA contamination was minimal. Three technical replicates were performed for each RT primer. Graphs show the average and standard deviation (SD).

Bioinformatics methods

Sequence extraction and annotation

For each sequence read, the first occurrence of the 6-mer or 7-mer perfectly matching the 5′ end of the 3′ linker was removed. The extracted inserts were then mapped to the female *Drosophila melanogaster* genome (Release R5.5, excluding chromosome YHet). Inserts that matched fully to a genomic sequence were collected, using an inhouse suffix tree-based software (Delcher et al., 1999; Gusfield, 1997) (Delcher et al., 2002). Sequences corresponding to pre-miRNAs or non-coding RNAs (ncRNAs) were identified using the same software. Genes and transposon annotations were retrieved from FlyBase (R5.5), and unannotated transposons identified using BLAST (Altschul et al., 1990) to query each transposon consensus sequence against the female genome, using an *e*-value cutoff of 10^{-10}.

Background-corrected sequence bias detection

For a set of sequences of various lengths, the frequency of each nucleotide at each position was computed as a foreground count matrix. The background frequencies at each position were computed by averaging over all possible *k*-mers (*k* = 23–29) in transposon sequences, strand-specifically (either sense or antisense), weighted by the length distribution of each foreground set. Binomial testing was performed for each nucleotide at each position in the foreground matrix, using the corresponding element in the background matrix as the probability parameter. The significance level of testing used was $\mathbf{a} = 2.2 \times 10^{-6}$. This stringency ($p = 0.001$) corrects for multiple testing (four nucleotides at 29 positions, in two orientations (sense and antisense) for two classes (total versus ping-pong pairs). Only the nucleotides significantly above background were displayed, with *y*-axis values corresponding to the relative frequency of foreground minus background. This background correction helps distinguish between an A or U bias that results from the inherent AU-richness of transposon sequences and A or U biases reflecting the mechanism of piRNA biogenesis. Sequences were analyzed as species and not weighted by their abundance (number of reads).

Consensus mapping of total piRNA

The 23–29 nt small RNAs from the total RNA sequence reads in each sample were mapped to the consensus sequence of each transposon. For each BLAST-hit transposon copy, all the ungapped segments of the consensus-hit sequence alignment were collected and used for coordinate conversion. The genomic coordinate of any insert that overlapped with an ungapped segment of a BLAST hit was converted to the position on the consensus sequence. Overlaps of an insert to multiple segments were corrected to avoid double-counting, as well as multiple hits on the same genomic position caused by repetitive sequences. An advantage of coordinate conversion over direct sequence matching to the consensus sequence is that it allows consistency in the small RNAs included in different analyses, i.e., the same set of the small RNAs mapped to transposon regions in the genome are included in the set mapped on the consensus sequence. For each piRNA, all the bases that cover that piRNA were counted as having the number of reads of that piRNA, normalized by the number of times they mapped to the genome and by the total abundance (number of reads) of ncRNAs in the sample.

Overlap analysis of total RNAs

For each genotype, all the piRNAs whose 5’ ends overlapped with another piRNA on the opposite strand were collected. For each length of overlap, the total number of reads were computed and normalized by the number of times the piRNA mapped to the genome and normalized to the abundance of ncRNA reads in that sample.

Chromosome and piRNA cluster density profiles

To generate chromosome profiles, we plotted the number of sense and anti-sense 23–29 nt small RNA reads within a 5 rolling window, as described by Brennecke et al. (2007).

Higher resolution piRNA cluster profiles were generated using a 100 bp sliding window with 10bp step size. Cluster coordinates are from Brennecke et al. (2007). Sequences mapping to the plus strand and minus strands were plotted separately on the top and bottom of the chromosome and cluster profiles.

Supplemental References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-10.

Austin, R., Orr-Weaver, T. and Bell, S. (1999). Drosophila ORC specifically binds to ACE3, an origin of DNA replication control element. *Genes and Development* **13**, 2639-49.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. and Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**, 1089-103.

Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, O. and Salzberg, S. L. (1999). Alignment of whole genomes. *Nucleic Acids Res* **27**, 2369-76.

Delcher, A. L., Phillippy, A., Carlton, J. and Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* **30**, 2478-83.

Gusfield, M. (1997). **Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology**. Cambridge, UK: Cambridge University Press.

Harlow, E. and Lane, D. (1999). Purifying Antibodies. In *Using Antibodies: A Laboratory Manual*, pp. 70-80: Cold Spring Harbor Laboratory Press.

Theurkauf, W. E. (1994). Immunofluorescence analysis of the cytoskeleton during oogenesis and early embryogenesis. *Methods Cell Biol* **44**, 489-505.

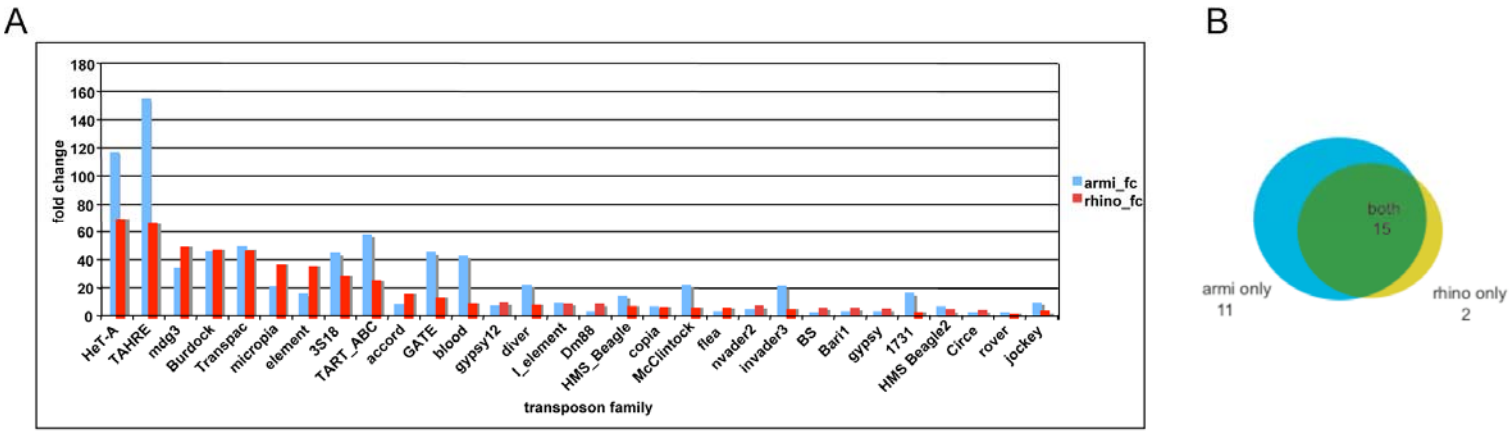


Figure S1. Transposon expression in *rhi* and *armi* mutants. A. Bar graph showing fold change in transposon family expression (relative to *w-1118*) in *rhi* (red) and *armi* (blue). The 30 most highly over-expressed families in *rhi* are shown. B. Venn diagram summarizing over-expressed families (FDR<0.02) in *armi* alone (blue), in *rhi* alone (yellow), and in both mutants (green). Number of elements in each class is indicated.

