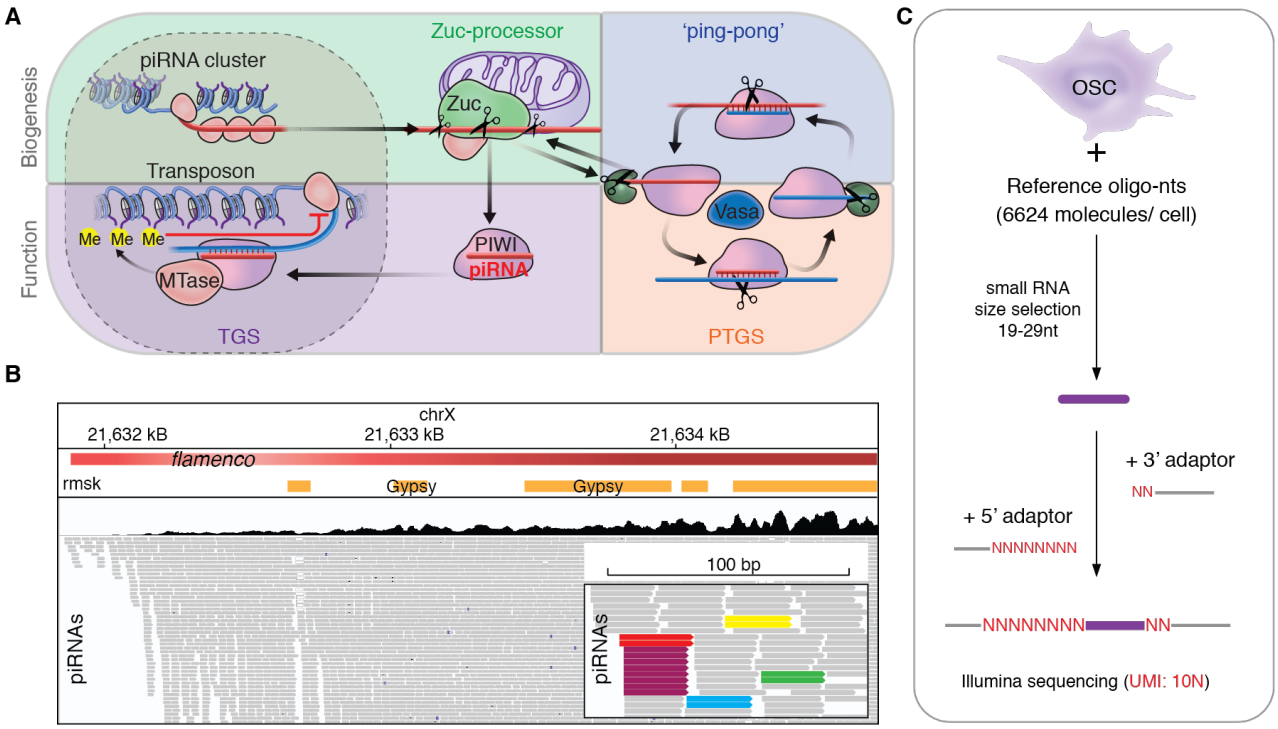
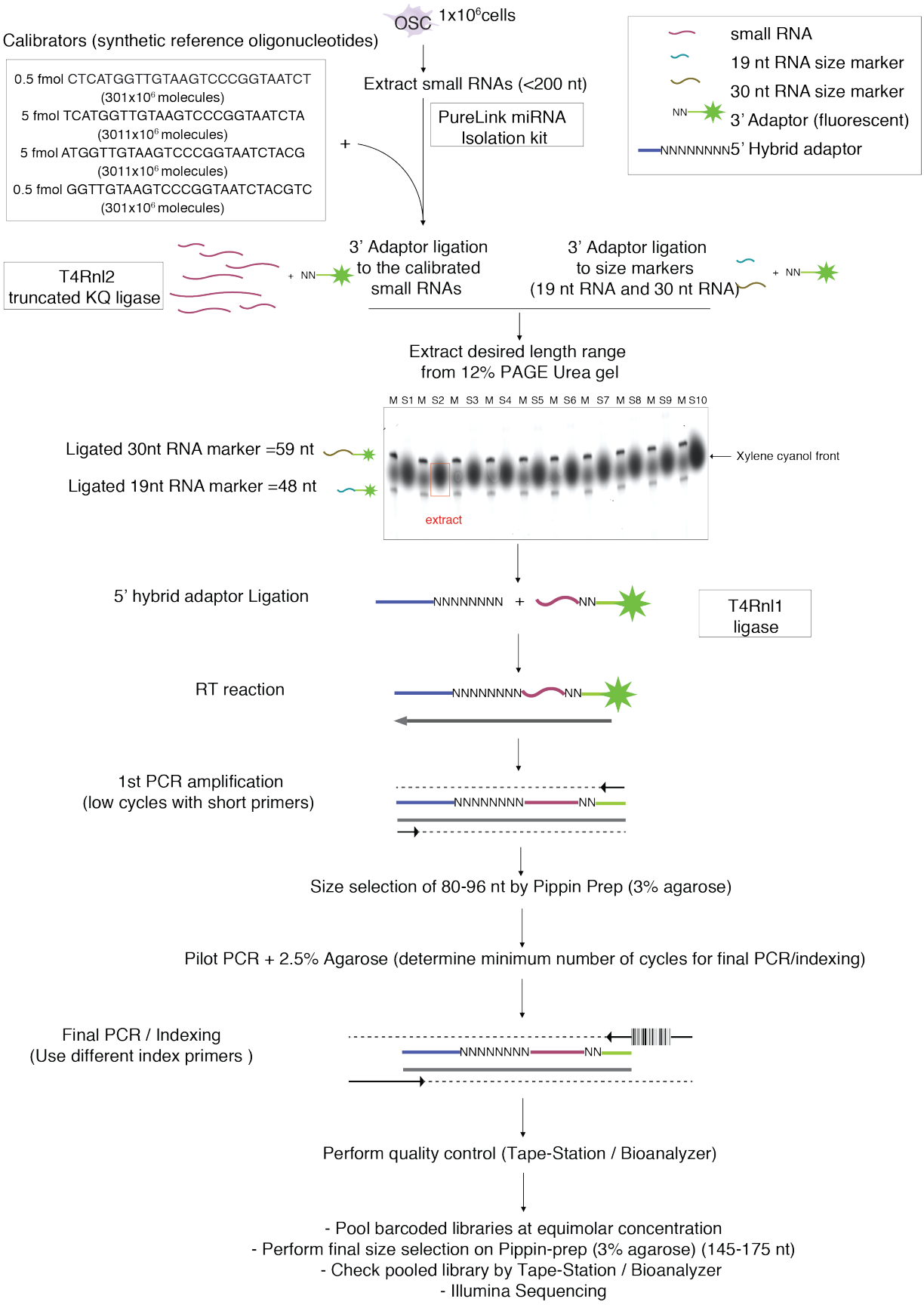
Supplemental Material

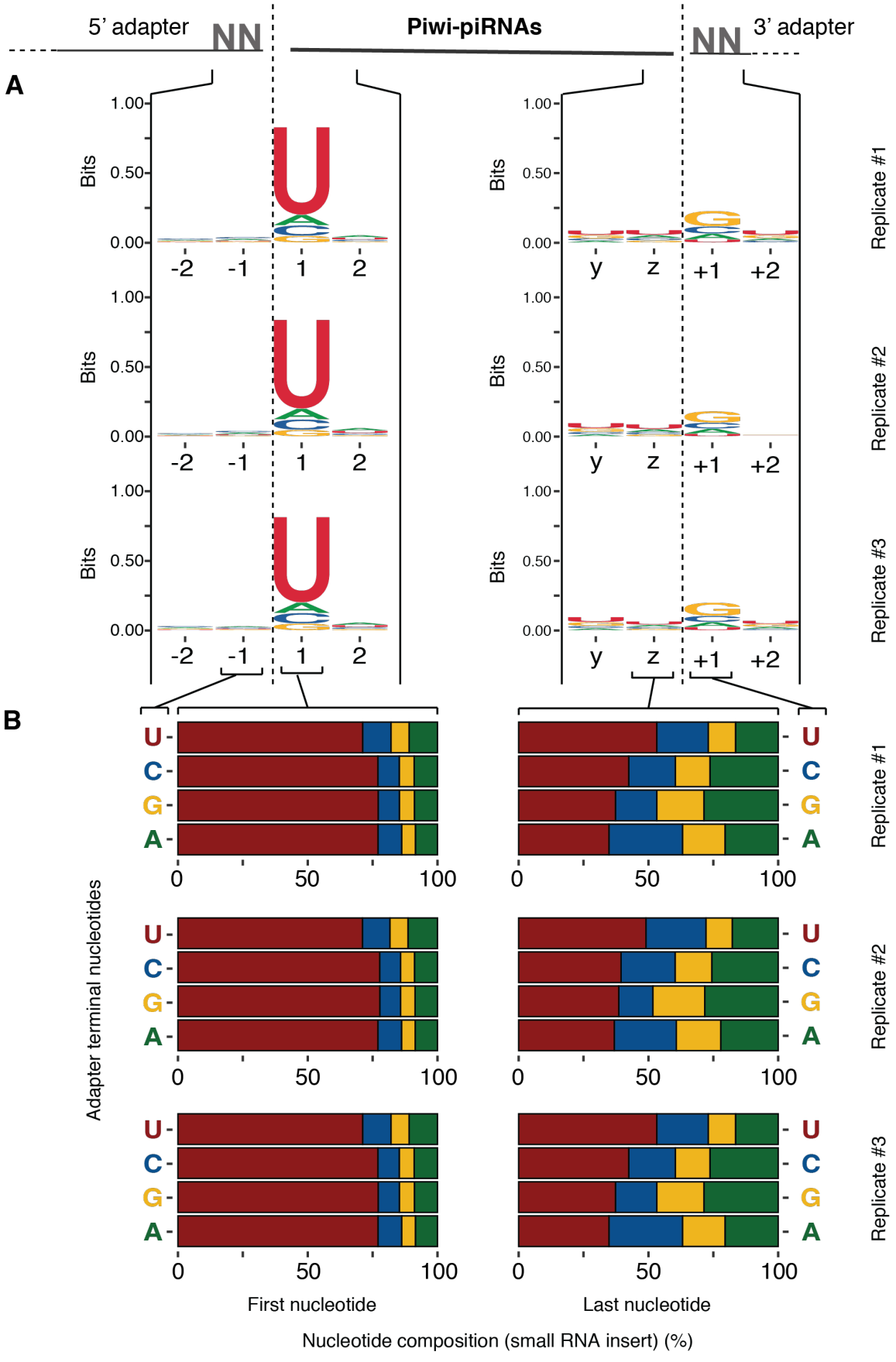
Supplemental Figures 1-7



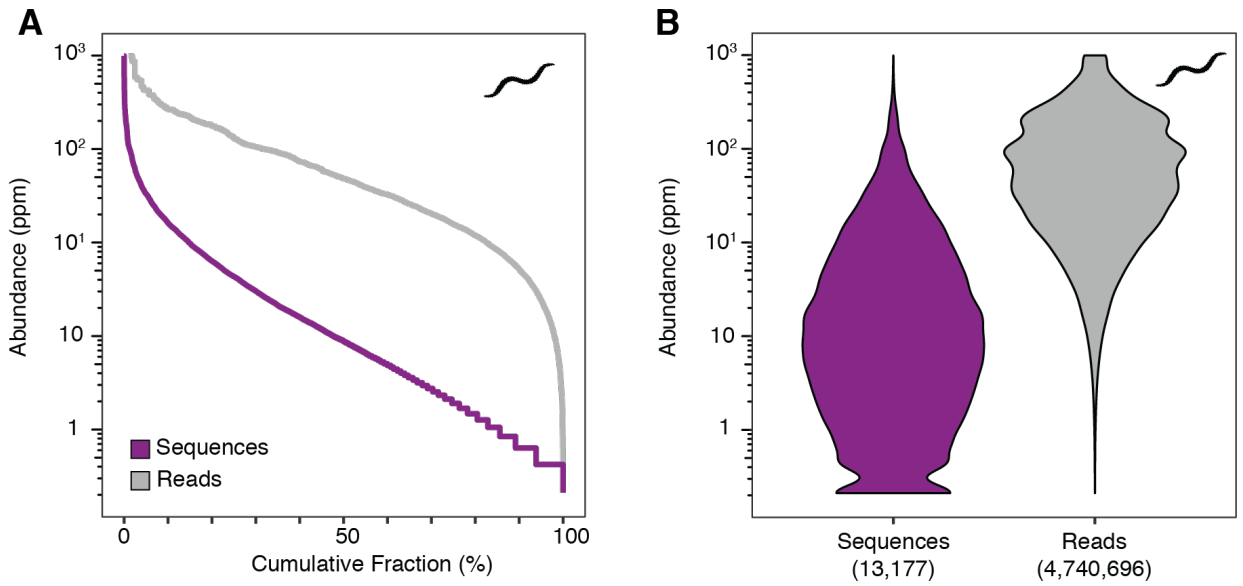
Supplemental Fig. S1. Illustration of piRNA biogenesis and sample preparation. (A) Schematic representation of conserved piRNA mechanisms at the example of piRNA silencing in the *Drosophila* ovary. PiRNA generating regions, piRNA clusters, produce long precursor transcripts that are sliced by the endonuclease Zucchini (Zuc) and loaded into PIWI proteins. The mature PIWI-piRNA complexes (piRISC) target transposon transcripts by complementary base-pairing. In the nucleus, piRISC recruits methyltransferases to induce lasting epigenetic restriction of transposons (TGS … transcriptional gene silencing). In the cytoplasm, piRISC cleaves target RNAs which results in post-transcriptional silencing (PTGS). Cytoplasmic RNA cleavage can result in generation of additional piRNAs in a