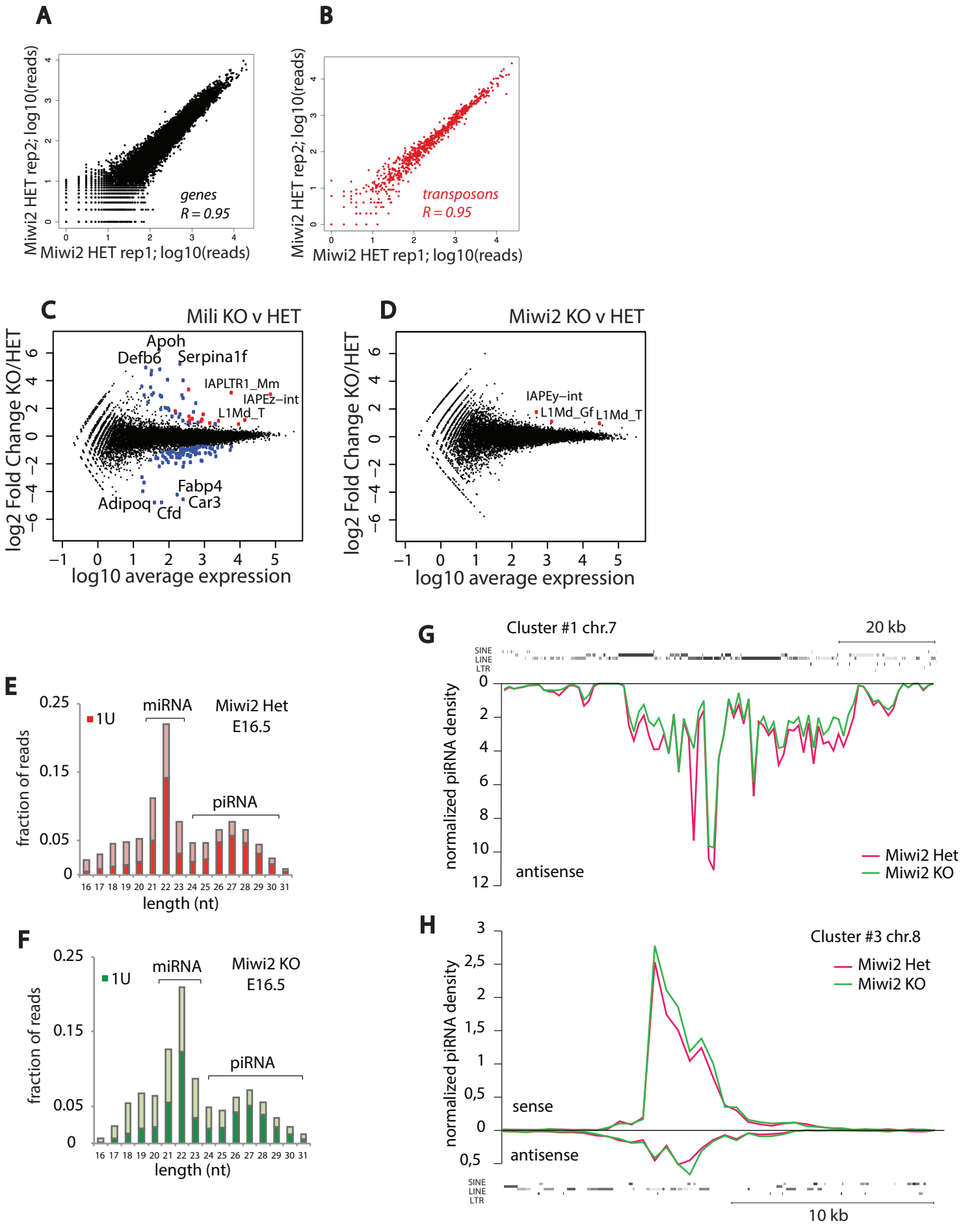
