Developmental Cell, Volume *24*

Supplemental Information

A Major Epigenetic Programming Mechanism Guided by piRNAs

Xiao A. Huang*, Hang Yin*, Sarah Sweeney, Debasish Raha, Michael Snyder and Haifan Lin

INVENTORY OF SUPPLEMENTAL MATERIALS

1. Supplemental Figures

Figure S1, related to Figure 1. Validation of Piwi antibody and comparison of Piwi binding in whole flies *vs* ovaries.

Figure S2, related to Figure 2. Genome-wide Piwi chromatin binding pattern and its correlation with Piwi-associated piRNA counts.

Figure S3, related to Figure 4. Sequence specificity for Piwi and HP1a recruitment.

Figure S4, related to Figure 6. No piRNA phasing was observed along euchromatic or heterochromatic genome.

Figure S5, related to Figure 7. Distributions of epigenetic regulators/marks across genome of *piwi*^{1}*/piwi*^{2}*,*

Figure S6, related to Figures 7. Causal relationships between epigenetic changes and adjacent localization of Piwi.

2. Legends for Supplementary Figures

3. Supplementary Tables

Table S1, related to Figure 1. Summary of Solexa Sequencing

Table S2, related to Figure 7. Enriched GO terms for repressed genes in *piwi*^{1}/*piwi*^{2}

Table S3, related to Figure 7. Enriched GO terms for activated genes in *piwi*^{1}/*piwi*^{2}

4. Supplementary Experimental Procedures

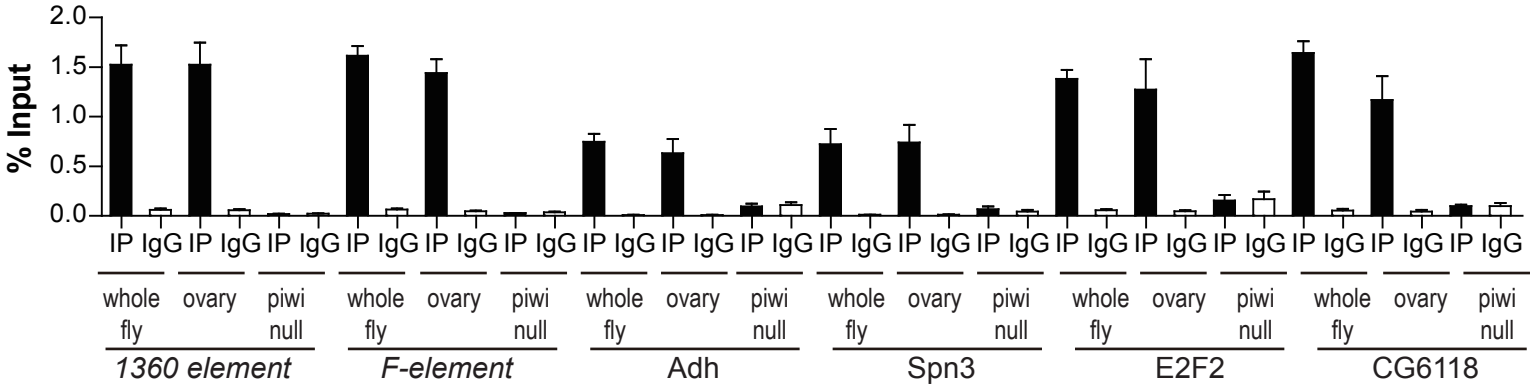
% Input
2.0
1.5
1.0
0.5
0.0
IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG IP IgG
whole fly
ovary
piwi null
1360 element
F-element
Adh
Spn3
E2F2
CG6118

