Supplemental Material

Planarians recruit piRNAs for mRNA turnover in adult stem cells

Iana V. Kim\textsuperscript{1}, Elizabeth M. Duncan\textsuperscript{2,4}, Eric J. Ross\textsuperscript{2,3}, Vladyslava Gorbovytska\textsuperscript{1}, Stephanie Nowotarski\textsuperscript{2}, Sarah A. Elliott\textsuperscript{2}, Alejandro Sánchez Alvarado\textsuperscript{2,3} and Claus-D. Kuhn\textsuperscript{1,*}

\textsuperscript{1} Gene regulation by Non-coding RNA, Elite Network of Bavaria and University of Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, Germany
\textsuperscript{2} Stowers Institute for Medical Research, 1000 East 50\textsuperscript{th} Street, Kansas City, MO 64110, USA
\textsuperscript{3} Howard Hughes Medical Institute, Stowers Institute for Medical Research, 1000 East 50\textsuperscript{th} Street, Kansas City MO, 64110, USA
\textsuperscript{4} Present address: Department of Biology, University of Kentucky, 202 TH Morgan Building, Lexington, KY 40506, USA

* Corresponding author: claus.kuhn@uni-bayreuth.de

List of supplemental material:

- Supplemental Methods (page 2 - page 11)
- Supplemental References (page 12 - page 15)
- Supplemental Figures S1 – S9
- Extended Supplemental Data
- Supplemental Tables 1-6
Supplemental Methods

Planarian culture

Asexual *S. mediterranea*, clone CIW4, were maintained at 20°C in Montjuic solution in the presence of 50 µg/ml gentamycin. Animals were fed weekly with homogenized calf liver and starved at least 7 days prior to experiments.

Expression of Strep-tagged SMEDWI proteins

Full-length SMEDWI-1, SMEDWI-2 and SMEDWI-3, each carrying a Strep-SUMO-Star tag, were expressed in High Five insect cells. Following 72 hours of protein expression, cells were pelleted, snap-frozen in liquid nitrogen and homogenized in RIPA buffer.

Western blotting

For protein sample preparation, 3 - 5 snap-frozen worms (4 - 5 mm) were homogenized in 50 µl homogenization buffer (30 mM HEPES, pH 7.7 (Sigma H989), 150 mM NaCl, 10 mM KCl, 4 mM MgCl₂, 1 mM DTT, 0.5% Triton X-100, Complete EDTA-free protease inhibitor (Roche)) using a micropestle. The lysate was then cleared by centrifugation at 16,000 rpm (21,130 g) for 15 min at 4°C in a tabletop centrifuge. The protein concentration of the obtained extracts was measured using the Bradford assay. Whole cell extracts were separated on 8% Bis-Tris gels before transfer onto nitrocellulose membranes. Not more than 15 - 20 µg of total protein was loaded per lane. Antibodies for Western blots were used with the following dilutions: rabbit polyclonal anti-SMEDWI-1 (gift from Dr. K. Bartscherer, Hubrecht Institute, Utrecht, Netherlands) 1:2,000 (Hauptmann et al. 2015), rabbit antiserum anti-SMEDWI-2 1:500, rabbit polyclonal anti-SMEDWI-3 1:4,000, mouse monoclonal anti-Strep-tag (iba, 2-1207-001) 1:1,000, mouse monoclonal anti-α-Tubulin (Sigma, T5168) 1:4,000, goat anti-rabbit IgG-peroxidase (Sigma, A0545) 1:20,000, goat anti-mouse IgG-peroxidase (Sigma, A9917) 1:20,000.

RNA interference

cDNA fragments for *smedwi*-2 (SMESG000069984.1) (nts 81 - 975) and *smedwi*-3 (SMESG000081970.1) (nts 1347 - 2332) were cloned into the pPR-T4P vector (gift of Dr. J. Rink,
Kim I.V., Duncan E.M., Ross E.J., Gorbovyska V., Nowotarski S.H., Elliott S.A., Sánchez A., Alvarado A., Kuhn C.D.

Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany). dsRNA synthesis and RNAi feeding experiments were carried out as described (Rouhana et al. 2013). dsRNA comprising nts 6616 - 7358 of unc-22 from C. elegans was used as control. For smedwi-2 RNAi experiments animals were fed twice (0 dpf (days post feeding), 3 dpf) and sacrificed on day 7 post feeding. For smedwi-3 RNAi experiments animals were fed three times (0 dpf, 3 dpf and 6 dpf) and sacrificed on day 11 post feeding. For SMEDWI-2 we found 7 dpf to be an optimal time point for a knockdown experiment, whereas 11 dpf was optimal for SMEDWI-3.

