Molecular Cell, Volume *50*

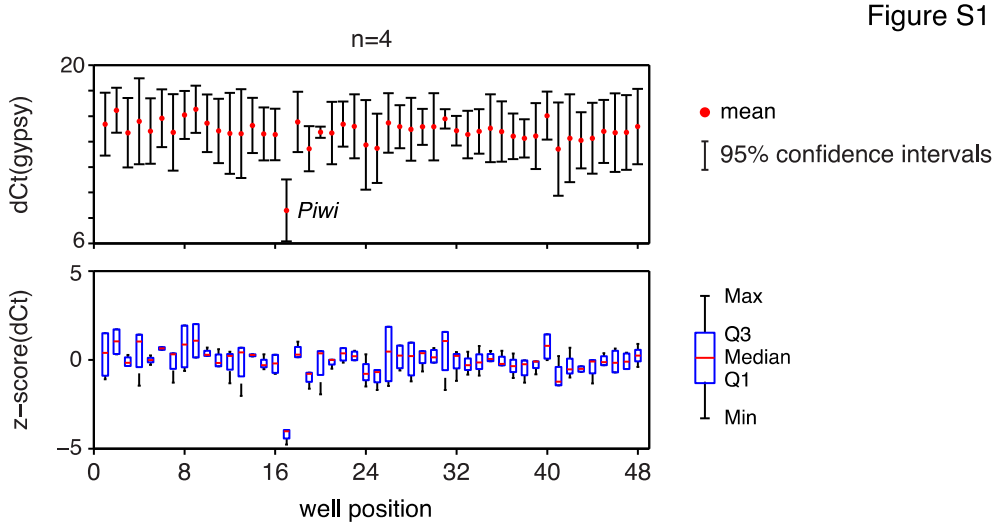
Supplemental Information

A Genome-wide RNAi Screen Draws a Genetic Framework for Transposon Control and Primary piRNA Biogenesis in *Drosophila*

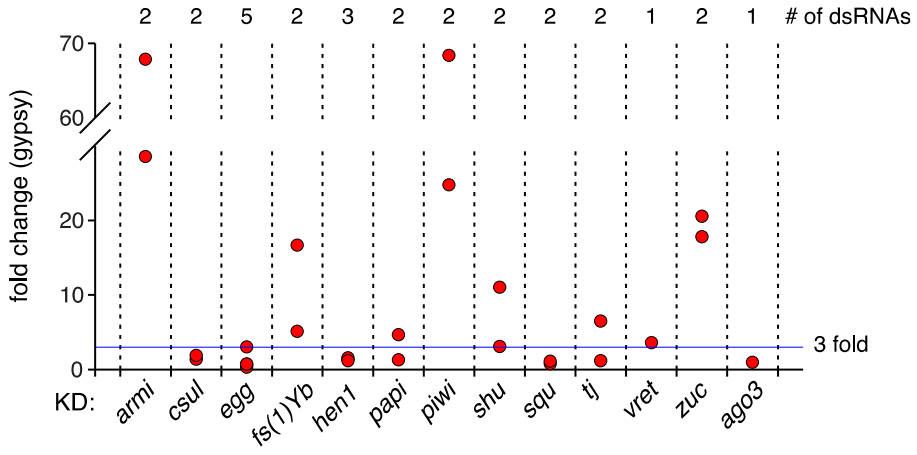
Felix Muerdter, Paloma M. Guzzardo, Jesse Gillis, Yicheng Luo, Yang Yu, Caifu Chen, Richard Fekete, and Gregory J. Hannon

Figure S1

A



B



C

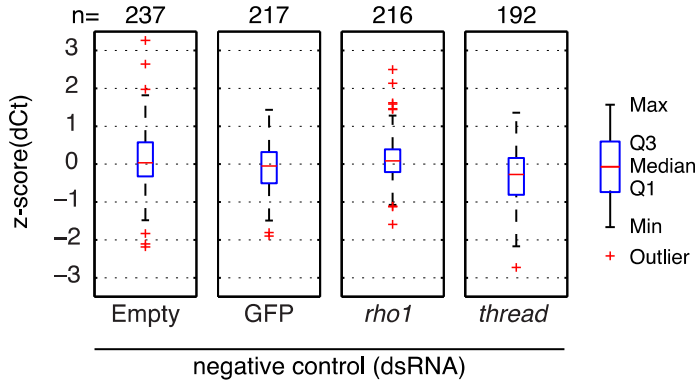


Figure S1. Performance and Controls for the Primary Screen, Related to Figure 1

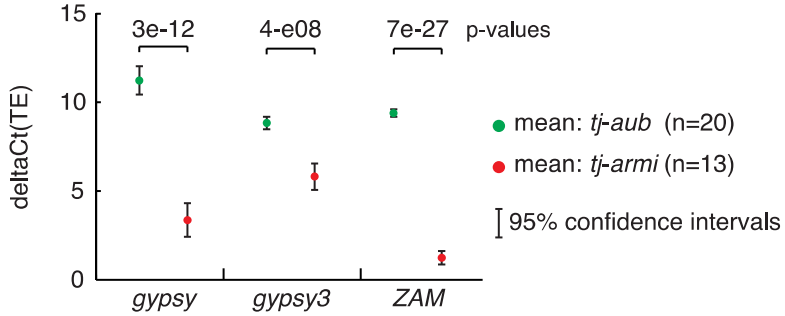
A) For 48 genes, including the positive control *piwi*, dsRNAs were transfected in four independent biological replicates. The upper graph shows the means and 95% confidence intervals of *gypsy* levels relative to a reference gene. The lower graph shows the individual z-scores as box plots for all 48 wells after normalization to the median of the plate. Piwi is a clear outlier in all four independent experiments.