Figure S1

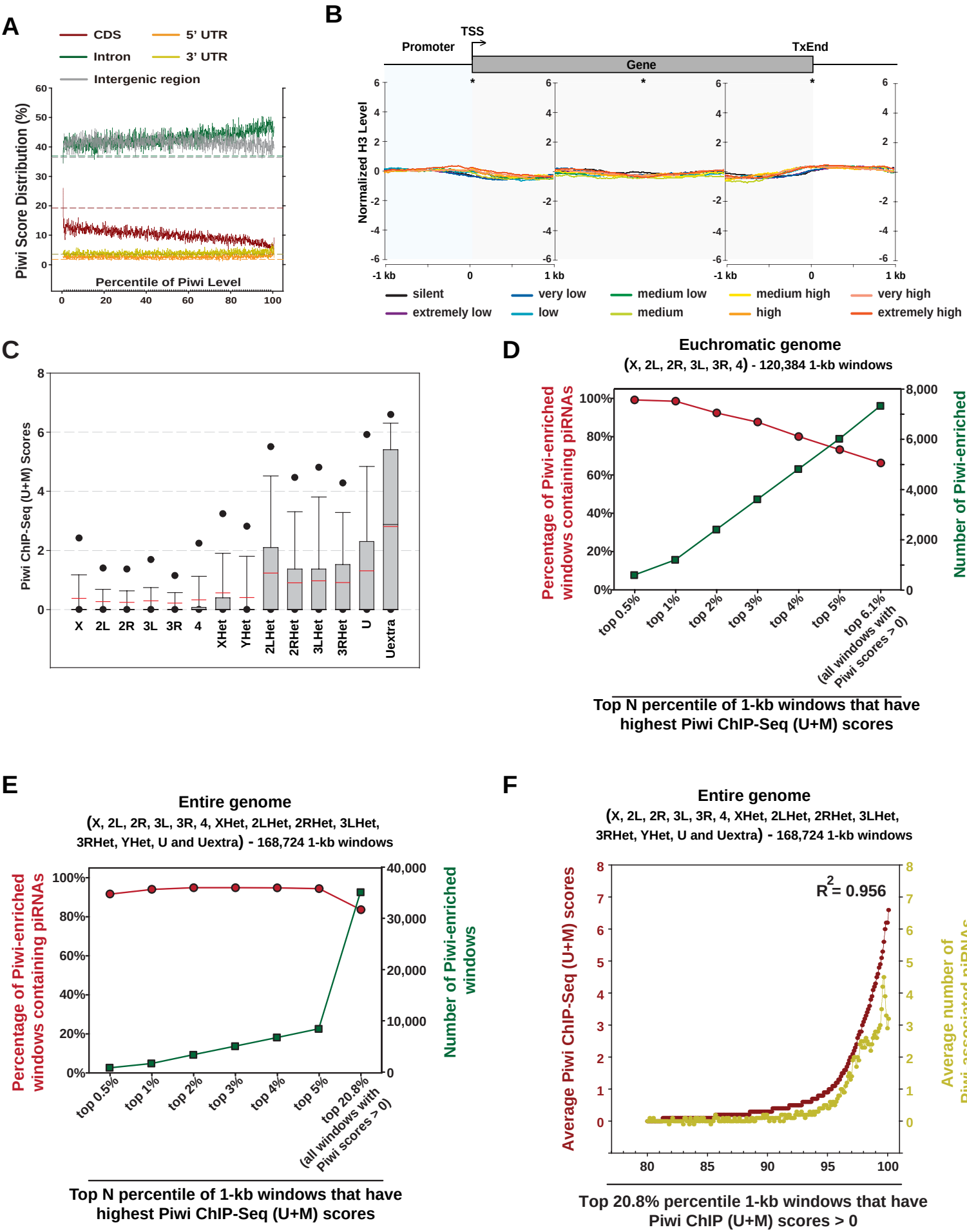
A
CDS
5' UTR
Intron
3' UTR
Intergenic region
Piwi Score Distribution (%)
Percentile of Piwi Level
B
TSS
Promoter
TxEnd
Gene
Normalized H3 Level
-1 kb
0
1 kb
silent
very low
medium low
medium high
very high
extremely low
low
medium
high
extremely high
C
Piwi ChIP-Seq (U+M) Scores
X
2L
2R
3L
3R
4
XHet
YHet
2LHet
2RHet
3LHet
3RHet
U
Uextra
D
Euchromatic genome
(X, 2L, 2R, 3L, 3R, 4) - 120,384 1-kb windows
Percentage of Piwi-enriched windows containing piRNAs
Number of Piwi-enriched windows
top 0.5%
top 1%
top 2%
top 3%
top 4%
top 5%
top 6.1% (all windows with Piwi scores > 0)
Top N percentile of 1-kb windows that have highest Piwi ChIP-Seq (U+M) scores
E
Entire genome
(X, 2L, 2R, 3L, 3R, 4, XHet, 2LHet, 2RHet, 3LHet, 3RHet, YHet, U and Uextra) - 168,724 1-kb windows
Percentage of Piwi-enriched windows containing piRNAs
Number of Piwi-enriched windows
top 0.5%
top 1%
top 2%
top 3%
top 4%
top 5%
top 20.8% (all windows with Piwi scores > 0)
Top N percentile of 1-kb windows that have highest Piwi ChIP-Seq (U+M) scores
F
Entire genome
(X, 2L, 2R, 3L, 3R, 4, XHet, 2LHet, 2RHet, 3LHet, 3RHet, YHet, U and Uextra) - 168,724 1-kb windows
R2 = 0.956
Average Piwi ChIP-Seq (U+M) scores
Average number of Piwi-associated piRNAs
Top 20.8% percentile 1-kb windows that have Piwi ChIP (U+M) scores > 0