**Slide-section immunofluorescence**

Slide-section immunofluorescence was performed by rinsing animals 3 times in planarian water, incubating in 5% N-acetyl-cysteine in 1x PBS for 5 min with light hand-rolling of tubes and then fixing in 4% PFA in 0.5x PBS overnight at 4°C with no rocking. Then dehydrated into ethanol by 30%, 50%, 70% dilutions in water and samples were stored for a minimum overnight and maximum 7 days at 4°C prior to embedding. Samples were automatedly infiltrated with paraffin (Thermoscientific, Richard-Allan Type 9) using a Delta Pathos processor (Milestone Medical, Sorisole Italy). The automation included a 4 min rinse in 85% ethanol, 100% ethanol for 25 min at 65 °C, 100% isopropanol for 35 min at 68°C, 2 changes of paraffin at 70°C each, 8 min and then 6 min (Tissue Infiltration Medium: Surgipath through Leica). Following a final step of 20 min in 65 °C paraffin, the tissue was embedded in paraffin for transverse sectioning onto Superfrost Plus slides (Fisher Scientific) at 5 microns using a Leica RM2255 microtome (Leica Biosystems Inc. Buffalo Grove, IL). Slides were deparaffinized by incubating in 100% Xylene for 3 min (repeated twice), 100% ethanol for 3 minutes (repeated twice), 80% ethanol rinse and a water rinse. Slides were fixed with 4% PFA for 3 hours at room temperature (RT) and washed with PBS three times. Antigen retrieval was performed in citrate buffer heated to 95°C for 15 min, cooled down for 20 min and then washed in TBST for 30 min. Slides were blocked with 4 drops of Background Buster (Innovex Biosciences) and incubated for 30 min at RT. Primary antibody was applied and incubated over night at 4°C. The following dilutions were used: anti-SMEDWI-1 at 1:200, anti-SMEDWI-3 at 1:4000 and anti-Y12 at 1:500. Slides were washed with TBST three times and secondary Alexafluor antibodies (647, 488) were applied at 1:500 at RT.
for 1 hour. Slides were washed three times with TBST and mounted using Prolong Gold Mounting media (Thermofisher). Sections were imaged using a Zeiss 710 with a Nikon 20x Plan-Apo 0.8 NA objective.

**Whole-mount in situ hybridization**

Whole-mount *in situ* hybridization were performed as described (Duncan et al. 2015). Investigated transcripts have the following gene identifiers: *ank1* (SMESG000076223.1), *dapk1* (SMESG000043474.1), *traf6* (SMESG000000294.1), *histone H2B* (SMESG000052758.1) (Rozanski et al. 2019).

**Flow cytometry**

FACS sorting of X1 and Xins cells was performed as described (Duncan et al. 2015).

**RNA library preparation**

RNA libraries from 100,000 FACS-sorted planarian X1 and Xins cells were constructed as described (Zhang et al. 2012). Next generation sequencing was carried out on an Illumina Next-Seq 500 platform (paired-end 75 nt or single-end 75 nt mode).

**Processing of small RNA libraries**

Statistical and graphical analysis were performed with R/Bioconductor (Gentleman et al. 2004). 3′-adapters were trimmed using cutadapt (Martin 2011) and filtered against planarian rRNA sequences (Supplemental Table 6) with SortMeRNA (Kopylova et al. 2012). Only sequences longer than 18 nt and shorter than 40 nt were kept for the following analysis. Processed reads were mapped against the planarian genome (Grohme et al. 2018) by using bowtie (1.1.2) (Langmead et al. 2009) with the following settings “-v 0 -a --best --strata”. The resulting alignment files were converted into bed2 format with piPipes (Han et al. 2014). To account for multi-mapping reads, the total read count was divided by the number of locations that they map to. These weighted counts were used to characterize piRNAs mapping to genomic features with BEDTools (2.25.0) (Quinlan and Hall 2010). To compare the abundance of mRNA-derived
piRNAs associated with SMEDWI-1, SMEDWI-2 and SMEDWI-3 and to compute the small RNA coverage of target genes, piRNAs were normalized to RPM sequences, and only uniquely mapping piRNAs were considered.

**Calculation of the ping-pong signature**
Homotypic and heterotypic ping-pong signatures were calculated with piPipes (Han et al. 2014).