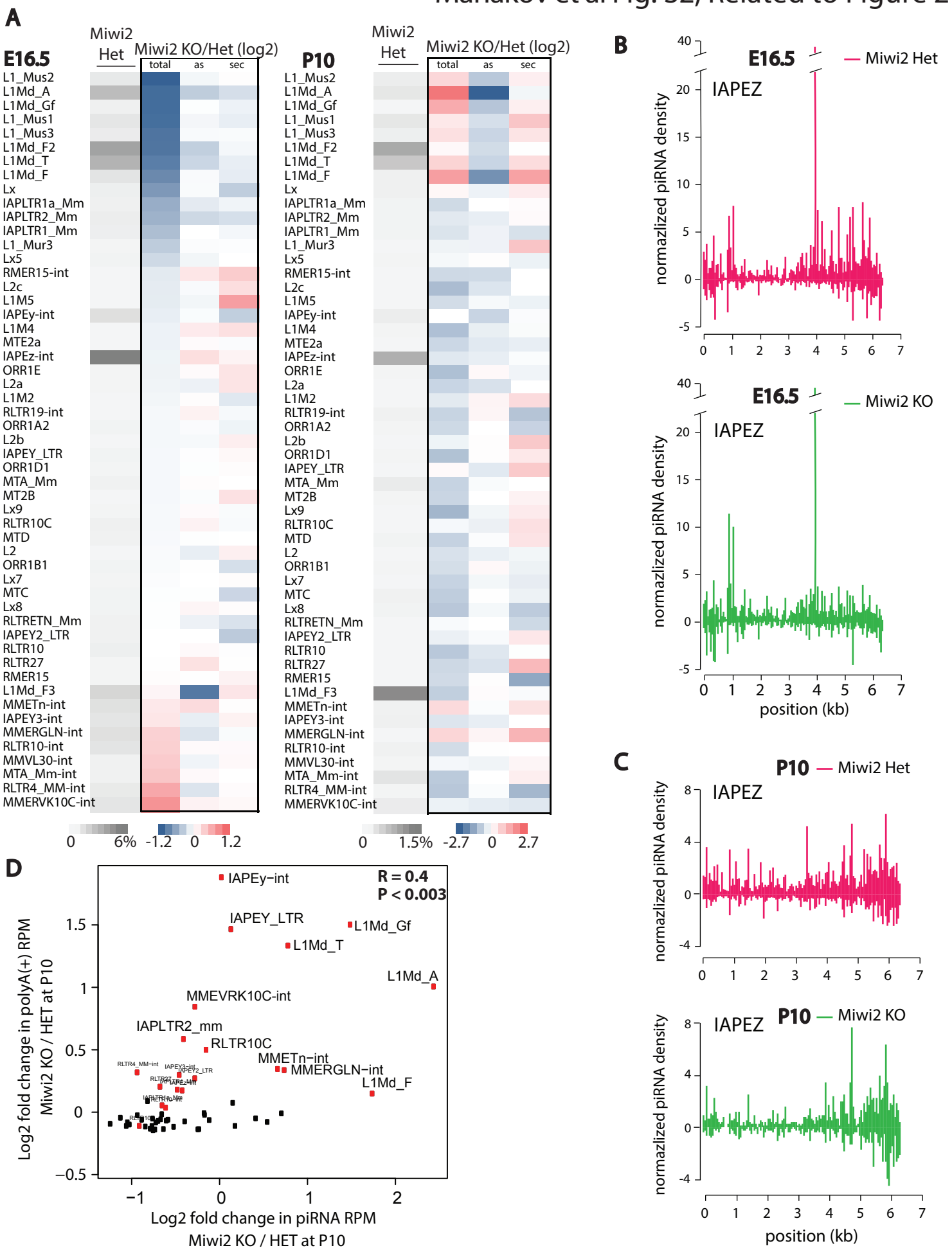
Inventory of Supplemental Information