B) The primary screen results in significant fold changes for all known somatic piRNA pathway components. Ago3 is shown as a negative control. The number of independent dsRNAs against each gene is indicated on top of the graph. The threshold for primary hit selection (3 fold up-regulation of *gypsy*) is marked in blue.

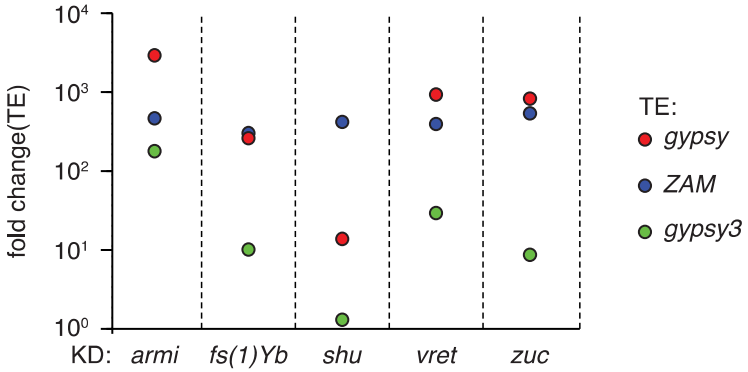
C) z-scores for 862 negative controls in the primary screen are shown as boxplots. Outliers are indicated as red crosses. The number of independent transfections of each dsRNA is indicated above.

Figure S2

A



B



C

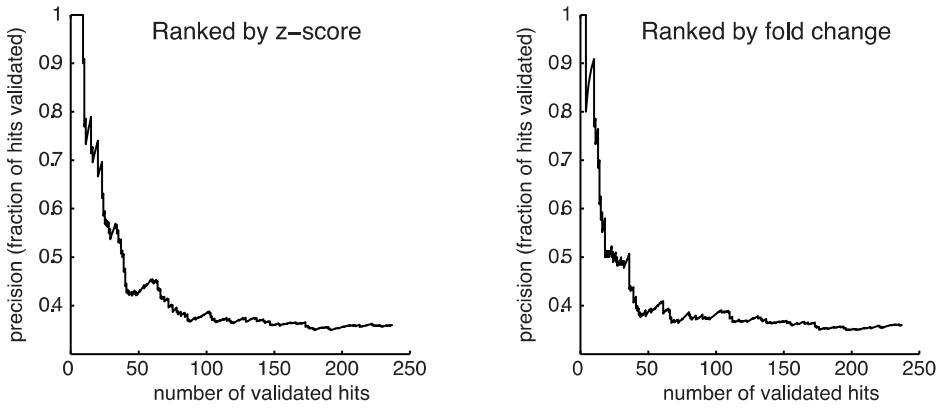


Figure S2. Performance and Controls for the Validation Screen, Related to Figure 2

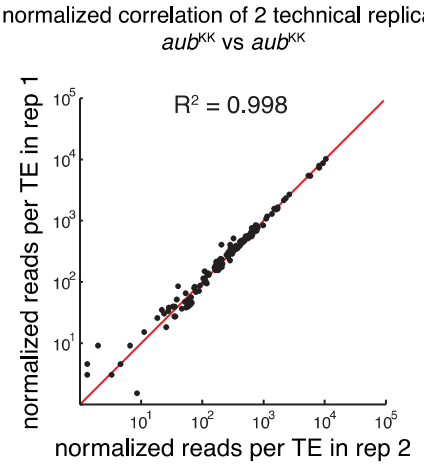
A) Knockdown of *armi* leads to highly significant differences in transposon expression when compared to a negative control (Aub). Shown are mean delta Ct values and 95% confidence intervals for three transposons assayed by qPCR. The number of biological replicates is indicates in brackets. The results of a t-test for significance are indicated as p-values for each transposon.

B) Knockdown of each component of the somatic piRNA pathway, which scored in the primary screen, has strong effects on the expression levels of three transposons *in vivo*. The fold change for each transposon upon knockdown is displayed on a log scale.

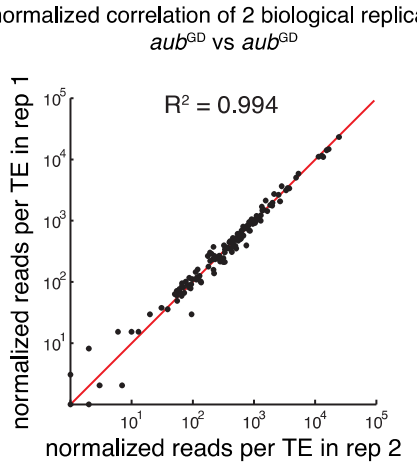
C) Z-scores and fold changes are a function of precision. Precision is the fraction of validated hits out of the total number of hits (validated and non-validated). The number of validated hits is shown on the x-axis. All dsRNAs for validated genes were used to cover the depicted range. Thus, if genes had dsRNAs producing z-scores or fold changes outside the range needed for primary hit selection, the genes' final annotation as validated or non-validated was assigned to those dsRNAs.