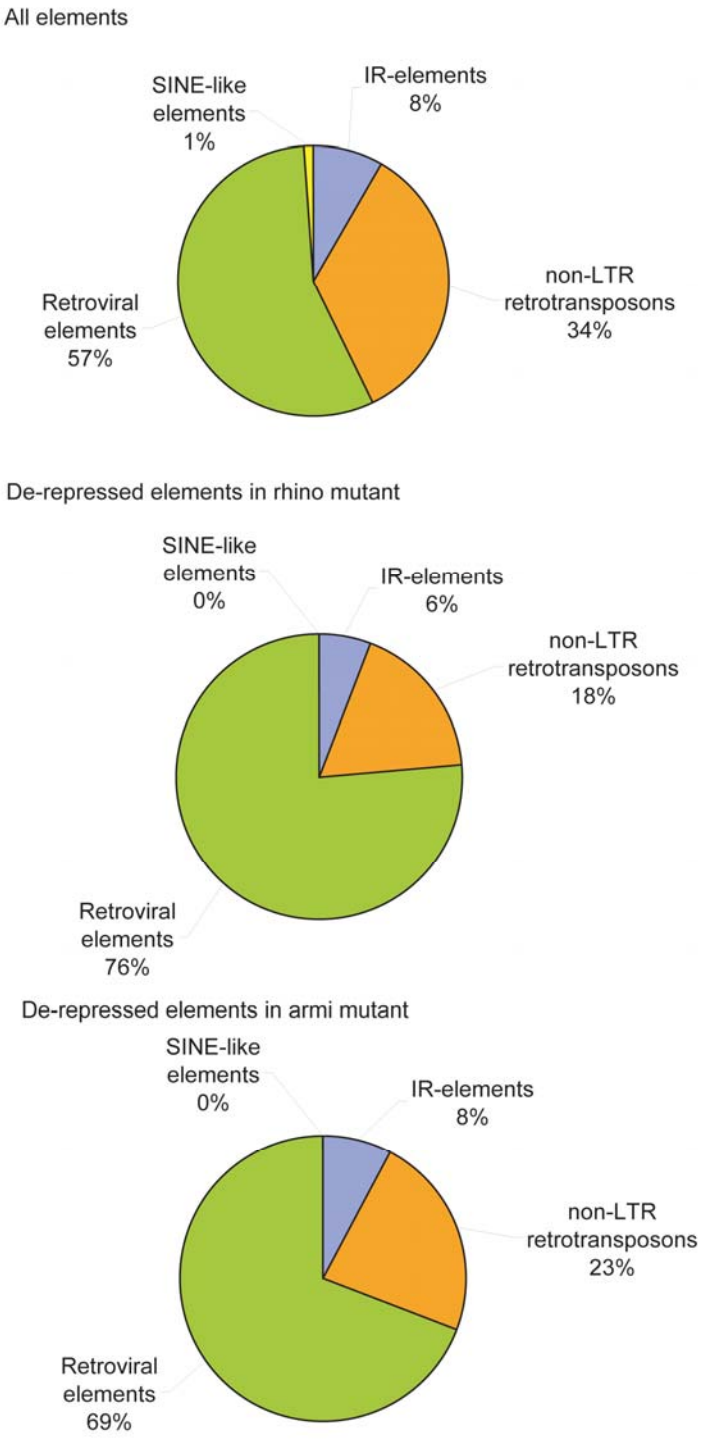


Figure S2. Class breakdown of over-expressed transposon families. Top. Pie chart showing the relative abundance of SINE-like (yellow), retroviral (green), Non-LTR (orange), and IR elements in the genome. Middle and bottom. Class breakdown of elements that are over-expressed in *rhi* and *armi* mutants, respectively.

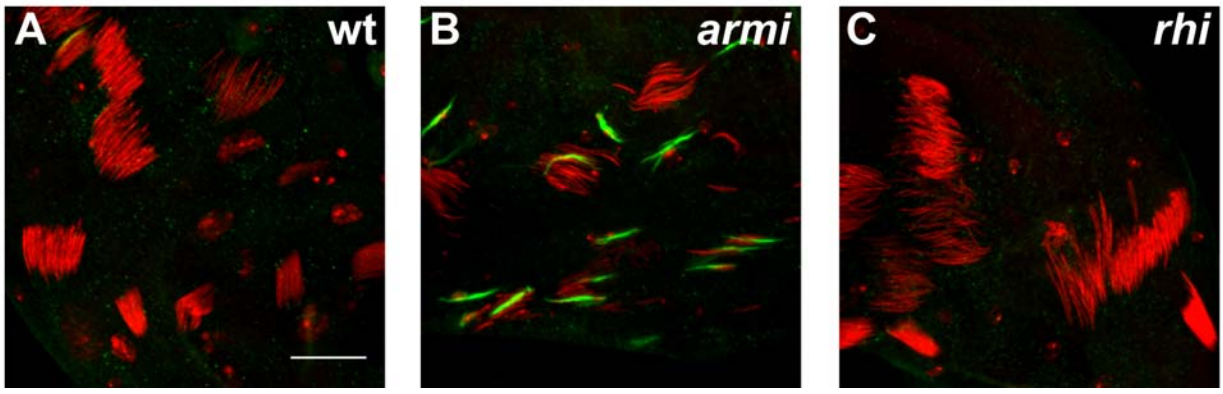


Figure S3. The *Stellate* locus is silenced in *rhi* mutant testes. The testes from (A) wild type, (B) *armi72.1/armi1*, (C) *rhiKG/rhi2* flies were stained for DNA (red) and Stellate protein (green). A and C. No Stellate protein is expressed in wild type and *rhiKG/rhi2* testes. B. In *armi72.1/armi1* mutant testes, Stellate protein accumulates and forms crystals. Projections of 5 serial 1mm optical sections are shown. Images were acquired under identical conditions. Scale bar is 20 mm.