Manakov et al Fig. S1, Related to Figure 1

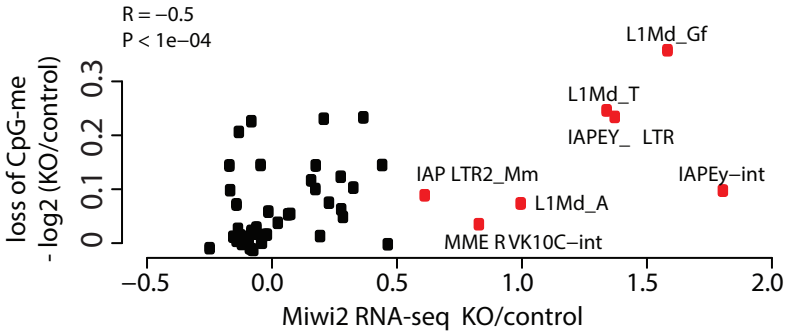


Manakov et al Fig. S2, Related to Figure 2



Manakov et al Fig. S3, Related to Figure 3

Correlation between the loss of CpG-methylation and expression of TE families in Miwi2 mutant at P10.



Manakov et al Table S1, Related to Figure 1

Library ID	Library type	Timepoint	Tissue	Genotype	Total reads	rRNA reads	unique aligned	multimapping	3'adaptor	other
DP63	poly(A)+	E16.5	testis	Miwi2 Het (control)	19,245,737	4,613,646	10,063,154	11,302,160	GATCGGAAGAGCACA	
DP74	poly(A)+	E16.5	testis	Dnmt3L Het (control)	24,408,372	6,173,251	12,159,845	14,201,684	GATCGGAAGAGCACA	
DP18	poly(A)+	P10	tesits	Miwi2 Het (control)	28,937,661	4,323,901	12,344,783	14,381,150	GATCGGAAGAGCACA	
DP19	poly(A)+	P10	tesits	Miwi2 KO	28,616,777	5,186,565	11,710,609	13,707,135	GATCGGAAGAGCACA	
DP54	poly(A)+	P10	tesits	Miwi2 Het (control)	13,748,354	1,430,788	7,366,377	8,625,092	GATCGGAAGAGCACA	
DP55	poly(A)+	P10	tesits	Miwi2 KO	20,908,192	2,133,400	11,164,525	13,060,032	GATCGGAAGAGCACA	
WP01	poly(A)+	P10	tesits	Mili Het (control)	53,163,538	908,358	35,921,310	34,876,422	GATCGGAAGAGCACA	
WP02	poly(A)+	P10	tesits	Mili Het (control)	53,667,678	709,977	31,764,016	29,350,806	GATCGGAAGAGCACA	
WP03	poly(A)+	P10	tesits	Mili KO	51,943,074	1,148,799	35,144,667	34,030,383	GATCGGAAGAGCACA	
WP04	poly(A)+	P10	tesits	Mili KO	29,349,050	498,144	17,526,051	16,466,881	GATCGGAAGAGCACA	
WP05	poly(A)+	P10	tesits	Mili KO	41,844,381	830,117	24,873,484	23,382,809	GATCGGAAGAGCACA	
DP76	small RNA	E16.5	tesits	Miwi2 Het (control)	29,122,945	6,454,940	7,591,535	13,631,188	TGGAATTCTCGGGTG	miRNA reads: 6,781,696
DP77	small RNA	E16.5	tesits	Miwi2 KO	36,284,521	11,161,013	8,417,337	14,281,448	TGGAATTCTCGGGTG	miRNA reads: 7,414,887
DP20	small RNA	P10	tesits	Miwi2 Het (control)	32,383,922	10,497,532	9,881,541	11,801,238	TGGAATTCTCGGGTG	miRNA reads: 8,772,017
DP21	small RNA	P10	tesits	Miwi2 KO	2,482,903	602,401	782,413	902,635	TGGAATTCTCGGGTG	miRNA reads: 665,991

Manakov et al Table S2, Related to Figure 1

Windows with more than 10 RPM of IP piRNAs: 48,134						
	E16 Small-RNAseq		**E16 Ribo(-) RNA-seq**		**E18 Ribo(-) RNA-seq**	
	Correlation	**P.value**	**Correlation**	**P.value**	**Correlation**	**P.value**
E16 Mili IP	0.625	< 2.2e-16	0.002	0.712	0.001	0.784
E16 Miwi2 IP (rep1)	0.503	< 2.2e-16	0.005	0.234	0.005	0.280
E16 Miwi2 IP (rep2)	0.505	< 2.2e-16	0.008	0.079	0.007	0.138
Windows with more than 10 RPM of IP piRNAs: 11,621						
	P10 Small-RNAseq		**P10 Ribo(-) RNA-seq**		**P10 Ribo(-) RNA-seq**	
	Correlation	**P.value**	**Correlation**	**P.value**	**Correlation**	**P.value**
P10 Mili IP	0.727	< 2.2e-16	0.005	0.571	0.005	0.582