Figure S3

A



B



C

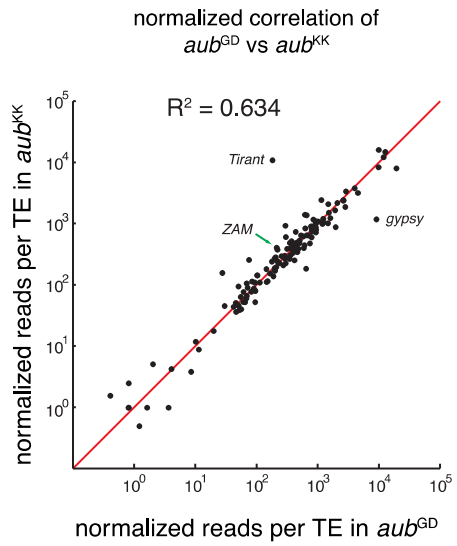


Figure S3. Two Negative Control Lines from Two Available VDRC Fly Libraries Show Different Transposon Expression Levels for Gypsy and Tirant, Related to Figure 3

A) Scatter plots of normalized reads mapping to TE consensus from RNA-seq data are shown. The squared correlation coefficient for two technical replicates of the KK line is indicated. The red line indicates where data points would show equal numbers in both samples.

B) The results for two biological replicates of *aub* flies from the GD library are shown.

C) Two Aub hpRNA lines from the KK and the GD library are compared. Data points for three transposons are highlighted: *gypsy* and *Tirant* as significantly differentially expressed and *ZAM* as the transposon used for hit calling in the validation screen.

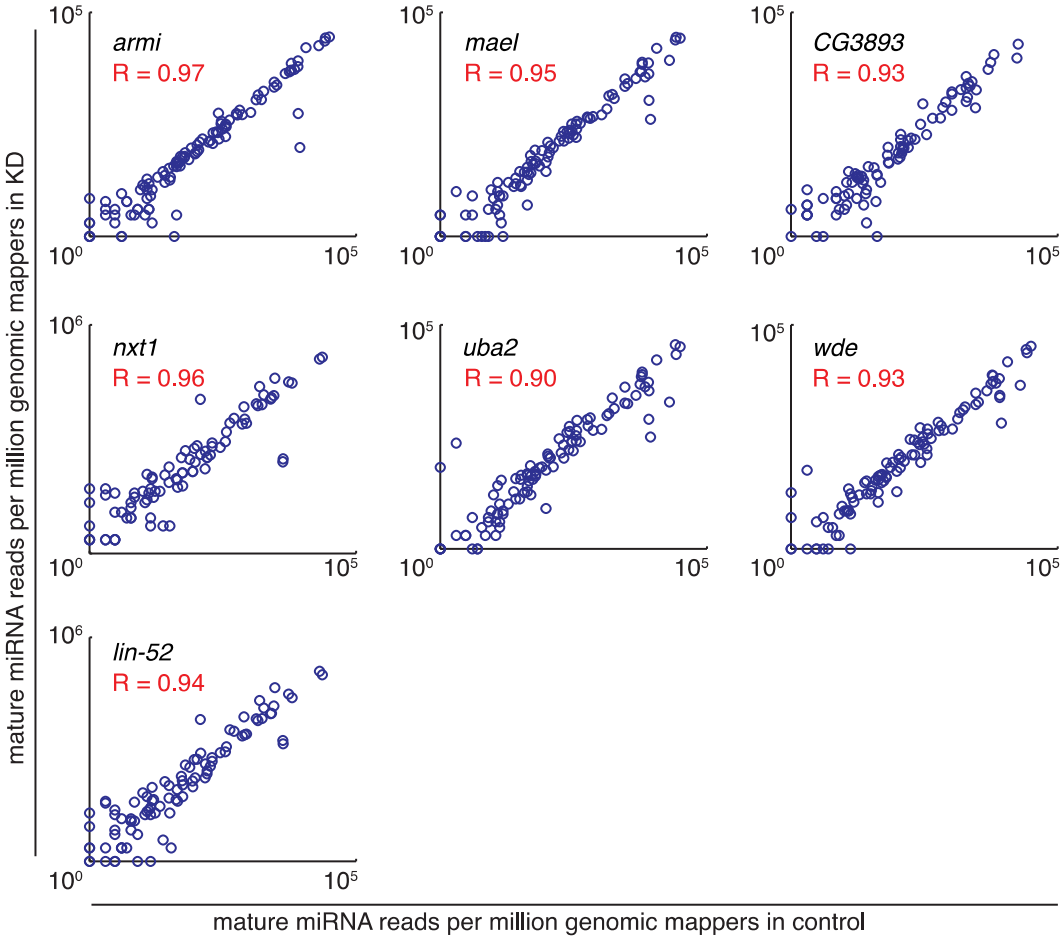


Figure S4. miRNA Populations in Knockdowns, Related to Figure 4

There are no significant changes in microRNA levels upon knock down of a number of validated hits. Scatter plots for mature microRNA in reads per million genomic mappers are shown for all follow-up genes compared to the Aub negative control. Pearson correlation coefficients are shown for each population.