Figure S2

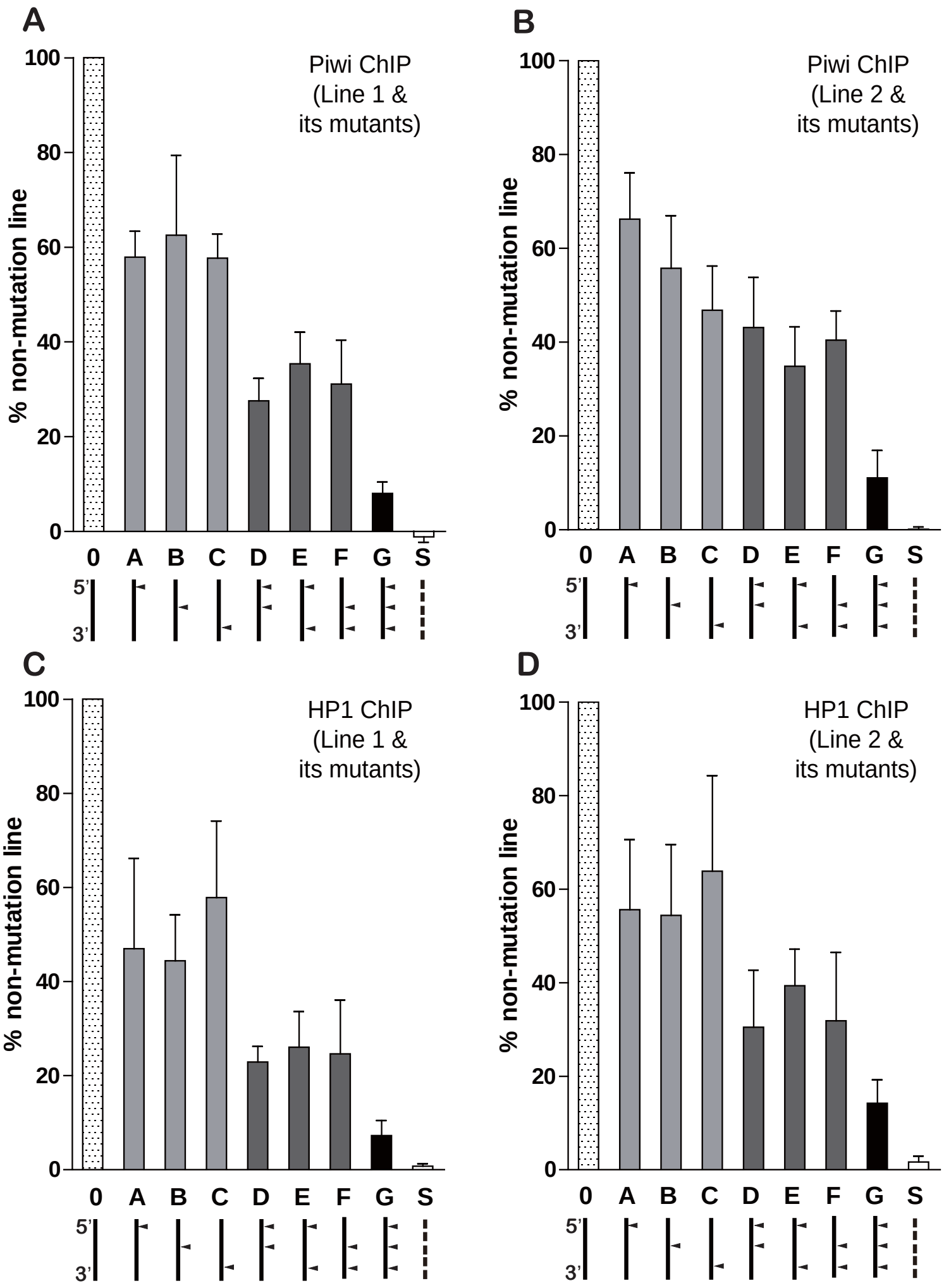
A
Piwi ChIP
(Line 1 &
its mutants)
% non-mutation line
100
80
60
40
20
0
0 A B C D E F G S
5'
3'
B
Piwi ChIP
(Line 2 &
its mutants)
% non-mutation line
100
80
60
40
20
0
0 A B C D E F G S
5'
3'
C
HP1 ChIP
(Line 1 &
its mutants)
% non-mutation line
100
80
60
40
20
0
0 A B C D E F G S
5'
3'
D
HP1 ChIP
(Line 2 &
its mutants)
% non-mutation line
100
80
60
40
20
0
0 A B C D E F G S
5'
3'

Figure S3

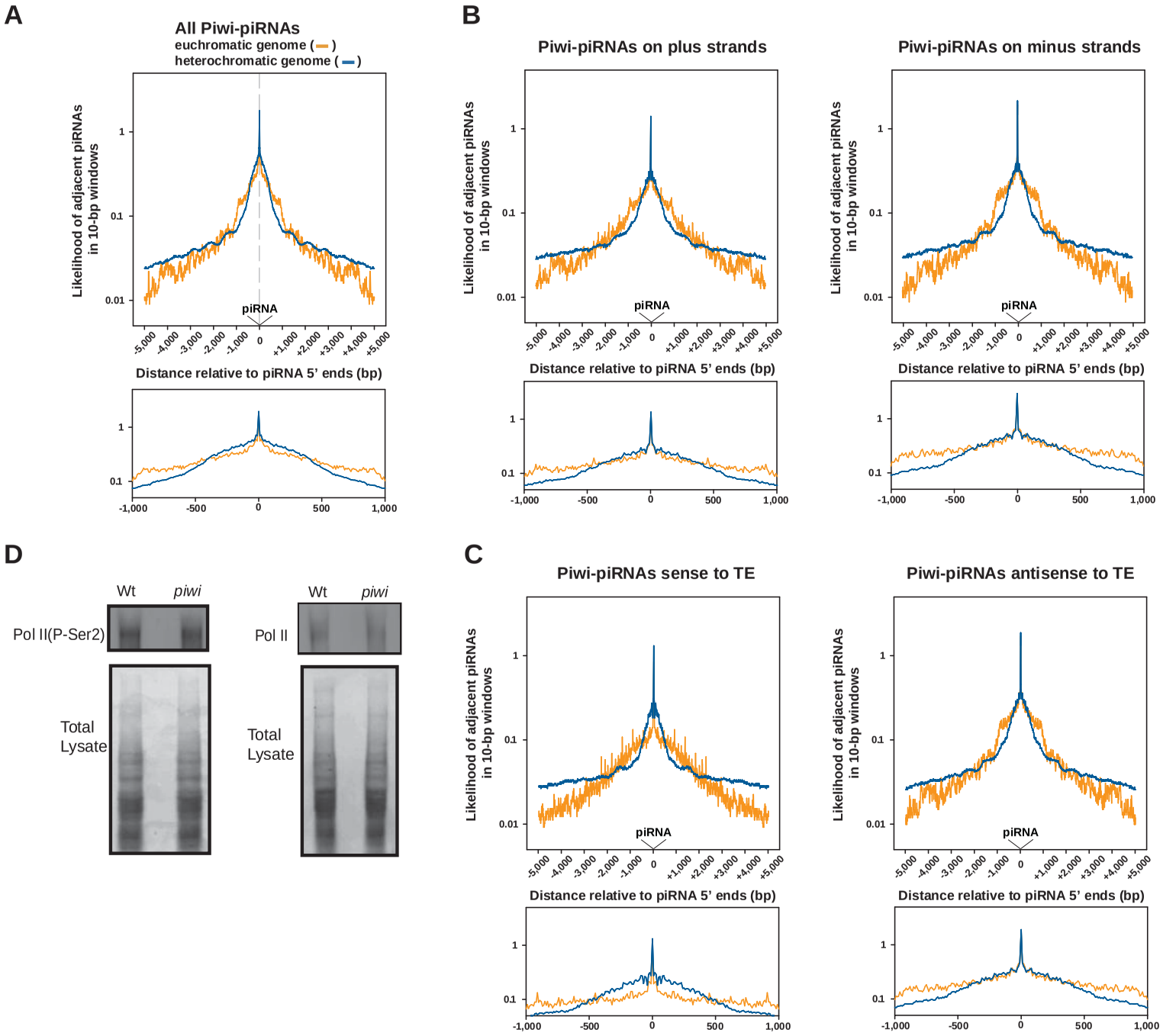
A
All Piwi-piRNAs
euchromatic genome ( )
heterochromatic genome ( )
Likelihood of adjacent piRNAs in 10-bp windows
piRNA
Distance relative to piRNA 5' ends (bp)
B
Piwi-piRNAs on plus strands
Piwi-piRNAs on minus strands
C
Piwi-piRNAs sense to TE
Piwi-piRNAs antisense to TE
D
Wt
piwi
Pol II(P-Ser2)
Pol II
Total Lysate