Manakov et al Table S3, Related to Figure 3

	PE (76bp)	PE (76bp)	PE (101bp)	PE (101bp)	PE (101bp)	SE (76bp)	SE (100bp)	SE (100bp)
	Molaro01 MiliKO	Molaro02 MiliKO	Molaro03 MiliKO	Molaro04 WT	Molaro05 WT	Molaro06 WT	DP107 Miwi2 HET	DP111 Miwi KO
TOTAL READS IN:	62,024,524	66,804,010	70,058,532	25,278,252	56,626,875	68,945,044	191,179,374	179,902,434
GOOD BASES IN:	8,029,550,911	8,637,301,472	16,273,901,915	3,780,727,856	8,615,406,597	4,493,042,076	18,227,408,984	17,110,781,041
TOTAL READS OUT:	**59,506,166**	**64,087,369**	**47,787,400**	**23,438,647**	**50,474,141**	**59,845,378**	**156,694,873**	**146,399,127**
GOOD BASES OUT:	7,606,377,852	8,185,353,724	11,289,443,429	3,454,751,405	7,585,217,880	4,036,302,273	14,876,200,906	13,864,542,088
DUPLICATES REMOVED:	2,518,358	2,716,641	22,271,132	1,839,605	6,152,734	9,099,666	34,484,501	33,503,307
READS WITH DUPLICATES:	1,202,772	1,287,938	14,449,793	1,088,806	4,032,071	4,833,177	24,551,391	23,935,925
SITES:	1,105,337,903	1,105,337,903	1,105,337,903	1,105,337,903	1,105,337,903	1,105,337,903	1,105,337,903	1,105,337,903
SITES COVERED:	642,055,481	664,225,639	776,450,379	436,673,879	691,082,493	482,149,690	909,012,367	894,676,197
FRACTION:	**0.58**	**0.60**	**0.70**	**0.40**	**0.63**	**0.44**	**0.82**	**0.81**
MAX COVERAGE:	27,449	27,739	32,910	30,652	37,630	2,198	4,313	4,275
MEAN COVERAGE:	1.21	1.30	1.69	0.59	1.30	0.71	2.74	2.54
MEAN (WHEN > 0):	2.08	2.17	2.41	1.49	2.07	1.64	3.33	3.14
MEAN METHYLATION:	0.05	0.05	0.04	0.04	0.04	0.05	0.05	0.05
SITES:	21,908,008	21,908,008	21,908,008	21,908,008	21,908,008	21,908,008	21,908,008	21,908,008
SITES COVERED:	18,393,567	18,614,602	19,497,109	14,916,356	18,903,655	16,258,758	19,384,910	19,293,116
FRACTION:	**0.84**	**0.85**	**0.89**	**0.68**	**0.86**	**0.74**	**0.88**	**0.88**
MAX COVERAGE:	44,660	45,546	56,073	48,195	58,157	3,137	7,722	7,661
MEAN COVERAGE:	2.99	3.21	3.86	1.45	3.21	1.95	5.46	5.03
MEAN (WHEN > 0):	**3.56**	**3.78**	**4.34**	**2.14**	**3.71**	**2.62**	**6.18**	**5.71**
MEAN METHYLATION:	**0.74**	**0.74**	**0.74**	**0.78**	**0.79**	**0.80**	**0.74**	**0.74**

Supplemental Figure Legends

Figure S1. **mRNA and piRNA profiles in *Miwi2* KO and control testis. Related to Figure 1.**

A and B. Reproducibility of gene and transposon expression measurement between biological replicates. The scatterplots demonstrate high correlastion ($R \geq 0.94$, p value $\leq 2 \bullet 10^{-16}$) in pairwise comparisons of RNA-seq libraries produced from biological replicate samples of *Miwi2* heterozygote control testis (P10). The x and y-axes show log transformed raw read counts attributed to protein coding genes (> 17,000 genes shown in black) and to all annotated LINE and LTR families of TEs (~800 types, shown in red; internal and LTR fragments were analyzed separately). Replicates of other *Miwi2* Het, as well as *Miwi2* KO, *Mili* Het and *Mili* KO are similarly consistent (data not shown).

C and D. Relatively small number of genes and TE families are significantly differentially expressed in *Mili* and *Miwi2* KO testis. The points on the scatterplots correspond to either protein coding genes or LINE and LTR families of TEs (> 17000 genes and ~800 types of TEs; in ERVs, internal and LTR fragments were analyzed separately). The x-axis shows library-depth-normalized average read count (expression level) and the y-axis shows fold change of expression in *Mili* and *Miwi2* KO testis (P10) relative to corresponding control samples. Differentially expressed genes and TE families that passed the significance threshold (p value < 0.01 at FDR < 20%) are shown as the blue points for genes (124 genes in *Mili* KO, 0 genes in *Miwi2* KO) and as the red points for TE families (14 families in *Mili* KO and 3 families in *Miwi2* KO).