feed-forward mechanism termed the 'ping-pong' cycle that is coordinated by the helicase VASA/DDx4. Ping-pong generated piRNAs can in turn inform piRNA processing by the ZUC-processor complex. **(B)** Genome track of Piwi-piRNAs that originate from *flamenco* (Flam) in OSC. The flamenco region contains a high density of dysfunct transposon fragments (rmsk … repeat masker). The single stranded Flam transcript captures most transposon insertions in antisense orientation and is processed into millions of diverse piRNA sequences by the endonuclease Zucchini (Zuc). PiRNA processing consumes the precursor and results in small RNAs that carry information about transposon sequences. **(C)** Schematic representation of piRNA quantification using spike-in calibrator oligonucleotides (reference oligo-nts) and small RNA sequencing in ovarian somatic sheath cell (OSC). Total RNA from one million cells was calibrated with a known number of synthetic reference RNA oligonucleotides of varying sequence and concentration (see methods). After size selection (19-29nt) and stepwise addition of adapter sequences that contained ten unique molecular identifiers (UMI), sequencing data were generated on an Illumina NextSeq platform. Our final estimate is based on eight biological replicates (n=8).


OSC
1x10⁶cells
Calibrators (synthetic reference oligonucleotides)
0.5 fmol CTCATGGTTGTAAGTCCCGGTAATCT (301x10⁶ molecules)
5 fmol TCATGGTTGTAAGTCCCGGTAATCTA (3011x10⁶ molecules)
5 fmol ATGGTTGTAAGTCCCGGTAATCTACG (3011x10⁶ molecules)
0.5 fmol GGTTGTAAGTCCCGGTAATCTACGTC (301x10⁶ molecules)
Extract small RNAs (<200 nt)
PureLink miRNA Isolation kit
small RNA
19 nt RNA size marker
30 nt RNA size marker
NN
3' Adaptor (fluorescent)
NNNNNNNN
5' Hybrid adaptor
T4Rnl2 truncated KQ ligase
3' Adaptor ligation to the calibrated small RNAs
3' Adaptor ligation to size markers (19 nt RNA and 30 nt RNA)
Extract desired length range from 12% PAGE Urea gel
M S1 M S2 M S3 M S4 M S5 M S6 M S7 M S8 M S9 M S10