Figure S4

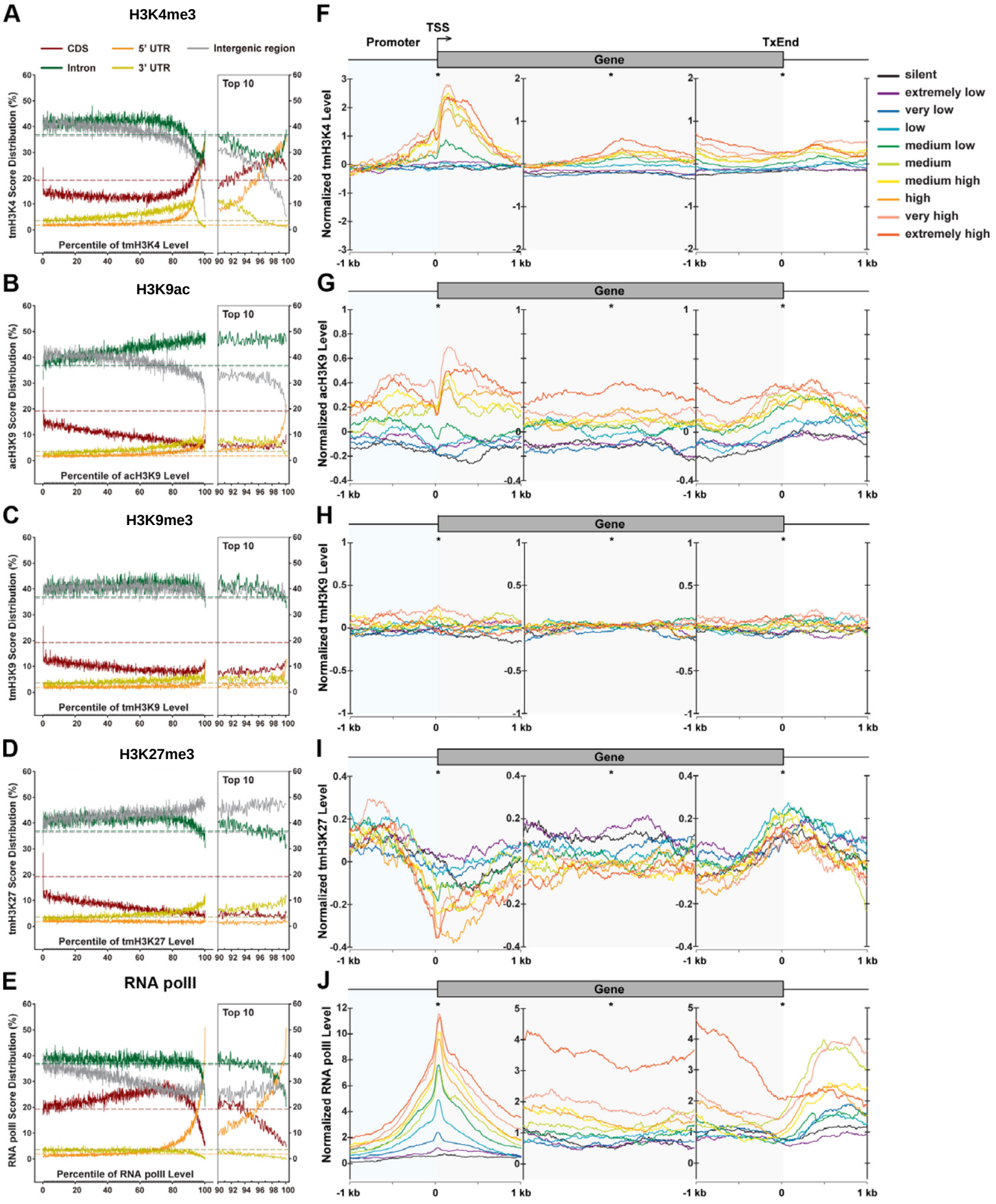
A
H3K4me3
CDS
5' UTR
Intergenic region
Intron
3' UTR
Top 10
tmH3K4 Score Distribution (%)
Percentile of tmH3K4 Level
B
H3K9ac
acH3K9 Score Distribution (%)
Percentile of acH3K9 Level
C
H3K9me3
tmH3K9 Score Distribution (%)
Percentile of tmH3K9 Level
D
H3K27me3
tmH3K27 Score Distribution (%)
Percentile of tmH3K27 Level
E
RNA polII
RNA polII Score Distribution (%)
Percentile of RNA polII Level
F
TSS
Promoter
TxEnd
Gene
Normalized tmH3K4 Level
silent
extremely low
very low
low
medium low
medium
medium high
high
very high
extremely high
G
Normalized acH3K9 Level
H
Normalized tmH3K9 Level
I
Normalized tmH3K27 Level
J
Normalized RNA polII Level
-1 kb
0
1 kb