E and F. piRNA fraction of small RNAs is enriched in species with uridine at position 1 (U1) in *Miwi2* Het. The histograms show fractions of (E) *Miwi2* Het and (F) *Miwi2* KO E16.5 libraries by reads of different lengths (the x-axis). Fraction of reads of each length that have U1 is indicated as a colored fraction of each bar.

G and H. Expression of piRNA from major embryonic clusters is not affected by *Miwi2* deficiency. The y-axis shows a binned count of reads aligned to genomic piRNA clusters (read counts are normalized to count of miRNAs in respective libraries) from *Miwi2* Het and Miwi2 KO embryonic testis. (G) Expression of piRNA from uni-directional, TE-rich cluster #1 on chromosome 7 is not affected in *Miwi2* KO at E16.5. (H) Expression of piRNA from bi-directional, TE-poor cluster #3 on chromosome 8 is not affected by *Miwi2* KO at E16.5.

Figure S2. Comparison of piRNA expression in *Miwi2* KO and control testis. Related to Figure 2.

A. piRNA from young L1 families are affected by *Miwi2* KO. The heat map shows expression of piRNA of selected TE families, overall piRNA abundance ("total"), fraction of antisense sequences ("as"), and fraction of secondary piRNA ("sec") in (A) E16.5 prospermatogonia and (B) P10 spermatogonia from *Miwi2* KO and *Miwi2* Het mice. Secondary piRNAs are defined as reads that have adenine at position 10 and do not have uridine residue at position one. The "total" column shows that abundance of piRNAs targeting several L1 families are ~2 fold decreased in *Miwi2* KO relative to *Miwi2* Het mice. The "as" and "sec" columns demonstrates that at E16 the fraction of libraries comprised of antisense and secondary piRNAs did not significantly differ between Miwi2 KO and control mice. At P10 the total abundance of L1 piRNAs is increased in *Miwi2* KO mice, whereas the antisense fraction is decreased. Ratios are represented on a log2 scale.

B and C. IAPEz-derived piRNA are not affected by *Miwi2* KO. piRNA densities (read counts in 50-nt windows) over the IAPEz consensus in *Miwi2* Het and *Miwi2* KO libraries from embryonic testes (E16.5, B) and postnatal spermatogonia (P10, C). Each bar represents a 5' end of a piRNA mapping to either sense or antisense strand of the element. The y-axis shows read counts normalized to total count of miRNAs in respective libraries. Depth of libraries and miRNA counts are shown in **TableS 1**.

D. Correlation in induction of TE expression and upregulation of piRNAs at P10. The x-axis shows fold change of piRNA read counts (normalized to counts of miRNA reads in the respective libraries, RPM) corresponding to selected families of LINE and LTR transposons (as in Figure 1A) in *Miwi2* KO and *Miwi2* Het total small RNA libraries at P10. The y-axis shows the fold change of long RNAs reads (cloned with polyA(+) protocol) in the same cells. Long-RNA-seq is normalized to the total number of reads mapped in the respective libraries (RPM).

Figure S3. Related to Figure 3. Correlation between the loss of CpG-methylation and expression of TE families in *Miwi2* mutant at P10

The y-axis shows a negative value of the log2 ratio of the mean CpG methylation level on selected TE families (as in Figure 3C) in *Miwi2* KO and control samples. The x-axis shows log2 ratio of TE transcript abundance of the same families. The distribution demonstrates the significant correlation ($p < 0.0001$, evaluated between fold changes on a linear scale) between upregulation of expression and loss of CpG methylation in *Miwi2* KO mice. TEs selected for Figure 3C and this plot have at least 50 genomic copies that passed the CpG detection threshold (each copy was required to harbor ten or more consensus CpG sites and to be detected by ten or more sequencing events in both mutant and control libraries).

Supplemental Table Legends

Table S1. RNA-seq and ChIP-seq libraries, related to Figure 1

Parameters of polyA+ and small RNA libraries

Table S2. Correlation between small RNA libraries, related to Figure 1

The table lists correlation of normalized read abundance (RPM) of small RNAs cloned from MILI and MIWI2 immonoprecipitates (IP), total small RNA and long RNA-seq reads in 1 kB genomic windows. Correlation between MILI and MIWI2 IP piRNAs and total small RNA libraries is significant at both E10 and P10, while correlation with long RNA-seq reads is not significant.