Ligated 30nt RNA marker =59 nt
Ligated 19nt RNA marker =48 nt
Xylene cyanol front
extract
5' hybrid adaptor Ligation
T4Rnl1 ligase
RT reaction
1st PCR amplification (low cycles with short primers)
Size selection of 80-96 nt by Pippin Prep (3% agarose)
Pilot PCR + 2.5% Agarose (determine minimum number of cycles for final PCR/indexing)
Final PCR / Indexing (Use different index primers)
Perform quality control (Tape-Station / Bioanalyzer)
- Pool barcoded libraries at equimolar concentration
- Perform final size selection on Pippin-prep (3% agarose) (145-175 nt)
- Check pooled library by Tape-Station / Bioanalyzer
- Illumina Sequencing

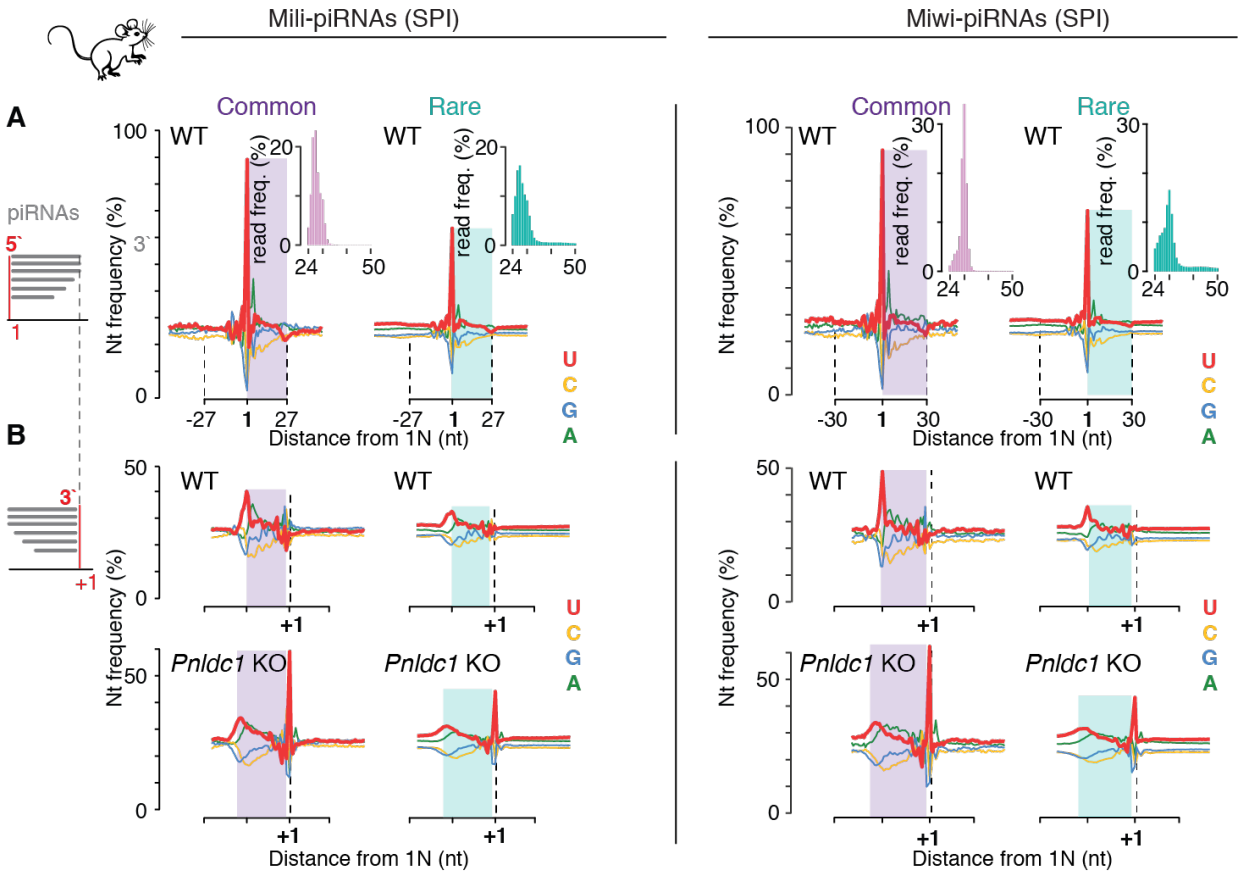
Supplemental Fig. S2. Preparation of calibrated of small RNA samples for sequencing. Construction of quantifiable small RNA libraries from ovarian somatic sheath cell (OSC). Small RNAs (<200 nt) were extracted from $1x10^6$ cells and were spiked with known amounts of 4 different reference oligonucleotides (calibrators, 26 nt each). Samples were ligated to a 3' adaptor (29 nt long) bearing 2 UMIs (variable nucleotides, N) in its 3' end and were size selected (48-58nt) through 12% Urea PAGE. Extracted RNA was then ligated to a 5' adaptor (34 nt long) bearing 8 UMIs (variable nucleotides, N) at its 3' end resulting in a total of 10 UMIs per cloned small RNA. Ligated RNA sequences were then reverse transcribed and amplified. After an initial low-cycle PCR, the products were size selected (80-96nt) using Pippin Prep to remove adaptor-adaptor byproducts. The recovered DNA was barcoded (3^{rd} read barcodes) during a second low-cycle PCR. Samples were evaluated using TapeStation (Agilent). Finally, the pooled samples were run on a Pippin Prep to select for 145-175nt (This second size selection is not need if individual libraries after barcoding do not contain adapter-adapter byproducts). Sequencing data were generated on an Illumina NextSeq platform. Our final estimate is based on eight biological replicates (n=8).