Figure S5

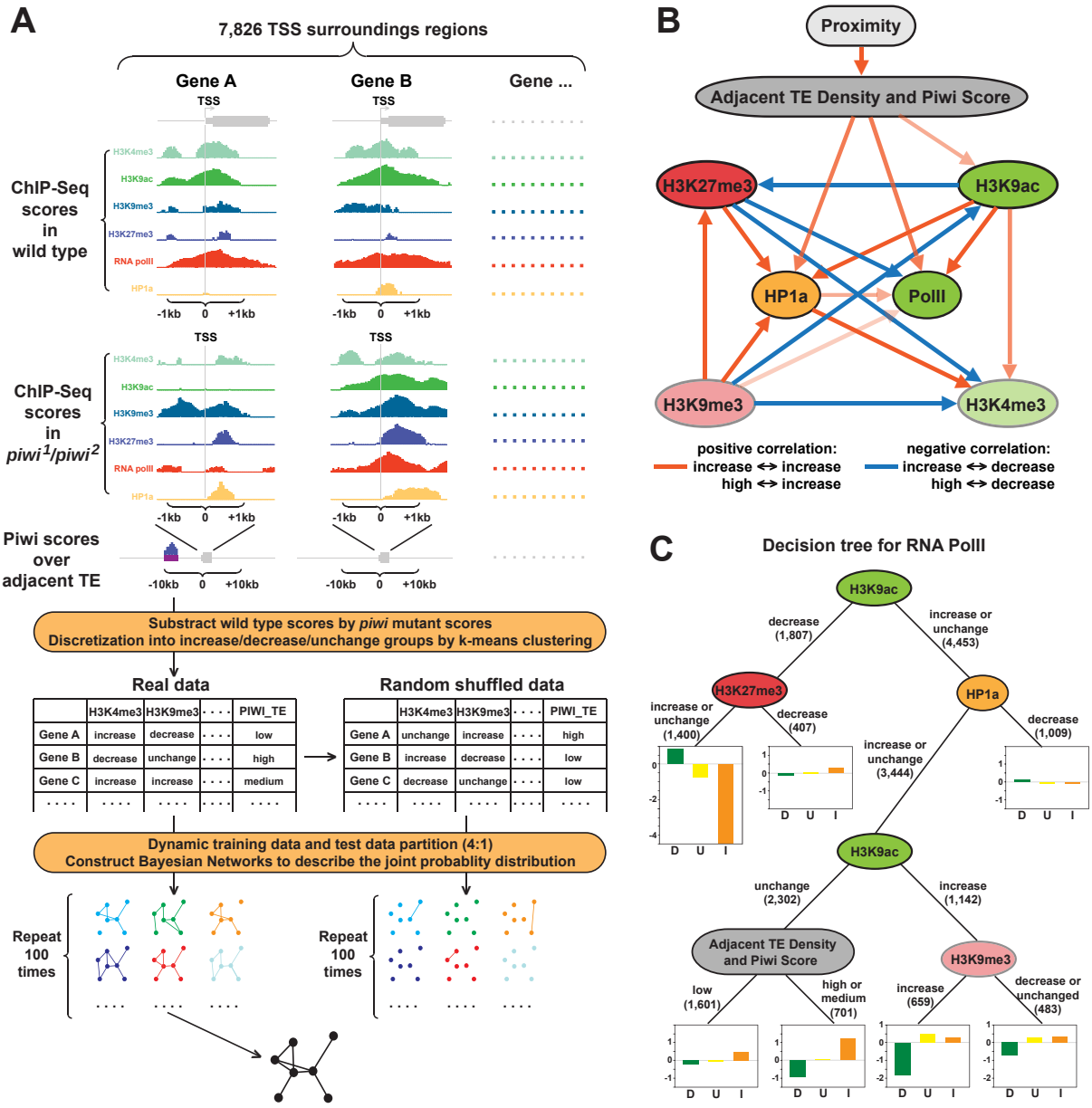
A
7,826 TSS surroundings regions
Gene A
Gene B
Gene ...
TSS
ChIP-Seq scores in wild type
ChIP-Seq scores in piwi1/piwi2
Piwi scores over adjacent TE
H3K4me3
H3K9ac
H3K9me3
H3K27me3
RNA polII
HP1a
-1kb 0 +1kb
-10kb 0 +10kb
Substract wild type scores by piwi mutant scores
Discretization into increase/decrease/unchange groups by k-means clustering
Real data
Random shuffled data
PIWI_TE
Gene A increase decrease low
Gene B decrease unchange high
Gene C increase increase medium
Gene A unchange increase high
Gene B increase decrease low
Gene C decrease unchange low
Dynamic training data and test data partition (4:1)
Construct Bayesian Networks to describe the joint probablity distribution
Repeat 100 times
B
Proximity
Adjacent TE Density and Piwi Score
H3K27me3
H3K9ac
HP1a
PolII
H3K9me3
H3K4me3
positive correlation:
increase ↔ increase
high ↔ increase
negative correlation:
increase ↔ decrease
high ↔ decrease
C
Decision tree for RNA PolII
H3K9ac
decrease (1,807)
increase or unchange (4,453)
H3K27me3
HP1a
increase or unchange (1,400)
decrease (407)
increase or unchange (3,444)
decrease (1,009)
D U I
H3K9ac
unchange (2,302)
increase (1,142)
Adjacent TE Density and Piwi Score
H3K9me3
low (1,601)
high or medium (701)
increase (659)
decrease or unchanged (483)