Figure S5

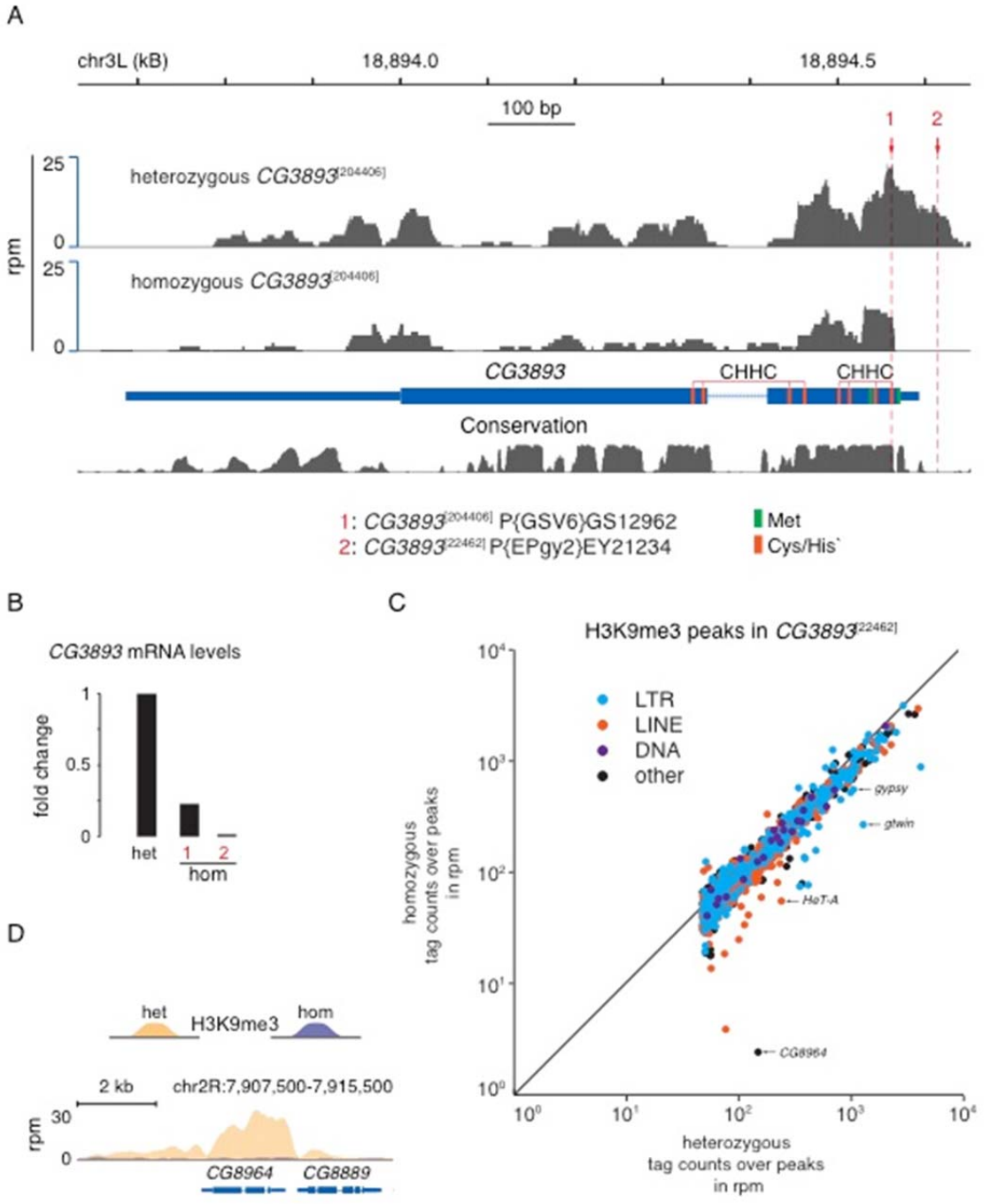

A
chr3L (kB)
18,894.0
18,894.5
100 bp
1
2
heterozygous CG3893[204406]
homozygous CG3893[204406]
rpm
25
0
CG3893
CHHC
CHHC
Conservation
1: CG3893[204406] P{GSV6}GS12962
2: CG3893[22462] P{EPgy2}EY21234
Met
Cys/His
B
CG3893 mRNA levels
fold change
1
0.5
0
het
1
2
hom
C
H3K9me3 peaks in CG3893[22462]
LTR
LINE
DNA
other
gypsy
gtwin
HeT-A
CG8964
homozygous tag counts over peaks in rpm
heterozygous tag counts over peaks in rpm
D
het
H3K9me3
hom
2 kb
chr2R:7,907,500-7,915,500
rpm
30
0
CG8964
CG8889

Figure S5. Two P Element Insertions Disrupt CG3893 Function, Related to Figure 5

A) Density plots of reads mapping to *CG3893* from RNA-seq libraries corresponding to the heterozygous and homozygous *CG3893*$^{[204406]}$ insertion line are shown. 1 and 2 shown in red designate the insertion points of P-elements in *CG3893*$^{[204406]}$ and *CG3893*$^{[22462]}$, respectively. Beneath, the FlyBase gene model for *CG3893* is shown with green boxes designating Met translation start sites and in red boxes are positions of Cys and His amino acids that make up the CHHC zinc fingers. Under the gene model, conservation is shown.