5' adapter
NN
Piwi-piRNAs
NN
3' adapter
A
B
Bits
1.00
0.50
0.00
-2
-1
1
2
y
z
+1
+2
Replicate #1
Replicate #2
Replicate #3
U
C
G
A
0
50
100
Adapter terminal nucleotides
First nucleotide
Last nucleotide
Nucleotide composition (small RNA insert) (%)

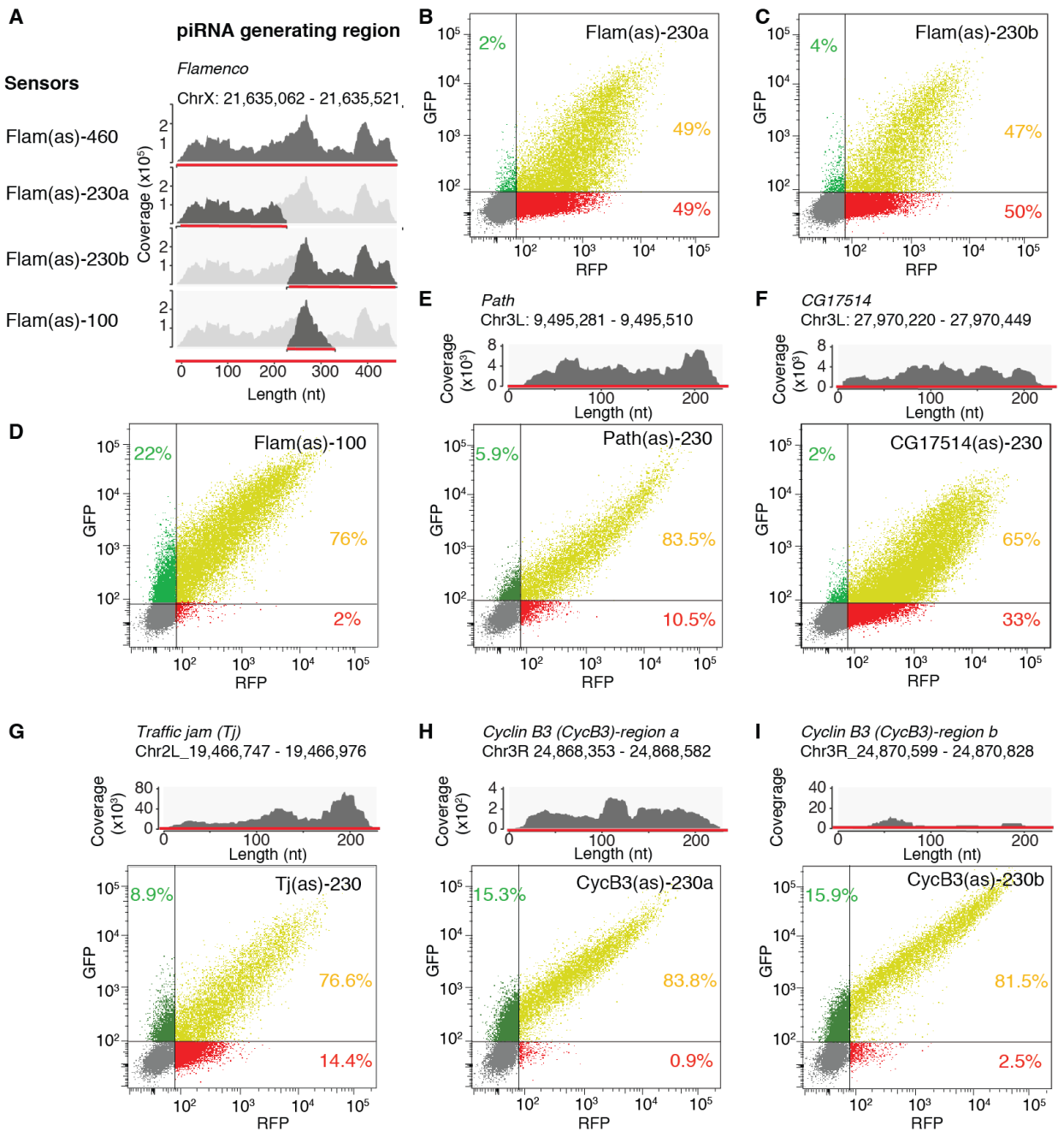
Supplemental Figure S3. Quality control for potential sequence specific artifacts during sample preparation: Randomized adapter-terminal nucleotides, unique molecular identifiers (UMI), and biological replicates. Piwi-piRNA were prepared for sequencing using adapters with random nucleotides (N) at their ligating termini to minimize potential sequence preferences during adapter ligation. Raw reads were trimmed of their constant adapter sequences and collapsed to remove PCR duplicates (using 10 UMIs). Analyses are shown for three biological replicates. (**A**) Sequence logos (ggseqlogo) for a 4nt sequence space surrounding the 5' and 3' ligation sites. The 1U-preference of Piwi-piRNAs (position 1) is not related to any sequence preference at position -1 (5'adapter: ligation-terminal nucleotide). Minimal sequence preferences are visible at the 3' ligation site. We speculate that the very slight avoidance of U in the +1 position (3'adapter: ligation-terminal nucleotide) might reflect a preference against homopolymers (runs of Us) during sequencing. Given the (A)/U-richness of piRNA-producing genomic intervals, A and U might result in homopolymers to a greater extend and results in a slight disadvantage in representation during sequencing. (**B**) Reads were split into groups based on the identity of the adapter-terminal nucleotides (5'adapter: ligation-terminal nucleotide = position -1; 3'adapter: ligation-temrinal nucleotide = position +1). The composition of the first and last nucleotide of the piRNA-insert are shown. The 1U-preference of Piwi-piRNAs can be observed with all possible adapter terminal nucleotides. [The slight reduction of the 1U fraction in 5' ligation events with Uridine (U) as adapter terminal nucleotide, might result from a preference of the system against homopolymers.] The general (A)/U-richness of piRNAs can be observed at their last nucleotides (position y and z). Slight variabilities can be observed with individual adapter-terminal nucleotides and between biological replicates. Our data support the current recommendation for randomized adapter-terminal nucleotides to minimize potential sequence specific effects, the use of unique molecular identifiers to correct for PCR-duplications and ensure accurate quantification, and the importance of biological replicates.