Figure S6

LEGENDS FOR SUPPLEMENTARY FIGURES 1-6

Figure S1. Validation of Piwi antibody specificity and comparison of Piwi binding in whole flies *vs* ovaries, Related to Figure 1.

Piwi association with the chromatin was determined in lysates from widltype adult flies and ovaries as well as from *piwi*-null adult flies. At six randomly selected genomic loci (1360 element, F-element, Adh, Spn3, E2F2 and CG6118) to represent Piwi binding sites in genes and heterochromatin, adult flies and ovaries show similar Piwi-immunoprecipitation (IP) signals, which are eliminated in the *piwi* mutant adult flies. Piwi-IP and control IgG-IP signals are shown as a percentage of Input as black and white bars, respectively. Each data point was repeated at least three times and error bar shows the standard error of the mean. Error bars denote S.D.

Figure S2. Genome-wide Piwi chromatin binding pattern and its correlation with Piwi-associated piRNA counts, Related to Figure 2

(**A**) Distribution of Piwi ChIP-Seq (U+M) scores across the euchromatic genome indicates that Piwi is relatively enriched in introns and intergenic regions, but very low abundant in gene coding sequences. The genomic regions on chromosome X, 2, 3 and 4 were divided into 2,407,635 50-bp windows, which were ranked into 100 percentiles based on Piwi ChIP-Seq (U+M) scores. For each score percentile, the percentages of windows overlapping with features of CDS, Intron, 5' UTR, 3' UTR and intergenic regions were drawn in different colors. The percentages of these genomic features in the euchromatic genome are indicated by dash lines in the corresponding colors.

(**B**) Profiles of Piwi ChIP-Seq (U+M) scores over models of protein coding genes in the wildtype genome show that Piwi is not particularly enriched in regions surrounding transcription start sites (TSS), transcript ends (TxEnd) or protein coding gene bodies. 20,738 transcripts from 9,338 protein coding genes were aligned at transcription start sites, mid points of transcript, and transcription end sites. Distribution of Piwi ChIP-Seq (U+M) scores within 2-kb regions around the aligning points were interrogated in 10-bp resolution. Genes were classified into 10 groups based on their expression levels.

(**C**) Distributions of Piwi ChIP-Seq (U+M) scores over different chromosomes or contigs demonstrate that Piwi is enriched in repetitive genomic regions. The top and bottom boarders of each box indicate 75th and 25th percentiles, respectively. The upper and lower ends of whiskers indicate 90th and 10th percentiles, respectively. The upper and lower black dots indicate 95th and 5th percentiles, respectively. Red lines and black line within boxes denote mean and medium values, respectively.

(**D-E**) Comparisons of genomic locations of Piwi and Piwi-associated piRNAs in euchromatic genome (D) and entire genome (E) reveal that Piwi-enriched regions overlap with piRNA annotations. The euchromatic portion of *D. melanogster* genome (X, 2L, 2R, 3L, 3R, 4) was divided into 120,384 1-kb windows, which were sorted based on their Piwi ChIP-Seq (U+M) scores. Of these, 7,329 (6.1%) 1-kb windows have Piwi ChIP-Seq (U+M) scores more than zero. 1-kb windows associated with highest Piwi ChIP-Seq scores (top 0.5%, top 1%, top 2%, top 3%, top 4%, top 5% from the 120,384 1-kb windows) and all 7,329 1-kb windows were overlapped with Piwi-associated piRNA annotations. The percentages of the Piwi-enriched windows overlapping with at least one piRNA annotation were plotted together with the number of Piwi-enriched windows in each group (E). Similarly, the entire *D. melanogster* genome (X, 2L, 2R,

3L, 3R, 4, 2LHet, 2RHet, 3LHet, 3RHet, YHet, U and Uextra) was divided into 168,724 1-kb windows, which were sorted based on their Piwi ChIP-Seq (U+M) scores. Of these, 35,107 (20.8%) 1-kb windows have Piwi ChIP-Seq (U+M) scores more than zero. 1-kb windows associated with highest Piwi ChIP-Seq scores (top 0.5%, top 1%, top 2%, top 3%, top 4%, top 5% from the 168,724 1-kb windows) and all 35,107 1-kb windows were overlapped with Piwi-associated piRNA annotations. The percentages of the Piwi-enriched windows overlapping with at least one piRNA annotation were plotted together with the number of Piwi-enriched windows in each group (E).

(**F**) A correlation analysis comparing genome-wide localizations of Piwi and Piwi-associated piRNAs indicate that Piwi-enriched regions are quantitatively associated with larger number of piRNAs. The entire *D. melanogster* genome was divided into 168,724 1-kb windows and averaged Piwi ChIP-Seq (U+M) scores and counts of Piwi-associated piRNAs were calculated for each window. These 1-kb windows were ranked into 100 percentiles by their Piwi ChIP-Seq scores and average scores for windows within a percentile were shown in dots for both Piwi ChIP-Seq scores (red) and Piwi-associated piRNA counts (red). As the majority of 1-kb windows have both Piwi scores and piRNA counts equal to zero, only the top 20.8% (35,107) of 1-kb windows that have Piwi ChIP-Seq scores more than zero were plotted and included in the calculation of Pearson Product-Moment correlation coefficient between Piwi scores and piRNA counts.