B) Both *CG3893* P-element insertion lines disrupt expression of its mRNA transcript but to different extents. qPCR for levels of *CG3893* transcript in heterozygous and homozygous flies are shown. Each homozygous fly is normalized to its corresponding heterozygous sibling. 1 corresponds to *CG3893*$^{[204406]}$ and 2 to *CG3893*$^{[22462]}$.

C) Read densities over a subset of H3K9me3 peaks identified in *CG3893*$^{[204406]}$ heterozygous flies are lower in homozygous siblings. The read densities are expressed as reads per million genomic mappers. All peaks are annotated and divided into four annotation categories: transposable elements (LTR, LINE and DNA elements) and 'other' (intra- and intergenic annotations). The three highlighted transposon peaks correspond to the plots shown in Figure 5H. A detailed view of the genic peak is shown in Figure S5D).

D) H3K9me3 peaks over genic regions are affected in *CG3893*$^{[204406]}$ homozygous flies. Read densities are expressed as reads per million genomic mappers (rpm).

Table S2. Top Enriched GO Terms in the Primary Screen, Related to Figure 1

GO term	p-value[1]	GO ID
Yb body	0	GO:0070725
negative regulation of growth of symbiont in host	0	GO:0044130
stem cell development	0	GO:0048864
negative regulation of multi-organism process	0	GO:0043901
positive regulation of Ras protein signal transduction	0	GO:0046579
dorsal appendage formation	0	GO:0046843
germ-line stem cell maintenance	8.E-10	GO:0030718
regulation of mRNA 3' end processing	6.E-09	GO:0031440
male germ-line stem cell division	6.E-09	GO:0048133
gene silencing by RNA	2.E-08	GO:0031047
negative regulation of transposition	2.E-07	GO:0010529
imaginal disc-derived wing expansion	3.E-07	GO:0048526

[1]after multiple test correction

Table S4. Primers and Probes Used in this Study, Related to Supplemental Experimental Procedures

Target	Sequence (5'-3')	Primer:Probe
Primary Screen		
gypsy fwd.	CCAACAATCTGAACCCACCAATCTA	
gypsy rev	AGTACCCGCCACAACCTTTAAG	
gypsy probe	CAAACAGGGTAGTTAAGTTAG	4.5:1
cib fwd.	GCCAGCATCCCAGCTTAGTAGT	
cib rev	GCTGGGGCGGCCATCTT	
cib probe	CGCTTCGCCAATCCA	1.5:1
Validation Screen		
gypsy fwd.	CAGGCGACAAACAGGGTAG	
gypsy rev	GTTCAAACACCAGCACATCC	
gypsy probe	AC ACAGGAATGTAGTTGGCATGCGA	4:1
gypsy3 fwd.	GACATACTGAAGGGCGAGAAC	
gypsy3 rev	TCAGGGTATCTAAGGGTGACG	
gypsy3 probe	CAAGGTAGAATTTTCCGAAGCGCAGC	4:1
ZAM fwd.	GGTATGGAAGATGTGGGTGTC	
ZAM rev	TCCTCTTCACCGTATCCCTAG	
ZAM probe	TCGCCGTAATACTCACCTGGACACT	4:1
rp49 fwd.	GTCGGATCGATATGCTAAGCTG	
rp49 rev	CAGATACTGTCCCTTGAAGCG	
rp49 probe	TTGTCGATACCCTTGGGCTTGCG	1:1
General qPCR primers and probes		
CG3893 fwd.	TCGTCATCCCAGTTCTCCT	
CG3893 rev	CATTTGATACCAGAGCCCCAG	
CG3893 probe	CGAAGACACCAGACACGCGAAGAT	1:1
2s-rRNA depletion antisense oligo		
2s-rRNA	AGTCTTACAACCCTCAACCA TATGTAGTCCAAGCAGCACT	

Supplemental Experimental Procedures

Cell Culture

OSS cell were cultured in Shields and Sang M3 Insect media (Sigma) supplemented with 10% FBS, 5% fly extract, 0.6mg/ml glutathione and 10mg/ml insulin as previously described (Niki et al., 2006). Cells were transfected using Xfect transfection reagent according to manufacturer's guidelines (Clontech, #631317).

DNA Plasmids

Expression vectors of CG3893:GFP, RFP:Piwi and RFP:ΔNTPiwi driven by an ubiquitin promoter were made using the Drosophila Gateway Collection (Terence Murphy, Carnegie Institute of Washington, Baltimore, MD). To construct expression clones, coding sequences of CG3893, Piwi and ΔNTPiwi (excluding the first 72 aa) were PCR-amplified from ovarian cDNA and cloned into pENTR/ D-TOPO, and then recombined with either destination vector pURW (DGRC1282), for Piwi and ΔNTPiwi or pUWG (DGRC 1284), for CG3893.