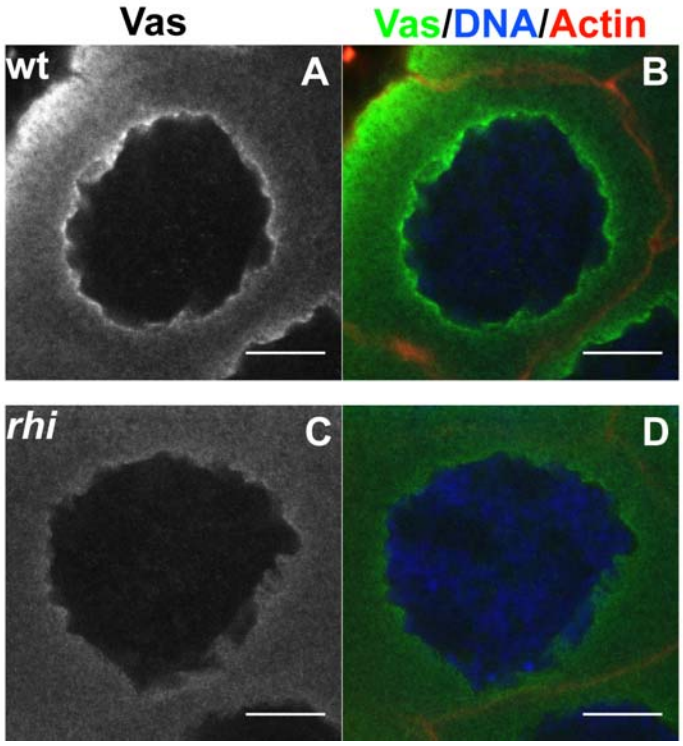


Figure S4. Localization of the nuage protein Vasa in *rhi* mutants.
(A, B). Vasa (Vas) localization to the perinuclear nuage in a wild type stage-6 egg chamber (A-B). (C, D). Vas is dispersed in the cytoplasm of a *rhi* mutant stage-6 egg chamber. Single optical sections are shown. Panel A and C show Vas alone. Panels B and D show a merge of Vas (green), Actin (red) and DNA (blue). Scale bar is 10 mm.

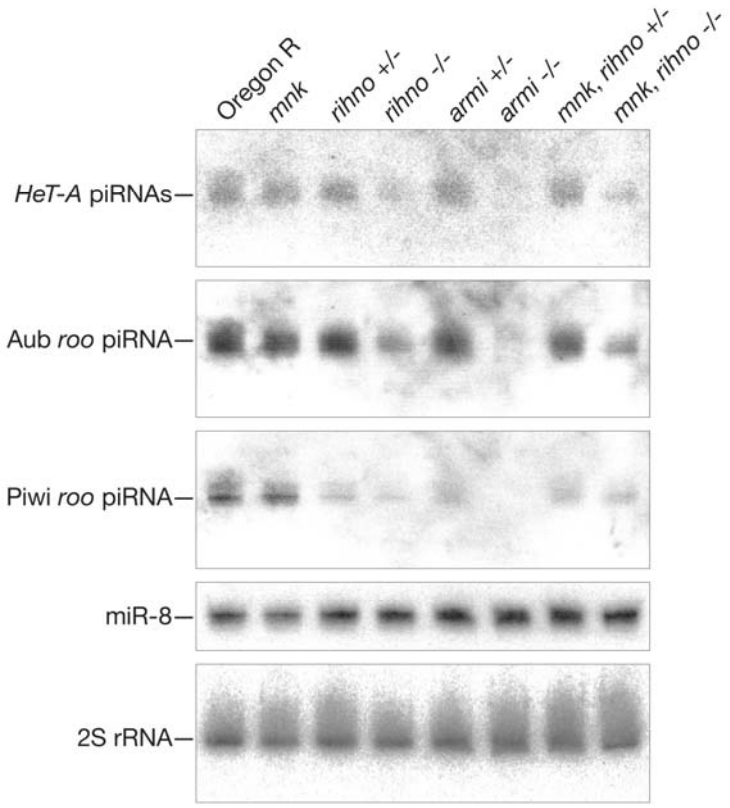


Figure S5. Northern blot analysis of piRNA and miRNA production in *rhi* mutants. Blots were probed for *HeT-A* piRNAs, miR8, and 2S RNA, as indicated. RNA was isolated from Oregon R, *rhi/+* , *rhi/rhi, armi/+*, *armi/armi,* and *mnk rhi/+* ovaries.

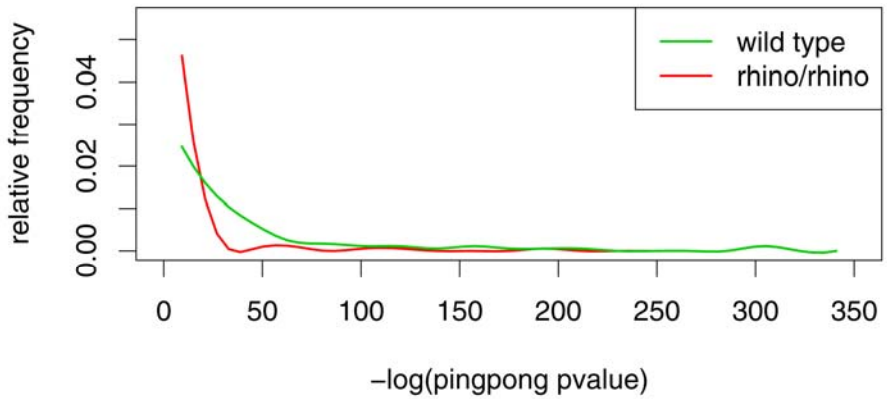


Figure S6. Frequency distribution of P-values for the 10 nt overlap bias for piRNAs derived from 95 transposon familes. The wild type distribution is in green and the distribution in *rhi* is in red. The decrease in ping-pong bias in *rhi* is highly significant (P-value $3e^{-10}$).