Supplemental Figure S4. Worm PRG1-piRNAs (21U-RNAs) are variable in sequence abundance but less skewed then fly and mouse piRNAs. In C. elegans, the main PIWI-clade Argonaute protein PRG1 associates with 21U-RNAs. Mechanisms of piRNA biogenesis in worms are not related to the mechanisms conserved in flies and mice . Each worm piRNA is determined by an individual transcriptional unit and produces a piRNA of defined sequence without sequence overlap with another piRNAs. PRG1-piRNAs were used from publicly available data (SRR538357), aligned to the C. elegans genome, and annotated using WS280/PRJNA13758. Sequence abundance was calculated as for fly and mouse piRNAs (Fig. 2). (**A**) Individual PRG1-piRNAs (21U-RNAs) were ranked by their abundance in parts per million (ppm). The cumulative fractions of sequences (purple) and reads (grey) are depicted. (**B**) Violin plots depict the distribution of sequence and reads with respect to their sequence abundance. The total number of sequences and reads in this data set are indicated.

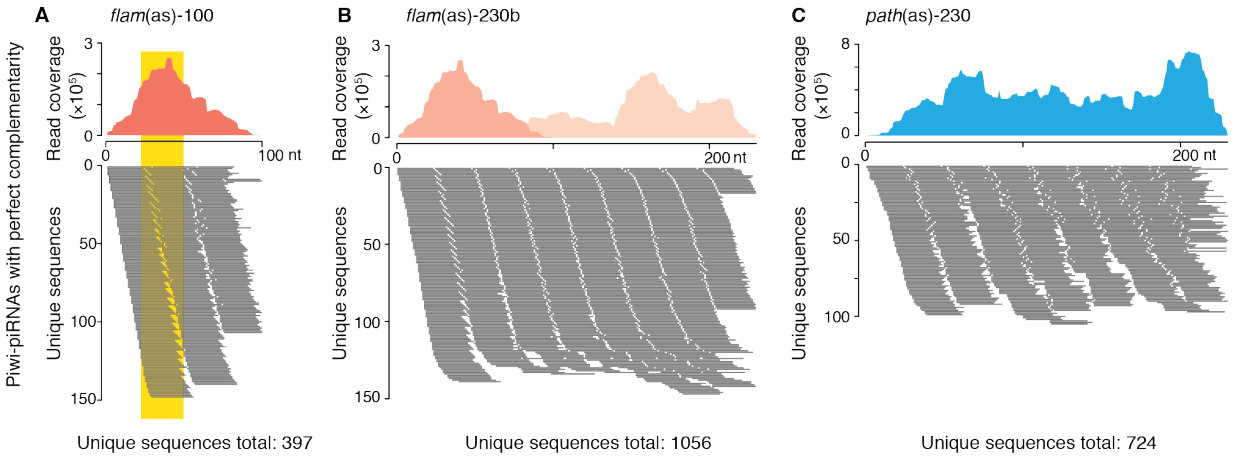


Supplemental Fig. S5. 5' and 3' end processing signatures of Mili- and Miwi-piRNAs in primary spermatocytes (SPI). Common and rare piRNAs exhibit similar processing signatures suggesting that they are generated by the same biogenesis machinery. Publicly available source data: SRA: PRJNA421205. **(A)** Common and rare piRNAs exhibit a preference for Uridine in the 5' most position (1U). Metagene analysis of uniquely mapping piRNAs that were aligned at their 5' end. The observed piRNAs are indicated by a colored box. An extended genomic interval is depicted. Insert: length distribution of common and rare piRNAs. **(B)** In mouse, the 3' ends of piRNAs are fist processed by enonucleolytic cleavage and then trimmed by the 3' to 5' exonuclease Pnldc1. Common and rare piRNAs show a preference for Uridine in the genomic position following the piRNA 3' end that only becomes visible upon knock-out (KO) of the exonuclease Plndc1 (+1U). Metagene analysis of uniquely mapping piRNAs that were aligned at their 3' end. The observed piRNAs are indicated by a colored box. An extended genomic interval is depicted.