Imaging of Fluorescent Fusion Proteins in OSS

OSS cells were co-transfected with plasmids expressing the indicated fusion proteins using Cell Line Nucleofector kit V (Amaxa Biosystems; program T-029). 48 hours after transfection, cells were plated on glass coverslips. 24 hours later, cells were stained with Hoechst 33342 (NucBlue live cell stain; Invitrogen, R37601) and immediately fixed in 2% formaldehyde/PBS at room temperature for 5min. After three 10min PBS washes, coverslips were mounted in proLong antifade (Invitrogen, P7481) and examined under a fluorescent microscope (Nikon Eclipse Ti). Z-stack images were taken with 40X magnification and the final images were de-convoluted under the default manufacturer settings.

RNAi Libraries

Two Drosophila dsRNA libraries were used in this study, the Open Biosystems (now Thermo Scientific) Drosophila RNAi Collection version 1.0/2.0 and the Drosophila RNAi Screening Center Genome-wide RNAi library (DRSC 2.0).

RNAi Screening

OSS cells were plated in 48-well dishes (79,000 cells/well). The following day cells were transfected with 500ng of dsRNA, 0.3µl Xfect reagent and 9.7µl Xfect Buffer. To do this procedure in a robust way the Epmotion robot (Eppendorf) was used to prepare the transfection mixture in a 96-well plate and to pipette the mixture onto the cells. Approximately 12 hours post transfection, cells were washed with PBS and media was replaced. An additional media change was done on day 3 post-transfection to avoid drying of wells. On day 5 post-transfection cells were lysed with 150ul of Lysis Buffer (10mM KCl, 10mM Tris pH8, 1.5mM MgCl2, 0.5% NP-40, 60 units RNasin) per well and shaken for 5min at 300 rpm. For the DRSC library, instead of the Lysis Buffer, Ambion Cells-to-Ct Lysis Reagent (Life Technologies cat 4391848M) was used to lyse cells. Following the 5 minutes of shaking, 15µl of Stop Solution (Life Technologies cat 4402960) was added to stop the lysis reaction, mixed by pipetting, and left for 2 minutes at room temperature. Lysates were transferred to a 96-well PCR plate. 22.5µl of the lysate was used as input for a 50µl reverse transcription (RT) reaction and then incubated at 37ºC for 1 hour and 95ºC for 5 min. The RT master mix and enzyme used were those provided in the TaqMan Gene expression Cells-to-CT kit (4399002). Both the transfer of the lysate to 96-well plates, as well as the RT reaction set-up was done using the Epmotion. After cDNA synthesis, 2µl of the cDNA was used as input in a qPCR reaction to assay levels of *gypsy* and cib in a multiplexed reaction, using TaqMan Fast Advanced Master Mix (Life Technologies cat 4444965) on an Eppendorf MasterCycler EP realplex machine. Levels of *gypsy* subgenomic transcript and the reference gene *ciboulot* were assayed using hydrolysis probes spanning splice junctions and the resulting Ct values were expressed as delta-Ct z-scores (distance in standard deviations from the plate median) and fold change (in relation to the plate median). For further analysis, we ignored extreme outliers for the reference gene and wells in which *gypsy* could not be detected after 38 cycles of qPCR. Targets were called a primary hit if the *gypsy* delta-Ct z-score was lower than or equal to -1.9 and the *gypsy* fold change higher than or equal to 3. We called an additional 22

genes a primary hit based on b-score normalization (Ramadan et al., 2007). After calling primary hits Primers and probes are listed in Table S4.

Drosophila Stocks and Husbandry

For crosses in the validation round we used *tj*-GAL4 (DGRC stock 10455); GS12962 (DGRC stock 204406) and EY21234 (Bloomington stock 22462) are P-element insertions into the CG3893 locus. The 328 fly stocks corresponding to the candidate hits were ordered from Vienna Drosophila Resource Center (VDRC) and the Drosophila RNAi Resource center. The trans-IDs used by VDRC are listed in Supplementary Table S3. Lines from the DGRC are indicated with the prefix TRIP. For all crosses performed during the validation screen, five tj-GAL4 females and three VDRC hpRNA males were crossed and left in vials for five days, when parental flies were removed from the vial. Eight days after, ten female and three male F1 flies were put into new vials with yeast. After two days, ovaries from female flies were dissected. Eight days later, we checked the vials for the presence of larvae to test for fertility.

RNA Isolation and qPCR Assays

Ovaries from 10 F1 flies were dissected for each cross. Ovaries were washed once with cold PBS and homogenized in 1 ml of Trizol reagent. Total RNA was purified by phenol chloroform extraction followed by isopropanol precipitation according to the Trizol protocol. RNA was then subjected to DNase treatment using Ambion Turbo DNA-free kit at 37ºC for 30 minutes according to the manufacturer's protocol (Life Technologies). cDNA was synthesized with 800ng RNA as input using oligo dT primers (dT20) and Superscript III Reverse Transcriptase (Life Technologies) at 50ºC for 50 minutes, followed by 15 minutes at 70ºC. Next, qPCR was performed to assay levels of *gypsy*, *ZAM*, *gypsy3* and rp49. Using hydrolysis probes with FAM and HEX fluorescent reporters, we multiplexed the qPCR for the transposon and rp49. Primers and probes are listed in Table S4. Fold changes for transposons were calculated using the delta Ct method (Livak and Schmittgen, 2001). In the case of the GD library we compared each knockdown to an average of 5 biological replicates of White negative controls, for the KK library

we used 5 biological replicates of Aub negative controls. All primers were tested for efficiency in single and multiplexed reactions. Only primers for which efficiency was not impaired in the multiplexed reactions were used.