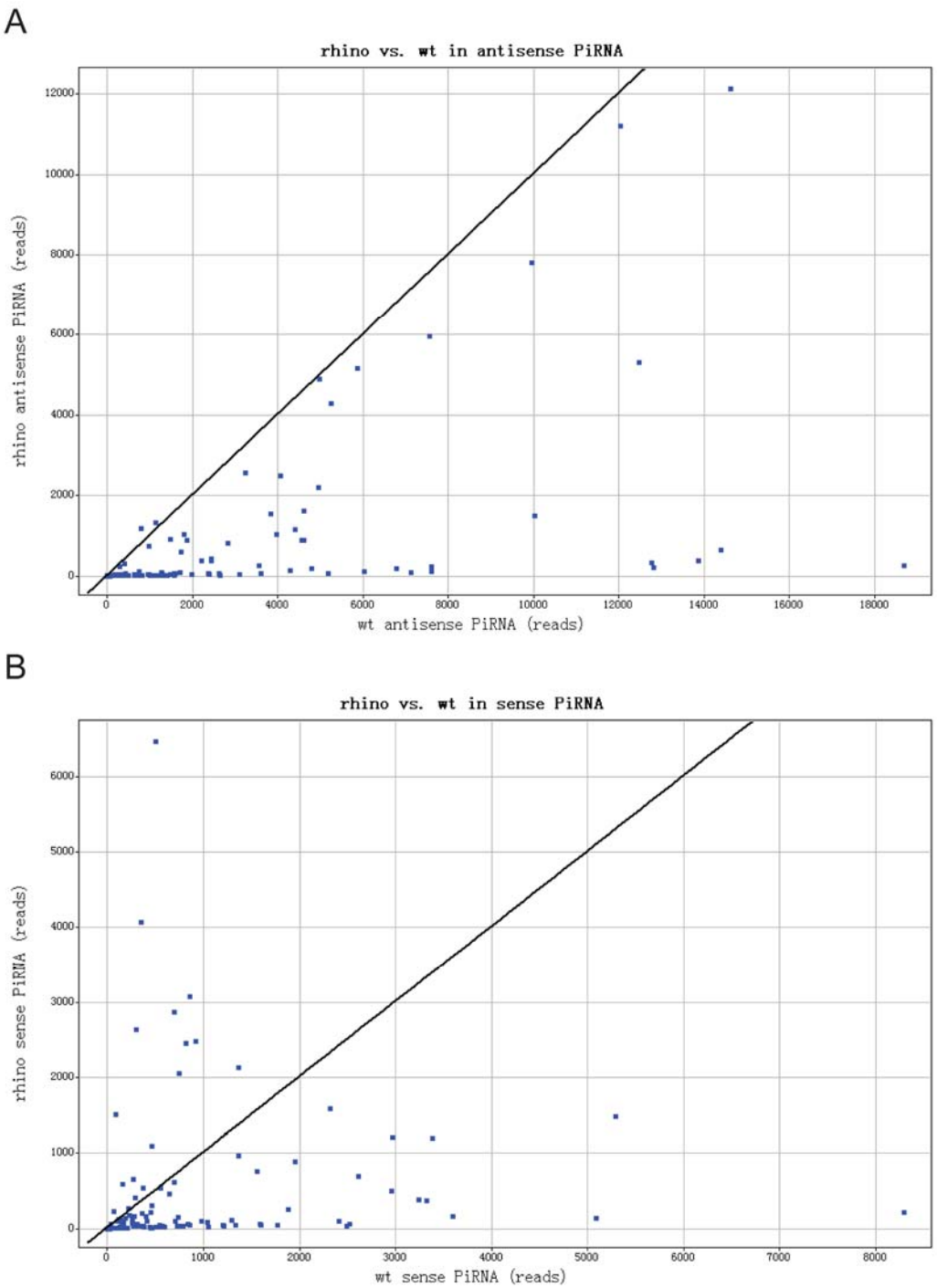


Figure S7. Sense and antisense piRNA abundance for 95 transposon families. (A). The abundance of antisense piRNAs from 95 transposon families in *rhi* mutants, plotted against the abundance in wild type controls. Each point represents one transposon family. (B) The abundance of sense piRNAs encoded by 95 transposon families in *rhi* mutants, plotted against abundance in wild type controls. *rhi* mutations reduce sense and antisense piRNAs encoded by the vast majority of elements. However, sense strand piRNAs from a limited number of families increases.

See separate PDF.

Figure S8. Detailed analysis of piRNAs encoded by major transposon families in wild type and *rhi* mutants. A and B. Sequence-bias plot for all piRNAs mapping to the transposon family (A) or for piRNAs with ping-pong partners (B). The graphs report only nucleotide biases significantly different from the intrinsic bias of the transposon family itself (see Supplemental Experimental Procedures). C and D. The abundance (per million) of piRNAs (23–29 nt) in wild type (*w*/+; *cn*/+) and *rhi* mutant (C) ovaries for each base-pair position of the transposon consensus sequence. Abundance was normalized by both the number of times a piRNA mapped to the genome and calibrated to the total abundance of sequences corresponding to non-coding RNAs in the sample. For each base, the total normalized abundance of piRNAs that fall on that position is reported. Here and below, blue denotes sense strand mapping piRNAs; red, antisense-strand mapping piRNAs. E and F. Length distribution of 19–29 nt RNAs mapping to the transposon family for *ago3*/TM6B (E) and Oregon R ovaries (F). The number of reads was normalized as in (C) and (D). G and H. Distribution of the length of 5 overlaps between sense and antisense piRNAs mapped to the transposon family for wild type (G) and *rhi* (H) ovaries. Only pairs that overlapped at their 5 ends were included. A peak at 10 bp indicates ping-pong pairing.