Table S3. BS-seq libraries, related to Figure 3

Parameters of BS-seq libraries

Supplemental experimental procedures, related to Experimental Procedures

Animals

Miwi2 knockout mouse model is described in (Kuramochi-Miyagawa, 2008). *Mili* knockout model is described in (Kuramochi-Miyagawa et al., 2004). All experiments were conducted in accordance with a protocol approved by the IACUC of the California Institute of Technology. Animals were group-housed, unless otherwise mentioned, at 22-24°C with *ad libitum* access to food and water in a 13-h/11-h light/dark cycle, in an AAALAC accredited housing facility. Mice were euthanized according to IACUC approved procedures.

RNA isolation, polyA+ selection, rRNA depletion, and RNA-Seq library cloning

Total RNA was isolated from total testis using Ribozol (Amresco, N580). For isolation of RNA from embryonic testes, embryos from several pregnant mothers were harvested, genotyped, and pooled according to genotype for RNA isolation. All P10 samples are either from individual animals, or from littermates. RNA was DNase-treated using Turbo DNA-free kit (Ambion, AM1907). For library cloning, 1 µg DNase-treated RNA was polyA+ selected using DynaBeads mRNA Purification Kit (Invitrogen, cat.no. 610.06). RNA-Seq libraries were cloned using either Illumina TruSeq RNA sample prep kit (version 2, 48 rxns, RS-122-2001) or NEBNext Ultra RNA library prep kit for Illumina (New England Biolabs, E7530). Cloning of small RNAs is described in a separate section.

Genomic DNA isolation, bisulfite conversion, and library cloning

Cells were isolated from testes as described in Pezic et al (2014), resuspended in 100 µl of TE, and lysed in 1 ml of genomic DNA lysis buffer (10 mM Tris-HCl pH 8.0, 100 mM EDTA pH 8.0, 0.5% SDS, supplemented with 250 U RNaseI (Ambion, AM2295)). Lysate was incubated at 37°C for 1 hour, transferred to a glass douncer and homogenized further by 15 strokes with tight pestle. Proteinase K (NEB) was added to final concentration of 100 µg/ml, and incubated at 50°C for 3 hours. After cooling to RT, equal volume of phenol pH 8.0 was added, mixed well, and incubated for 10 min at RT with agitation. Water phase was recovered after centrifugation, and phenol extraction repeated one more time, followed by phenol/chloroform and chloroform extraction. Water phase was then mixed with 2 volumes of absolute ethanol, with addition of 5M NaCl to final concentration of 300 mM. DNA was precipitated for 2 hours at -80°C, or overnight at -20°C. Pelleted DNA was washed twice with 70% EtOH, dried, and resuspended in water. A260/280 and A260/230 ratios were measured to check purity. Genomic DNA (200 ng) was sonicated using Bioruptor (Diagenode) in 100ul 1X TE, and libraries cloned using NuGEN Ovation UltraLow MethylSeq Multiplex System 1-8 kit (NuGEN 0335-32)

according to manufacturer's instructions. Bisulfite conversion of DNA is performed after adapter ligation using this kit, and it was done using Active Motif MethylDetector Kit (#55001) as instructed by the manufacturer's protocol.

Small RNA cloning

Total RNA was isolated from embryonic testes using Ribozol. Thirty ug total RNA was loaded on 12% urea-polyacrylamide (PAA) gel. Fraction 19-30 nt was excised and snap-frozen in liquid nitrogen in 400 ul 0.4M NaCl. RNA was eluted from the gel over night at 16C, shaking at 1,000 rpm and precipitated in 3 volumes of absolute ethanol with added GlycoBlue. Pre-adenylated 3' linker (/5`Phos/TGGAATTCTCGGGTGCCAAGGAACTC/3`ddC/; 5'DNA adenylation kit, NEB) was ligated to RNA over night at 4C using Truncated RNA Ligase 2 (NEB). Ligation reactions were loaded onto 10% urea-PAGE, 45-56 nt fraction was excised and nucleic acids extracted as above. 5' linker (rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC) was ligated to the samples using RNA Ligase 1 (NEB) over night at 4C. Ligation reactions were loaded on 10% Urea-PAA gel, 72-83 nt fraction was excised and nucleic acids as above. Extracted samples were reverse-transcribed (primer sequence: GGAGTTCCTTGGCACCCGAGA) and library amplified by PCR using standard Illumina primers. Final libraries were excised from 1.5% agarose gel, and sequenced on Illumina platform.