**Sequence logos and Venn diagrams**
To generate sequence logos with seqLogo (1.46.0) piRNA sequences mapped to the reference genome were collapsed to avoid biases due to the presence of highly abundant piRNAs. Only mapped piRNAs that appear in at least two replicates were used to generate Venn diagrams with the eulerr package (4.1.0).

**piRNA cluster prediction**
proTRAC version 2.4.2 (Rosenkranz and Zischler 2012) was used to assign 405,725 sequence reads to 270 piRNA clusters, covering 3.8 million base pairs. ProTRAC was run with “-clsize 1000, -pimax 40, pimin 26, -pdens 0.09”.

**Processing of the CHIP-Seq libraries**
H3K4me3 CHIPseq reads were aligned to the *Schmidtea mediterranea* genome with bowtie (version 1; parameters: -k 100 –strata) and formatted with samtools (version 1.8) (Grohme et al. 2018; Li et al. 2009; Langmead et al. 2009; Rozanski et al. 2019). Peaks were called with macs2 (version:2.1.1.20160309; parameters: -q 0.01) (Zhang et al. 2008). The R library DiffBind was used to identify peaks differentially bound between *smedwi-2(RNAi), smedwi-3(RNAi)* and *unc-22 (C. elegans) RNAi* as control. BEDTools (version: v2.26.0) was used to calculate the overlap of differential peaks with annotated repeat families (Quinlan and Hall 2010; Grohme et al. 2018; Rozanski et al. 2019).
Processing of RNA-seq libraries

Reads after removal of 3' adapters and quality filtering with Trimmomatic (0.36) were trimmed to a length of 50 nt. Next, sequences mapped to planarian rRNAs were removed with SortMeRNA. Reads were assigned to the reference genome or consensus transposable element sequences (Bao et al. 2015) in strand-specific mode and quantified with kallisto (Bray et al. 2016) with “--single -l 350 -s 30 -b 30” for single-end libraries and “-b 30” for paired-end libraries. Differential expression analysis of transposable elements upon RNAi knockdown was performed with edgeR (3.20.9) using the TMM normalization method (McCarthy et al. 2012; Robinson et al. 2010). Differentially expressed planarian transcripts were identified by sleuth (0.29.0) (Pimentel et al. 2017).

Processing of HITS-CLIP reads

Before mapping reads to the planarian genome, 3' adapters were removed with cutadapt, sequences mapping to planarian rRNA were discarded with SortMeRNA, and all duplicated reads were collapsed with the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit). Large and small CLIP reads from separately prepared upper and down band (Fig. 3B) libraries were pooled for their subsequent joint analysis. CLIP reads were mapped to the planarian genome using the bwa aligner with the following settings “bwa aln --n 0.06 --q 20” (Li and Durbin 2010). CLIP sites were determined employing CIMS analysis as implemented in the CTK tool kit (Zhang and Darnell 2011; Shah et al. 2017). Briefly, sites with aligned CLIP reads carrying substitutions and deletions (FDR ≤ 0.001) that appeared in at least in two libraries and with a total number of overlapping unique tags k at the aligning position ≥ 30 were considered as SMEDWI-3 targeted CLIP regions.

For the annotation of CLIP regions to genomic features reads were mapped to the planarian genome using STAR (Dobin et al. 2013) with the following settings: --outFilterMatchNmin 15 --outFilterMatchNminOverLread 0.72. Uniquely mapped CLIP libraries were annotated to transposable elements, piRNA clusters and coding genes with BEDTools.
Analysis of genic piRNA abundance across SMEDWI-3 HITS-CLIP target sites

To estimate the abundance of genic piRNAs across SMEDWI-3-targeted CLIP-sites, we calculated the density of the mapped SMEDWI-3 immunoprecipitated piRNAs per base pair of respective CLIP sites. Briefly, CLIP sites were defined as merged overlapping CLIP regions. Next, piRNAs were mapped to the CLIP sites with the following parameters “-v1 --a --best --strata”. Weighted counts for piRNAs mapping to CLIP regions in sense orientation were normalized to the length of the respective CLIP region (in bp).

Gene-set enrichment analysis (GAGE) of ping-pong and protected mRNAs

To analyze the coordinated differential expression of SMEDWI-3 ping-pong or -protected transcripts we carried out gene-set enrichment analysis with GAGE (2.28.2) (Luo et al. 2009). Expression changes of the two groups of genes upon RNAi knockdown in X1 and Xins cells were assessed with edgeR.