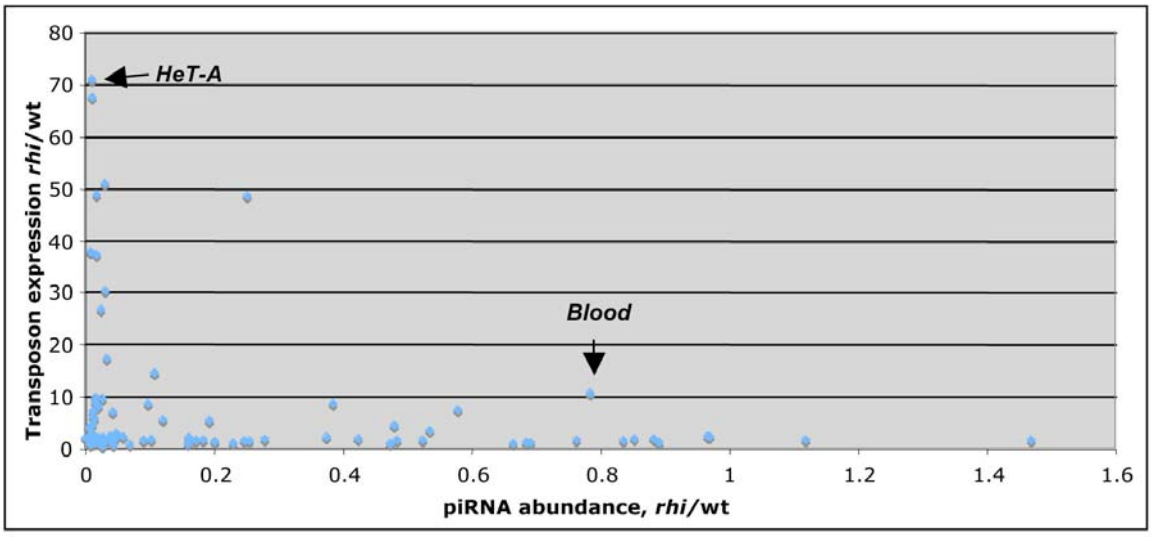
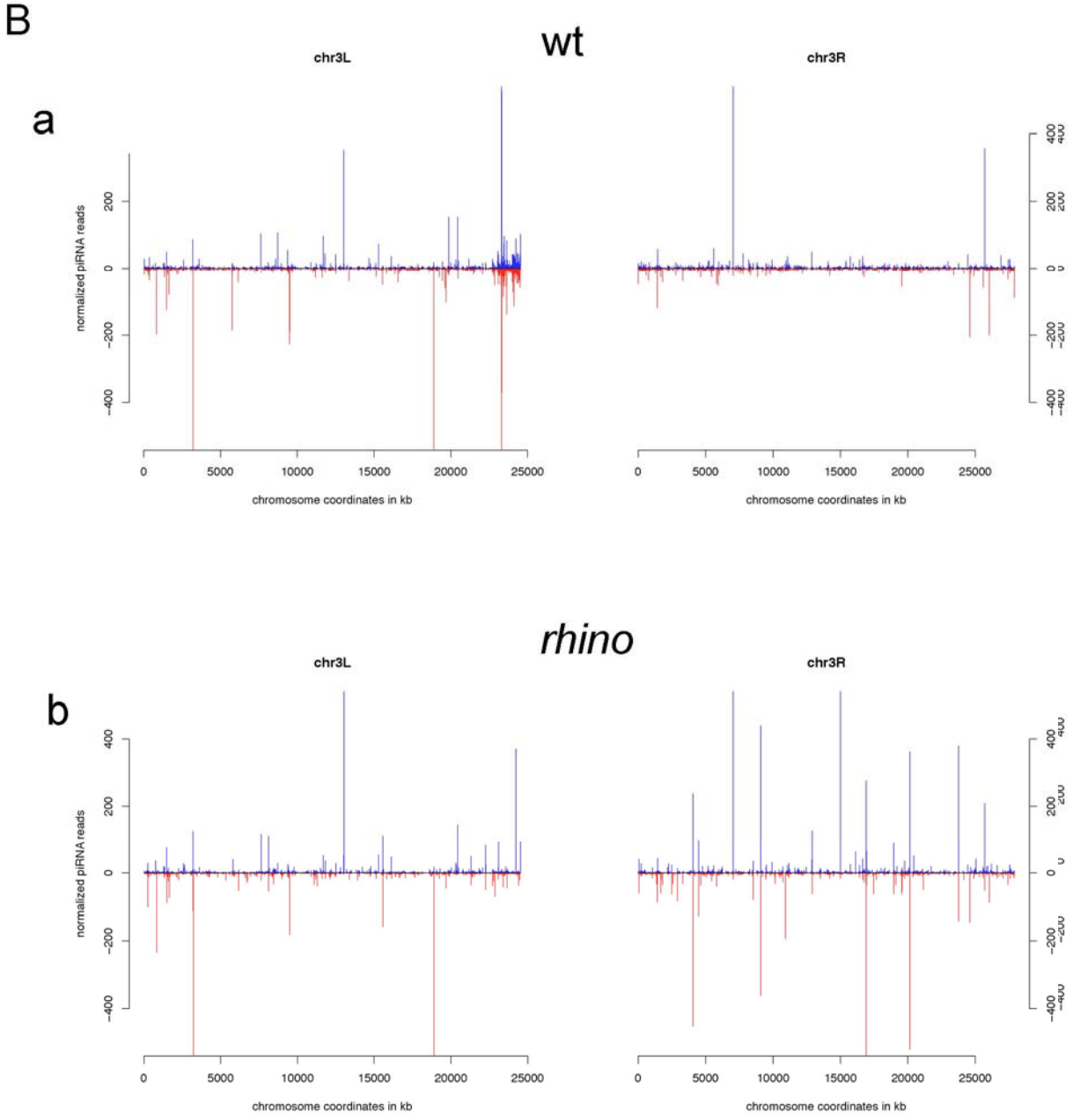


Figure S9. Transposon expression vs. piRNA abundance. A. The graph shows fold change in transposon family expression in *rhi* (y axis, *rhi/ w-1118*) plotted against fold change in antisense piRNAs encoded by the same transposon family (x axis, *rhi/w-1118*). The majority of highly over-expressed elements also show significant reductions in antisense piRNAs (*Het-A*). However, expression of *blood* increase 10 fold while the total antisense piRNA pool is reduced by approximately 20%.

B


wt
a
chr3L
chr3R
normalized piRNA reads
chromosome coordinates in kb
rhino
b
chr3L
chr3R
normalized piRNA reads
chromosome coordinates in kb