Bioinformatics analysis of gene and transposon expression

We used of polyA(+) non-stranded protocol to generate RNA-seq libraries. Trimming of Illumina adaptors with Reaper was the first step in library processing (Davis et al., 2013). Next, library reads were first aligned to sequences of rRNA transcripts (GenBank identifiers: 18S, NR_003278.3; 28S, NR_003279.1; 5S, D14832.1; and 5.8S, K01367.1) with Bowtie 0.12.7 with up to three mismatches (Langmead et al., 2009). Reads with valid alignments to rRNA were considered contamination and removed from father analysis. The remaining reads were mapped to mm10 reference genome either in unique or multimapping mode. In the unique alignment mode reads with more than one valid alignment to the genome were discarded, while in the multimapping mode up to 10,000 valid alignments were allowed.

The alignments in the unique mode were used to evaluate expression of protein coding genes, which we approximated as a sum of reads with 5' coordinates intersecting gene exonic features. Counts of reads mapped to exons of overlapping genes were evenly dived between corresponding genes, and reads overlapping with small non-coding RNAs (e.g. miRNAs, snoRNAs, tRNA etc.) were excluded. RefSeq transcriptome annotation (Pruitt et al., 2009) was used to define exonic features of protein

coding and non-coding RNAs, including small non-coding RNAs. Annotation of small non-coding RNAs was manually curated and supplemented with species annotated in Rfam (Nawrocki et al., 2015), miRBase (Kozomara and Griffiths-Jones, 2014), RepeatMasker (http://www.repeatmasker.org) of mm10 genome as available via UCSC Genome Bioinformatics Site (Karolchik et al., 2014) and Ensembl (Cunningham et al., 2015). Reads overlapping with small non-coding RNAs were excluded from the analysis of protein coding gene expression. Custom software (for assembly of the genome annotation and for per-gene read counting), the software documentation as well as the description of the curation procedure is hosted on GitHub (https://github.com/getopt/EXPRESSION_BY_FEATURE).

The alignments in the multimmaping mode were used to evaluate expression of TE families. Annotation of individual TE copies and families were taken from RepeatMasker (http://www.repeatmasker.org) annotation of mm10 genome as available via UCSC Genome Bioinformatics Site (Karolchik et al., 2014). Per-family read count was computed as previously described (Pezic et al., 2014). Briefly, reads with 5' coordinates intersecting individual copies of TEs in a family were totaled up weighing each read alignment instance by the total number of alignments of that read.

DESeq procedure was used to evaluate statistical significance of the fold change in expression levels between knockout and control samples (Anders and Huber, 2010). We assumed that the error in measurements of gene and TE family expression are due to the same underlying technical and biological variability and therefore follow the same distribution. This allowed us to combine per-gene and per-family read counts to model dispersion and evaluate statistical significance of expression fold changes with DESeq. We implemented DESeq procedure via its R (v 3.1.2) interface as described in DESeq documentation. DESeq parameter "sharing" was set to "fit-only" and genes and TE families with 0 read count in either knockout or control samples were excluded from the test.

Bioinformatics analysis of bisulfite sequencing libraries to evaluate CpG-methylation

We used the same computational procedure to analyze Miwi2 knockout and heterozygote control libraries that we generated and previously published Mili knockout and wild type control libraries (Molaro et al., 2014). First, Illumina adaptors were trimmed using Trim Galor! tool (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Trimmed reads were then aligned to mm10 genome with Bismark tool (Krueger and Andrews, 2011). Default Bismark parameters and Bowtie2 (v 2.2.2) engine (Langmead and Salzberg, 2012) were used to align both single-end and paired-end libraries. Following the alignment, mean CpG-methylation level of individual CpG cites and of regions of interests (e.g. TE copies or fragments of TEs) was determined using MethPipe

computational tool suite (Song et al., 2013). Since only uniquely aligned reads were considered, repetitive regions in the genome may have a detection bias toward paired-end libraries. Therefore we limited our analysis to those CpG sites that were covered in the most shallow single-end library (~19 million sites or 88% of genomic sites) which we refer to as “consensus” CpG sites. Reads aligned to the same genomic position were collapsed and replicate samples were merged prior to evaluation of the mean CpG methylation of individual CpG sites and TE copies. To evaluate mean CpG methylation of TE families we considered only those TE copies that contained at least ten consensus CpG sites and at detected by at least ten sequencing events.