Supplemental Fig. S6. Dual-fluorescence reporter assay for Piwi-piRNA silencing in ovarian somatic sheath cells (OSC). The reporter contained target sites for endogenous piRNAs in the 3'UTR of GFP. mCherry was expressed as non-targeted control from the same plasmid (as in Fig. 4). GFP reports on targeting by endogenous piRNAs that reduces the expression of GFP and results in the appearance of 'red only' cells. The fraction of green, yellow (cells expressing red and green) and red cells is shown in % of all transfected cells. **(A)** PiRNA sensors with antisense complementarity to different regions of *flamenco* (*lncRNA:flam*) (*flam*) (length (-n) in nucleotides (nt)). Endogenous piRNAs that originate from the corresponding fragments of *flamenco* (*lncRNA:flam*) (*flam*) are shown as coverage track. **(B-D),** The strength of piRNA silencing -indicated by the fraction of 'red-only', fully silenced cells- varies between different reporter constructs. Flow cytometry analysis of ovarian somatic sheath cells that were transfected with varying *flamenco* (*lncRNA:flam*) (*flam*) sensors (indicated in **A**). Cells were analyzed 48 hours after transfection. **(E-I)** Sensors targeted by piRNAs produced from *pathetic* (*path*) (**E**), CG17514 (**F**), *traffic jam* (*tj*) (**G**),

and two different intervals of *cyclin B3* (*cycB3*) (**H, I**). Production of endogenous piRNAs by the sensed region is indicated as coverage track (top panel). Cells expressing the corresponding antisense sensors were analyzed by flow cytometry 48h after transfection. GFP shows piRNA-dependent silencing with varying strength for different target regions.



Supplemental Figure S7. Design of Piwi-piRNA sensors. Multiple piRNAs are produced from long single stranded transcripts in varying register in flies and mice. Reporter constructs are designed with antisense complementarity to parts of piRNA-generation regions. Endogenous piRNAs with perfect complementarity to the sensors are depicted. Unique sequences are shown sorted by their 5'start site. The total number of fully complementary sequences is indicated. (**A**) The shortest sensor, Flam(as)-100, is only 100nt long, which corresponds to the length of ~4 Piwi-piRNAs. However, this 100nt-long element can be targeted by 397 unique piRNA sequences with perfect sequence complementarity in overlapping intervals. (**B**) There are 1056 fully complementary endogenous Piwi-piRNA sequences for the 230nt long Flam(as)-230b, and (**C**) 724 fully complementary piRNA sequences for the 230nt long Path(as)-230 sensor. The dense coverage of diverse piRNA sequences and unknwon rules for target engagement hamper the design of an unambiguous target-site for a single piRNA (one piRNA length indicated in yellow). Further studies are required to establish the rules of piRNA:target engagement and enable more precise sensor design and quantification of targeting piRNAs.