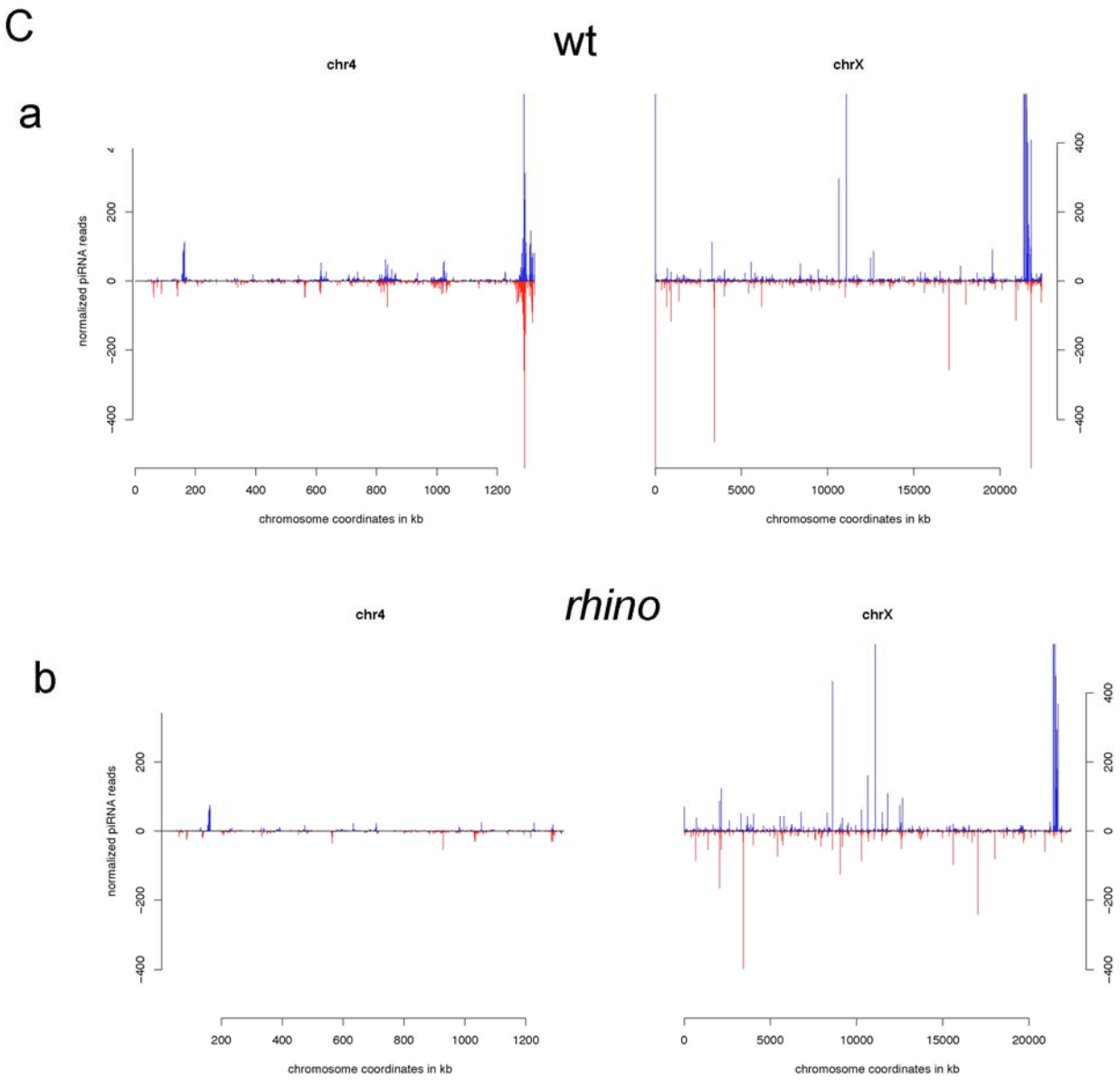


Figure S10. Chromosome profiles of uniquely mapping piRNAs. A. Chromosome 2 profiles of uniquely mapping piRNAs in wt (a) and *rhi* (b). B. Chromosome 3 profiles of uniquely mapping piRNAs in wt (a) and *rhi* (b). C. Chromosome 4 and X profiles of uniquely mapping piRNAs in wt (a) and *rhi* (b). The number of sense and antisense piRNA reads within a rolling 5 kb window are shown as a function of position on the major chromosome arms. Mutations in *rhi* reduce piRNAs mapping to pericentromeric heterochromatin on all of the autosomes. By contrast, piRNAs mapping to X-linked pericentromeric heterochromatin are retained.

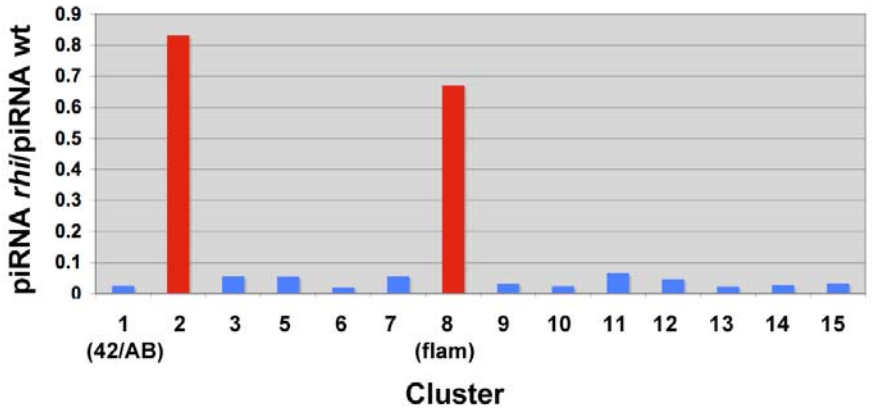


Figure S11. piRNA production by uni-strand and dual-strand clusters. Blue bars represent the ratio of normalized piRNA reads in *rhi*/wt for the indicated dual-strand clusters. Red bars represent the ratio of normalized piRNA reads in *rhi*/wt for the indicated uni-strand clusters.

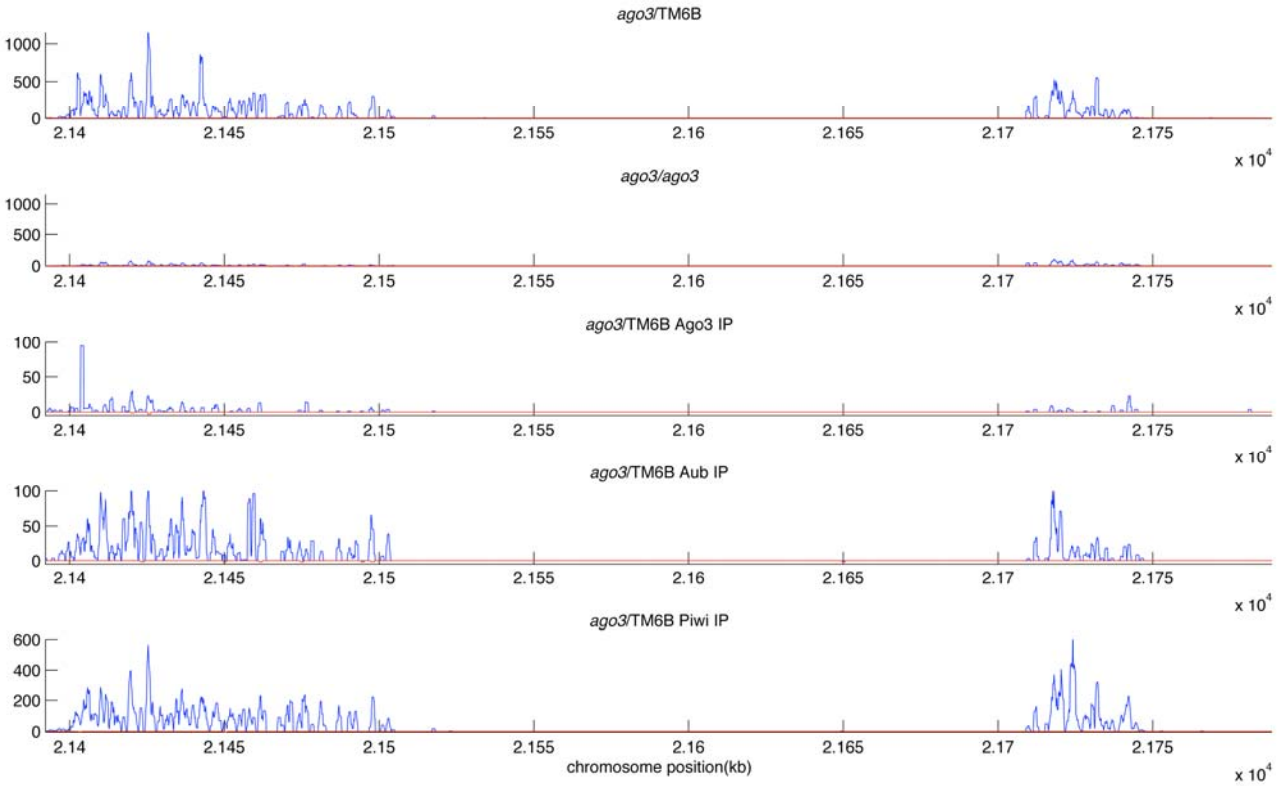


Figure S12. Cluster 2 piRNA profiles. Total piRNAs mapping to cluster 2 are dramatically reduced in *ago3/ago3* mutants, relative to *ago3*/TM6B controls. piRNAs that immunoprecipitated with Ago3, Aub and Piwi map to cluster 2. Note that the scales are different, and that the majority of piRNAs mapping to this cluster bind to Piwi.