To profile CpG methylation along sequence of individual TE copies of L1-A and IAPEz elements (Figure 4D) we followed the same procedure as described above for genome wide analysis, but instead of alignment to mm10 genome the reads were aligned to sequences of a copy of L1-A element (mm10 chr1:164723807-164730175) and a fusion of IAPLTR1_mm (mm10 chr8:121405292-121405674) and IAPEz-int (mm10 chr8:121405675-121412154) elements. These TE copies were manually selected on the basis of their full length and low divergence from consensus according to RepeatMasker track in the UCSC Genome Browser (Kent et al., 2002). Even though a complete IAPEz provirus normally consists of a single internal part flanked by two LTR sequences, we have truncated the right flanking LTR to 65 nt (i.e. shorter than the length of sequencing reads) in order to avoid redundant read alignments since reads with multiple valid alignments are filtered out by Bismark.

Bioinformatics analysis of MIWI2 and MILI IP and total small RNA libraries

Illumina adaptors were removed using custom developed dynamic programming algorithm (Olson et al., 2008). Next, custom suffix array procedure was used to align reads longer than 24 nt to mm9 genome (Olson et al., 2008). Reads overlapping with TEs were identified by the overlap of alignments with RepeatMasker (http://www.repeatmasker.org) TE annotation of mouse genome as available through UCSC Genome Bioinformatics Database (Karolchik et al., 2014). Reads with multiple alignments to different types of TEs were annotated following the majority rule based on ten alignments selected at random. When ties were encountered, they were resolved following hierarchy principle (Olson et al., 2008).

Supplemental References

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

Cunningham, F., Amode, R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S.*, et al.* (2015). Ensembl 2015. Nucleic acids research *43*, D662-D669.

Davis, M., van Dongen, S., Abreu-Goodger, C., Bartonicek, N., and Enright, A. (2013). Kraken: a set of tools for quality control and analysis of high-throughput sequence data. Methods (San Diego, Calif) *63*, 41-49.

Karolchik, D., Barber, G., Casper, J., Clawson, H., Cline, M., Diekhans, M., Dreszer, T., Fujita, P., Guruvadoo, L., Haeussler, M.*, et al.* (2014). The UCSC Genome Browser database: 2014 update. Nucleic acids research *42*.

Kent, J., Sugnet, C., Furey, T., Roskin, K., Pringle, T., Zahler, A., and Haussler, D. (2002). The human genome browser at UCSC. Genome research *12*, 996-1006.

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic acids research *42*, D68-D73.

Krueger, F., and Andrews, S. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics (Oxford, England) *27*, 1571-1572.

Kuramochi-Miyagawa, S. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes & Development *22*, 908-917.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W.*, et al.* (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development (Cambridge, England) *131*, 839-849.

Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

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

Molaro, A., Falciatori, I., Hodges, E., Aravin, A., Marran, K., Rafii, S., McCombie, R., Smith, A., and Hannon, G. (2014). Two waves of de novo methylation during mouse germ cell development. Genes & development *28*, 1544-1549.

Nawrocki, E., Burge, S., Bateman, A., Daub, J., Eberhardt, R., Eddy, S., Floden, E., Gardner, P., Jones, T., Tate, J.*, et al.* (2015). Rfam 12.0: updates to the RNA families database. Nucleic acids research *43*.

Olson, A.J., Brennecke, J., Aravin, A.A., Hannon, G.J., and Sachidanandam, R. (2008). Analysis of large-scale sequencing of small RNAs. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing, 126-136.
Pezic, D., Manakov, S., Sachidanandam, R., and Aravin, A. (2014). piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes & development.
Pruitt, K., Tatusova, T., Klimke, W., and Maglott, D. (2009). NCBI Reference Sequences: current status, policy and new initiatives. Nucleic acids research *37*, D32-D36.
Song, Q., Decato, B., Hong, E., Zhou, M., Fang, F., Qu, J., Garvin, T., Kessler, M., Zhou, J., and Smith, A. (2013). A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. PloS one *8*.