RNA-Seq and Analysis

For RNA-seq libraries, 2.5-5ug of total RNA was depleted of ribosomal RNA using the Epicenter Ribo-Zero rRNA Removal Kits (Human/Mouse/Rat), following the manufacturer's directions. Libraries were prepared using the Illumina Script Seq v2 RNA-Seq library preparation kit and were sequenced on an Illumina HiSeq platform for 36 cycles in a single end run. After collapsing all reads into a non-redundant list (cloning counts were preserved), they were mapped to Drosophila viral, tRNA and miscRNA (rRNA, snoRNA etc) sequences using the short read aligner Bowtie (Langmead et al., 2009). Only sequences in each library that did not map to either of these contaminants were then mapped to the Drosophila genome with up to two mismatches. Additionally, only uniquely mapping sequences were considered for further analysis. The same reads were mapped to a custom index of transposon consensus sequences with up to 2 mismatches (Kaminker et al., 2002). Reads mapping to up to 2 locations were considered for further analysis. For differential expression analysis of transposons we aggregated read counts mapping to these consensus sequences in sense orientation. For differential expression analysis of genes, we used htseq-counts (Part of the 'HTSeq' framework, version 0.5.3p3) to assess read counts per gene. In both cases we used the R package DESeq to call differential expression at a FDR cutoff of 0.05 based on two biological replicates (Anders and Huber, 2010).

Small RNA Cloning and Analysis

For small RNA libraries, 2.5 µg of total RNA was depleted of the 2S rRNA by annealing an antisense primer (Table S4, 95ºC to 25ºC in ~1h) followed by RNase H digestion at 37ºC for 30 minutes in 5X FS buffer (from Superscript III Reverse Transcriptase Kit, Life Technologies; RNase H was from NEB, M0297S). The remaining RNA was used as input. Libraries were

constructed using the Illumina TruSeq small RNA sample Prep kit following the manufacturer's protocol. For analysis of sRNA populations of CG3893 heterozygous and mutant animals, we used 50 ng of size selected RNA (19-28nt) as input. After sequencing on a Illumina HiSeq single-end 36 run, the TruSeq adapter (TGGAATTCTCGGGTGCCAAGGAACTCCAGTCAC) was clipped from the 3' end of the read and sequences shorter than 15 nt were discarded from further analysis. The remaining sequences were collapsed into a non-redundant list and mapped to Drosophila viral, tRNA and miscRNA (rRNA, snoRNA etc) sequences using the short read aligner Bowtie (Langmead et al., 2009). Only non-mapping reads were consequently mapped to the D. melanogaster genome (D. melanogaster Apr. 2006 [BDGP R5/dm3]). Up to two mismatches were allowed. Read counts of uniquely mapping reads were normalized to reads per million genomic mappers and compared to a negative control: in the case of knockdowns using long hpRNAs from the VDRC KK libraries we used Aub (106999^{KK}), in the case of GD libraries we used White (30033^{GD}). The same reads were mapped to a custom index of transposon consensus sequences with up to 2 mismatches (Kaminker et al., 2002). Reads mapping uniquely were considered for further analysis. The percentages of flamenco mappers displayed in Figure 4A are based on read counts normalized to 42AB. The rankings displayed in Figure 4B are calculated based on aggregated read counts of unique mappers to piRNA clusters defined in Brennecke et al. (2007). For size profiles, we used the same negative control libraries for comparison, which were normalized to the same scale in order to accurately compare across knockdowns. For analysis of transposons we aggregated read counts mapping to consensus sequences in sense orientation and normalized to the counts of three germline dominant transposons (*roo*, *Rt1b* and *Het-A*).

ChIP-Seq

ChIP was done as described in Ram et al. (2011) and Garber et al. (2012), with some modifications. Approximately fifty ovaries were dissected from heterozygous or homozygous flies into cold PBS and washed once with PBS. Ovaries were then fixed in 1.8% formaldehyde