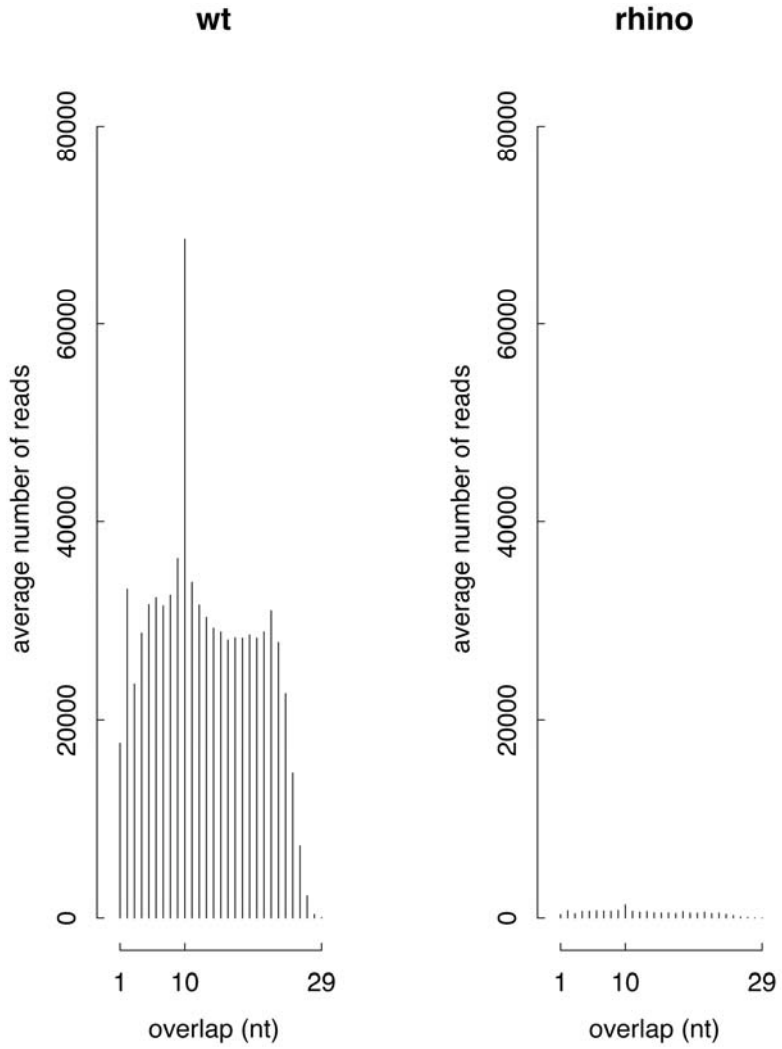


Figure S13. Ping-pong amplification of cluster 1/42AB piRNAs. Histogram of overlapping unique sense and antisense piRNAs mapping to cluster 1/42AB. In wild type ovaries, piRNAs derived from this cluster show a pronounced peak at 10 nt, characteristic of ping-pong amplification. The 10 nt peak is nearly eliminated in *rhi* mutants.

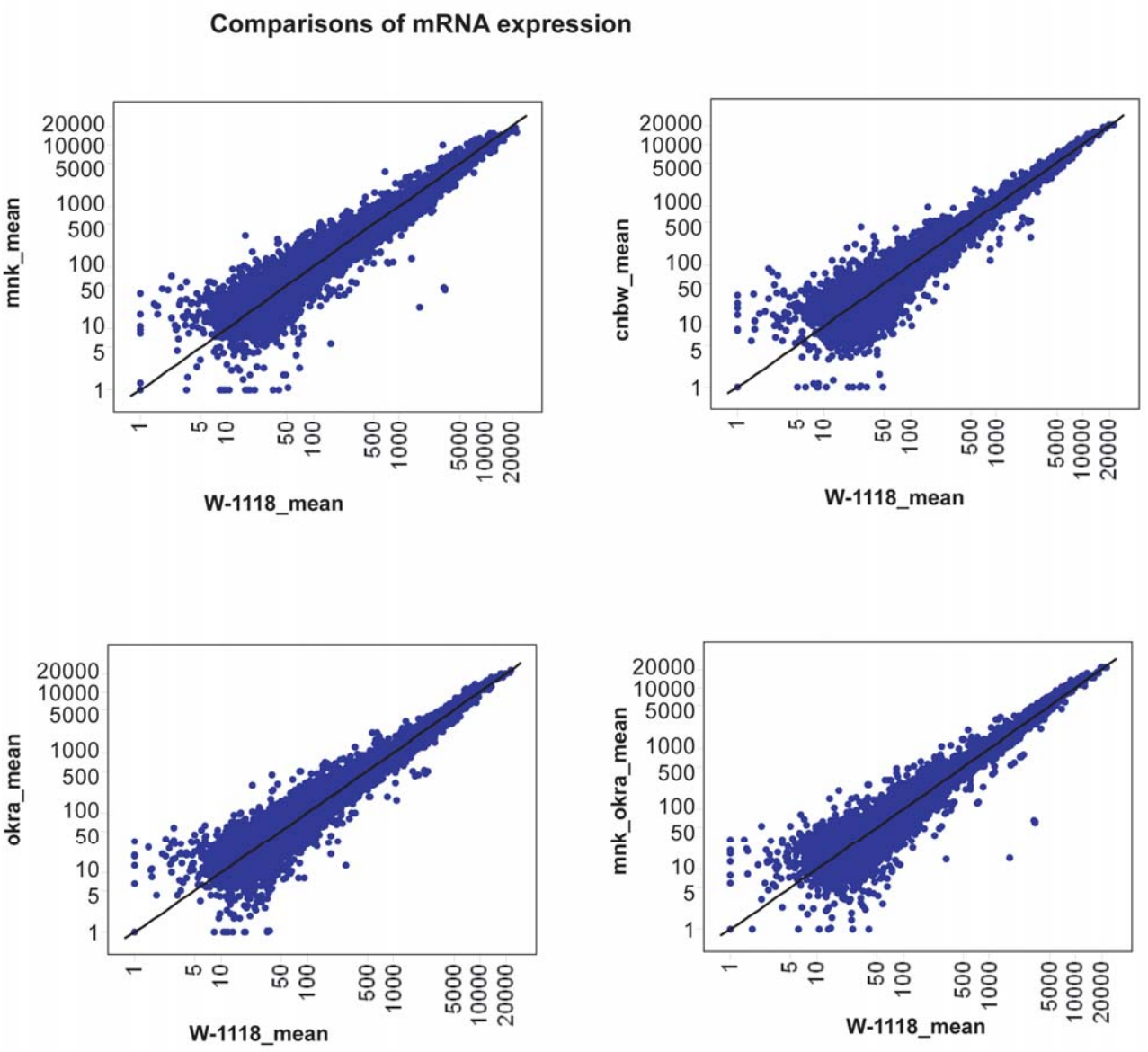


Figure S14. Transposon and protein coding gene expression in control genotypes. Tiling arrays were performed on total ovarian RNA isolated from females with two common background genotypes (*w-1118* and *cn,bw*), homozygous mutant for meiotic repair mutation *okra*, homozygous for the DNA checkpoint mutation *mnk,* and double mutant for *okra* and *mnk.* A. Log2 protein coding gene expression in the indicted genotypes plotted against Log2 protein coding gene expression in *w-1118.* B. Log2 transposon family expression in the indicated genotypes plotted against Log2 transposon family expression in *w-1118.* Log2 expression is plotted.