Figure S3. Sequence specificity for Piwi and HP1a recruitment, Related to Figure 3

(**A-B**) Piwi ChIP in transgenic lines with point mutations and scrambled sequences for Lines 1 and 2, respectively. Transgenic lines A-C carry single point mutations in the ectopic target

sequence of Line 1 or 2 as schematically indicated in the bottom of each panel. Transgenic lines D-F carry double point mutations and line G has a triple mutation. Line S has the wildtype piRNA target sequence in Lines 1 and 2 replaced by non-related bacterial sequences of the same length and repeat copy number correspondingly. For each line, Piwi ChIP signals subtracted by IgG at the ectopic sites was normalized to the corresponding signal at the original sites. The resulting normalized Piwi binding score for each mutational line was shown as a percentage of Line 1 (A and C) or 2 (B and D). Error bars in A-D denote S.D.

(**C-D**) HP1 ChIP in the same transgenic lines as in Panels A and B.

Figure S4. No piRNA phasing was observed along euchromatic or heterochromatic genome, Related to Figure 5

The distance between a piRNA (5' end) and its adjacent 100 piRNAs (5' ends) were calculated and binned into 10-bp windows. The accumulative numbers of adjacent piRNAs in every 10-bp windows for all piRNAs in euchromatic (yellow) and heterochromatic genome (blue) were normalized to the total number of piRNAs in euchromatic or heterochromatic regions, respectively. Distributions of adjacent piRNAs were plotted in ranges of $\pm$5 kb (the upper charts in all panels) and $\pm$1 kb (the lower charts in panels).

(A) Distribution of all PIWI-piRNAs with respective to their adjacent PIWI-piRNAs.

(B) Distribution of PIWI-piRNAs on DNA plus strands (left) and minus strands (right) with respect to their adjacent piRNAs on corresponding DNA strands.

(C) Distribution of PIWI-piRNAs on sense and antisense strands of transposable elements (TE) with respect to their adjacent piRNAs on the corresponding TE strands.

(D) Levels of phosphorylated Ser2 PolII and total PolII in wildtype and *piwi*^{1}/*piwi*^{2} mutant flies. The level of Ser2-phosphoylated RNA Pol II (A) and total RNA Pol II (B) in wildtype and *piwi* null mutants was determined by western blotting. The total lysate loading was shown by Ponceau S staining.

Figure S5. Distributions of epigenetic regulators/marks across genome of *piwi*^{1}/*piwi*^{2}, Related to Figure 6

(**A-E**) Distributions of H3K4me3, H3K9me3, H3K27me3, H3K9ac and RNA PolII ChIP-Seq scores across genome of *piwi*^{1}/*piwi*^{2}. Genomic regions on chromosome X, 2, 3 and 4 were divided into 2,407,635 50-bp windows, which were ranked into 100 percentiles by ChIP-Seq scores. For each score percentile, the percentages of windows overlapping with features of CDS, introns, 5' UTR, 3' UTR and intergenic regions were drawn in different colors. The percentages of these genomic features in the genome are indicated by dash lines in the corresponding colors. Right panels show zoom-in distribution profiles for windows with top 10 percent of ChIP-Seq scores.

(**F-J**) Profiles of H3K4me3, H3K9me3, H3K27me3, H3K9ac and RNA PolII ChIP-Seq scores over features of protein coding transcripts in genome of *piwi*^{1}/*piwi*^{2}. 20,738 transcripts from 9,338 protein-coding genes were aligned at transcription start sites (TSS), mid points of transcript, and transcription ends (TxEnd). Distribution of ChIP-Seq scores within 2-kb regions surrounding the aligning points were interrogated in 10-bp resolution. Genes were classified into 10 groups based on their expression levels.

Figure S6. Causal relationships between epigenetic changes and adjacent localization of Piwi, Related to Figures 6 and 7

(**A**) A flowchart illustrates the Bayesian network construction using ChIP-Seq scores within TSS ± 1kb regions. See text for details.

(**B**) The common Bayesian network consisted of only compelled edges agreed by all 100 trials. Nodes are colored by their correlation to gene expression with exception of HP1a, which is reported to have opposite epigenetic functions in gene expression. Color intensities of nodes are scaled related to their correlation to gene expression. A directed edge denotes the occurrence of the target node is dependent on that of the source node, or that the occurrence of the source node is predictive of the target node. The edge colors indicate the positive correlations (red) and negative correlations (blue). The edge color intensities denote correlation levels (measured by Pearson Product-Moment correlation coefficient).

(**C**) Decision tree of RNA PolII within the Bayesian network shown in panel B. Tree bifurcates based on epigenetic changes of the root node. Numbers of gene classified into branches are shown in bracket. Over-representation (positive values) and under-representation (negative values) of increased (I), decreased (D) or unchanged (U) fractions of RNA PolII are shown at the ends of the tree in $\log_2$ scale.

Table S1. Summary of Solexa Sequencing, related to Figure 1.