for 10 minutes, then quenched by adding glycine to 0.125M and immediately placed on ice. Tissue was then homogenized by douncing five times with pestle A (Kontes). Washed once with PBS supplemented with protease inhibitors (Roche) and pellet was flash frozen in liquid nitrogen. Pellets were then thawed on ice and resuspended in 1mL Lysis Buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl, pH 8.1) and lysed for 10 minutes in ice. Chromatin was sheared to 200-800bp using a Branson sonifier (model S-450D). After clearing lysate by centrifugation, 9mLs of Dilution Buffer (0.01% SDS, 1.1%Triton X-100, 1.2mM EDTA, 16.7mM Tris-HCl, pH 8.1, 167mM NaCl) were added to the lysate and 5mLs of the lysate were incubated with a 50ul of an equal mixture of conjuagted protein A and G Dynabeads (Invitrogen). To conjugate beads, they first had been washed once in Blocking Buffer (1X PBS, 0.5% TWEEN 20, 0.5% BSA), then coupled for 1 hour at 4ºC with 5ug of H3K9me3 antibody (Abcam 8898) and finally washed twice with Blocking Buffer to remove excess antibody. Lysate and conjugated magnetic beads were rotated at 4ºC overnight. Beads were then resuspended in 200ul cold RIPA buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 14 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% DOC) and transfer to a 96-well plate. All further separation steps were performed in the 96-well plate magnet. Beads were washed five times with 200ul cold RIPA, two times with RIPA buffer supplemented with 500 mM NaCl, two times with LiCl buffer (10 mM TE, 250mM LiCl, 0.5% NP-40, 0.5% DOC), and once with TE (10Mm Tris-HCl pH 8.0, 1mM EDTA). Samples were eluted in 50 µl of 0.5% SDS, 300 mM NaCl, 5 mM EDTA, 10 mM Tris HCl pH 8.0. The eluate was reverse cross-linked at 65ºC for 4 hours and then treated with 2ul of RNaseA (Roche, 11119915001) for 30 min followed by 2.5 µl of Proteinase K (NEB, P8102S) for two hours. Library preparation was done as indicated in Garber et al. (2012), but without automation. In brief, to purify DNA 120ul of Ampure XP beads (Agencourt) were added to the reverse cross-linked samples, mixed by pipetting and incubated for 2 minutes. Samples were then placed on the magnetic stand for 4 minutes to separate beads, followed by 2 washes with 70% ethanol and air dried for 4 minutes and eluted in 10mM Tris-HCl pH 8.0. Library was constructed by

performing DNA end-repair, A-base addition, adaptor ligation and enrichment PCR. After each step DNA was purified by adding 20% PEG and 2.5 M NaCl to the reaction, to allow DNA to bind to Ampure XP beads already in the tube. Samples were not moved form their original well position, until after PCR enrichment. The libraries were sequenced on the Illumina HiSeq platform for 76 cycles in a pair-end run. The resulting reads were mapped with Bowtie 2 with the preset option --sensitive (Langmead and Salzberg, 2012). Only read pairs mapping concordantly were used for further analysis. For calling H3K9me3 peaks, annotation of called peaks and visualization, we used the HOMER software package (Heinz et al., 2010). Enrichments of H3K9me3 signal were calculated using input libraries as a control signal. All peaks within 8kb distance from each other were merged into regions. We used annotatePeaks from Homer to then calculate the tag counts in heterozygous and homozygous libraries over those regions, normalizing each tag to reads per million genomic mappers.

Statistical Procedures

Enrichment analysis was conducted using 2714 gene sets from the gene ontology (Ashburner et al., 2000; Barrell et al., 2009). This constituted the complete complement of gene sets in GO with between five and 100 Drosophila genes annotated to them in either the cellular component or biological process branch of GO. Molecular function substantially overlapped with biological process in many top functions and was excluded to diminish redundancy. Significance was calculated using an adaptation of the ROC-based approach described in (Gillis et al., 2010) and elsewhere. To obtain a ranking for the genes, dsRNA z-scores and fold changes were independently averaged for each gene. These scores were then converted into ranks and averaged (effectively weighting them equally). Based on the ROC_{50} approach first described in (Gribskov and Robinson, 1996), all scores outside of the top 50 were regarded as tied. Statistical enrichment of the GO functions was then calculated (Mann-Whitney U test) with multiple test correction (Benjamini, 1995).

Supplemental References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics *25*, 25-29.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.

Barrell, D., Dimmer, E., Huntley, R.P., Binns, D., O'Donovan, C., and Apweiler, R. (2009). The GOA database in 2009--an integrated Gene Ontology Annotation resource. Nucleic acids research *37*, D396-403.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met *57*, 289-300.

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*, 1089-1103.

Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., Robinson, J., Minie, B., Chevrier, N., Itzhaki, Z.*, et al.* (2012). A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. Molecular cell *47*, 810-822.

Gillis, J., Mistry, M., and Pavlidis, P. (2010). Gene function analysis in complex data sets using ErmineJ. Nature Protocols *5*, 1148-1159.

Gribskov, M., and Robinson, N.L. (1996). Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching. Computers & chemistry *20*, 25-33.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription

factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell *38*, 576-589.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M.*, et al.* (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome biology *3*, RESEARCH0084.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359.

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 biology *10*, R25.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Niki, Y., Yamaguchi, T., and Mahowald, A.P. (2006). Establishment of stable cell lines of Drosophila germ-line stem cells. Proceedings of the National Academy of Sciences of the United States of America *103*, 16325-16330.

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M.*, et al.* (2011). Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. Cell *147*, 1628-1639.

Ramadan, N., Flockhart, I., Booker, M., Perrimon, N., and Mathey-Prevot, B. (2007). Design and implementation of high-throughput RNAi screens in cultured Drosophila cells. Nature Protocols *2*, 2245-2264.