Table S1. *mnk* , *mei-41*, and caffeine suppress D-V patterning defects in *rhi* mutants. Genotypes are as indicated. For caffeine treatment, flies were fed yeast paste made with a 2% w/v solution.

Maternal genotype	**2 (wild type)**	**1 (fused)**	**0 (absent)**	**Hatch Rate(%)**	**n**
mnkP6 / mnkP6	93.6	2.5	3.9	72.6	827
mei41D3 / mei41D3	100	0	0	0	920
meiW681 / meiW68K05603	94.3	4	1.7	67.2	1281
rhi02086 / rhiKG00910	17.2	66.5	16.3	0	700
mnkP6 rhi02086 / mnkP6 rhiKG00910	80.5	14.5	5	0	689
mei41D3 / mei41D3 ; rhi02086 / rhiKG00910	32.9	53.3	13.8	0	732
meiW68K05603 rhi02086 / meiW681 rhiKG00910	18.7	58.5	22.8	0	244
rhi02086 / rhiKG00910 2% caffeine	88.4	6.6	5	0	473
armi72.1 / armi1	1.8	24.7	73.6	0	227
armi72.1 / armi1 2% caffeine	10.6	29.6	59.8	0	477
mei41D3 / mei41D3 ; armi72.1 / armi1	56	38.4	5.6	0	575
mei41D3 / mei41D3 ; armi72.1 / armi1 2% Caffeine	83.2	11.9	4.9	0	226
rhi02086 / rhiKG00910; GFP-Rhino/Gal4 nos	96.1	3.6	0.3	74	483

Table S3. Sequencing statistics. miRNAs were depleted from total RNA by oxidation "Total small RNA species" correspond to genome-matching species after excluding annotated non-coding RNAs (ncRNAs), such as rRNA, snRNA, or snoRNA. "Transposon-matching species" correspond to small RNAs mapped to *Drosophilia melanogaster* transposons.

Genotype	Total species	Genome-matching species	species matching annotated ncRNAs	Small RNA species (excluding ncRNAs)	Pre-miRNA–matching species	Small RNA species excluding pre-miRNA–matching species	23–29 nt small RNA species			
							Total	Transposon-matching		
								Total	Sense	Antisense
wild-type	1,303,753	855,559	22,303	833,256	568	832,688	605,723	438,722	141,983	300,206
rhino	1,140,120	579,207	31,153	548,054	537	547,517	406,087	216,694	74,697	143,490

Table S5. Localization of Rhi-independent transposon families to the major uni-strand clusters. Column A. Transposon family. Column B. Fold change in antisense strand piRNAs (*rhi*/wt). Column C. Presence of annotated copies in cluster 2 (2) or *flamenco* (*flam*).

Transposon	*Fold change anti-sense piRNA*	*Cluster 2/flamenco?*
Stalker3	1.47	2
gypsy5	1.12	flam
Tabor	0.97	
Stalker2	0.97	flam
297	0.89	2
Stalker4	0.88	2, flam
Stalker	0.85	2
mdg1	0.84	2
blood	0.79	2, flam
412	0.76	2,flam

Table S6. qPCR and RT-PCR primers

RT primers

name	genomic coordinates	sequence
cl1-A-rt-plus	2R:2210762..2210782	CGAAGCCTTAGATCTCGCTCC
cl1-A-rt-minus	2R:2210578..2210599	ACATCAGGAACACAGCGAGGTG
cl1-32-rt-plus	2R:2337054..2337080	GGTGCAAATGTCTCATCATAATCAGTC
cl1-32-rt-minus	2R:2336830..2336858	GATGAAATTGAATTTCGTGATGACAGATC
cl2-A-rt-plus	X:21394981..21394991	CGTGGGTCCAG
cl2-A-rt-minus	X:21394876..21394888	TGAAAATCGCATC
flam-rt-plus	X:21507389..21507400	TACGCTAGTCCG
flam-rt-minus	X:21507136..21507149	TATGAGGATCAGAC
rp49-rt	3R:25871130..25871140	GGAGGAGACGCCG

qPCR primers

name	coordinate	sequence
cl1-A-f	2R:2210633..2210652	CGTCCCAGCCTACCTAGTCA
cl1-A-r	2R:2210705..2210724	ACTTCCCGGTGAAGACTCCT
cl1-B-f	2R:2193107..2193126	GCAGATGAGCTGAAACGAAA
cl1-B-r	2R:2193166..2193187	TCGCAGTCGTGTAATCCAAA
cl1-C-f	2R:2325464..2325483	CGCTGTTGAAAGCAAATTGA
cl1-C-r	2R:2325544..2325563	GAGACCTTCGCTCCAGTGTC
cl1-32-f	2R:2336913..2336932	GTGGAGTTTGGTGCAGAAGC
cl1-32-r	2R:2337004..2337023	AGCCGTGCTTTATGCTTTAC
jing-f	2R:2389861..2389880	CTTACCCGTTTCGGTTTTGA
jing-r	2R:2389934..2389953	ACACCCGAAATTCATTCTGC
pld-f	2R:2136668..2136687	ATGGCTGGGAAACTGACAAC
pld-r	2R:2136771..2136792	TTGCGGCATGTGTTATTTGT
cl2-A-f	X:21394901..21394920	GCCTACGCAGAGGCCTAAGT
cl2-A-r	X:21394956..21394975	CAGATGTGGTCCAGTTGTGC
cl2-B-f	X:21402737..214028756	CTGCTTTGTGCTTGGAGATG
cl2-B-r	X:21402808..21402830	TCTGCACAGATTCTGAAATTGAA
flam-f	X:21507195..21507215	TGAGGAATGAATCGCTTTGAA
flam-r	X:21507312..21507339	TGGTGAAATACCAAAGTCTTGGGTCAAC
ry-f	3R:8860469..8860497	TCTTCACTGGCTATCGCAGGAATGTTATC
ry-r	3R:8860648..8860675	CACCAAAAGCCATCGAGATCTCCGCCAC
rp49-f	3R:25871281..25871302	CCGCTTCAAGGGACAGTATCTG
rp49-r	3R:25871147..25871165	ATCTCGCCGCAGTAAACGC