Sequencing Scale		Non-redundant Solexa tags			Solexa reads		
		total tag #	unique tag #	multiple tag #	total read #	unique read #	multiple read #
wild type	Non-ChIP	3,381,912	2,525,205	856,707	4,038,814	2,574,146	1,464,668
	PIWI	2,649,033	1,962,740	686,293	2,882,370	1,980,498	901,872
	HP1a	2,677,024	2,085,608	591,416	2,967,286	2,114,682	852,604
	RNA polII	4,366,407	3,861,312	505,095	5,626,556	4,805,130	821,426
	H3	5,693,771	4,747,317	946,454	6,351,188	5,046,304	1,304,884
	tmH3K4	5,131,829	4,234,712	897,117	5,650,921	4,339,650	1,311,271
	tmH3K9	3,824,497	2,970,904	853,593	4,543,831	3,157,785	1,386,046
	acH3K9	2,536,989	2,355,671	181,318	3,054,914	2,556,881	498,033
	tmH3K27	7,722,000	6,176,152	1,545,848	8,979,461	6,392,159	2,587,302
	Total:	**37,983,462**	**7.9 X coverage***		**44,095,341**		
$\frac{piwi^1}{piwi^2}$	Non-ChIP	1,747,367	1,302,053	445,314	2,084,210	1,334,144	750,066
	HP1a	1,649,867	1,177,396	472,471	1,946,699	1,198,838	747,861
	RNA polII	2,122,762	1,687,459	435,303	3,142,038	1,846,311	1,295,727
	H3	7,595,185	5,912,409	1,682,776	8,776,206	6,055,057	2,721,149
	tmH3K4	5,941,201	4,876,957	1,064,244	6,591,340	5,020,897	1,570,443
	tmH3K9	5,816,577	4,570,579	1,245,998	6,595,148	4,686,450	1,908,698
	acH3K9	5,565,954	4,689,872	876,082	5,934,698	4,902,120	1,032,578
	tmH3K27	4,037,124	3,396,685	640,439	4,745,884	3,636,411	1,109,473
	Total:	**34,476,037**	**7.2 X coverage**		**39,816,223**		

Note*: 1 X coverage = 168,717,020 bp (total genome size) / 35 (tag length) = 4,820,486 Solexa tags

Table S2. Enriched GO terms for repressed genes in *piwi*^{1}/*piwi*^{2}, related to Figure 7.

Category	GO term	Count	%	P-value	Fold of enrichment
SP_PIR_KEYWORDS	hydrolase	243	10.2	5.20E-38	2.3
SP_PIR_KEYWORDS	oxidoreductase	127	5.3	9.90E-29	2.9
SP_PIR_KEYWORDS	membrane	181	7.6	9.90E-25	2.2
SP_PIR_KEYWORDS	metal-binding	155	6.5	3.40E-22	2.2
SP_PIR_KEYWORDS	ribosomal protein	46	1.9	2.60E-17	4.2
SP_PIR_KEYWORDS	transferase	135	5.7	1.00E-15	2
SP_PIR_KEYWORDS	ribonucleoprotein	46	1.9	3.90E-14	3.5
SP_PIR_KEYWORDS	protease	70	2.9	3.00E-13	2.6
SP_PIR_KEYWORDS	transmembrane	144	6.1	4.40E-13	1.8
SP_PIR_KEYWORDS	endoplasmic reticulum	38	1.6	1.90E-11	3.5
SP_PIR_KEYWORDS	cytoplasm	68	2.9	2.60E-11	2.4
SP_PIR_KEYWORDS	signal	78	3.3	5.20E-10	2.1
SP_PIR_KEYWORDS	zinc-finger	76	3.2	5.60E-10	2.1
SP_PIR_KEYWORDS	heme	34	1.4	5.80E-10	3.4

Table S3. Enriched GO terms for activated genes in *piwi*^{1}/*piwi*^{2}, related to Figure 7.

Category	GO term	Count	%	P-value	Fold of enrichment
SP_PIR_KEYWORDS	nucleus	34	9.6	4.20E-13	4.4
SP_PIR_KEYWORDS	dna-binding	25	7.1	5.10E-12	5.7
SP_PIR_KEYWORDS	alternative splicing	24	6.8	2.10E-10	5.1

SUPPLEMENTARY EXPERIMENTAL PROCEDURES

Primers used for ChIP-qPCR and RT-qPCR experiments are summarized below:

Ectopic Site before insertion:

Forward 5'-AATCAATCAACCGGCAAAAA

Reverse 5'- ATTGATATCGCCGATTACGC

Line 1 Original Site:

Forward 5'- CAGCCTTCTGTCCAGGGTTA

Reverse 5'- ACTGGATCCATGGCTGAAAC

Line 2 Original Site:

Forward 5'- GGGAGATGAAAATTCTGTCGAG

Reverse 5'- CGGCAAGAACAAAACGATTA

Line 1, Line 2 and Line 3 Ectopic Site:

Forward 5'- GGCAAGGGTCGAGTCGATAG

Reverse 5'- TGGTGCTATGTTTATGGCGC

Line 1 Primers for quantifying cDNA bridging insertion site upstream:

Forward 5'- GGATCTGATATCGAGGCCTGT

Reverse 5'- TTGCCTCTGGGGTGAAGG

Line 1 Primers for quantifying cDNA bridging insertion site downstream:

Forward 5'- GGATCCATGGCTGAAAAGG

Reverse 5'- TGTTCTCAACTTCAAAGGCAGA

Line 2 Primers for quantifying cDNA bridging insertion site upstream:

Forward 5'- TGACTGTGCGTTAGGTCCTG

Reverse 5'- TATTTGGTGCGGGTGAAGG

Line 2 Primers for quantifying cDNA bridging insertion site downstream:

Forward 5'- CGCACCAAATAATAAACTGAAAGG

Reverse 5'- TGTTCTCAACTTCAAAGGCAGA