Degradome-Seq analysis

Degradome libraries were processed using piPipes (Han et al. 2014). Only genes that had at least 5 unique pairs of 5’ to 5’ 10 nt overlap between degradome reads and antisense piRNAs were considered as degraded in the ping-pong cycle.

GO term enrichment analysis

GO term enrichment analysis was conducted using topGO (2.30.1) with Fisher’s exact test (Alexa et al. 2006). GO-term gene annotations were obtained from PlanMine (Brandl et al. 2016).

Identification of chimeric reads

Chimeric reads were identified using the CLIP-chimaeric pipeline with default parameters (Alexiou et al. 2018; Vourekas et al. 2016). To calculate the base-pairing density for Group 1, Group 2 and Group 3 only chimeric reads with piRNAs part mapping within ±20 bp from the midpoint of mRNAs fragments were considered.
SHAPE-map data analysis

Sequenced reads from 1M7 and DMSO control libraries were analyzed using ShapeMapper2 (Busan and Weeks 2018). The implemented --correct.seq option was used to detect sequence variants and to correct target transcript sequences prior to SHAPE analysis. The software was executed with the following global parameters: " --min-depth = 1000, --indiv -norm". For histone H2B and MYH15 the default bowtie2 mapper was used, for smedwi-3 the star-aligner (star aligner). Calculated shape reactivities per nucleotide were used as constraints for RNA secondary structure prediction performed with RNAstructure 6.0.1 (Reuter and Mathews 2010). Base-pairing probabilities ($p_{i,j}$) were obtained from the .pfs output files from RNAstructure 6.0.1 and were used to calculate the Shannon entropy for each nucleotide (Mathews 2004).

$$S = - \sum_{i,j} p_{i,j} \times \log (p_{i,j})$$

Median SHAPE reactivities and median Shannon entropies were calculated over a centered 25-nt sliding window, allowing partial windows at the boundaries of the SHAPE data. Nucleotides with a median shape reactivity < 0.3 and a median Shannon entropy of < 0.04 (corresponding to a base-pairing probability of 0.95) were considered structured, otherwise they were defined as flexible (Mustoe et al. 2018). piRNAs co-immunoprecipitated with SMEDWI-3 were mapped to all mRNAs. For mapping a custom script was used that allowed up to 9 mismatches for piRNA mapping (for histone H2B only one mismatch was allowed) aside from a mandatory seed match for nucleotides 2-8. (We note here that respecting the experimentally determined piRNA targeting rules for C. elegans did not alter the observed distribution of predicted piRNA binding sites (Shen et al. 2018; Zhang et al. 2018)).

Primers used in this study

| Transcript ID    | Gene name | feature | Primer sequence                        |
|------------------|-----------|---------|---------------------------------------|
| SMESG000036375.1 | smedwi-1  | FWD     | ATGACCGGTGGCCAGCAGATGGGCATGGATT       |
| Accession        | Gene    | Type | Primers                                                                 |
|------------------|---------|------|-------------------------------------------------------------------------|
| SMESG000036375.1 | smedwi-1| REV  | CAACTAATGTTACAAGTG                                                     |
|                  |         |      | GTACTTTCGACAAGCTTCTACTACTGAGACA CCAAGTAAAATTCATAAG                   |
|                  |         | FWD  | ATTGAGGCTCACAGAGAACAGATTTGGTGGA ATGGAAGAAATCCCGGTGAAA                 |
|                  |         | REV  | GTACTTTCGACAAGCTTCTACTACAGATAAA ACAGGCGGTT                          |
|                  |         | FWD  | ATTGAGGCTCACAGAGAACAGATTTGGTGGA ATGGAAGAAATCCCGGTGAAA                 |
|                  |         | REV  | GTACTTTCGACAAGCTTCTACTACAGATAAA ACAGGCGGTT                          |
| Primers used for cloning RNAi constructs into pPR-T4P vector |
| * part of primer corresponding to vector overhang is in bold |
| SMESG000069984.1 | smedwi-2| FWD  | GGCTAGCGGCCATTACCATCCCGAGGTAAAAT GGGAAATCGTG                         |
|                  |         | REV  | TACCGGGCCGCGCAATTCTACCGGTCCCTGTG GATTTTGTAAACAGC                     |
|                  |         | FWD  | GGCTAGCGGCCATTACCATCCCGGAACACGTA TGGGGCTTACAA                       |
|                  |         | REV  | TACCGGGCCGCGCAATTCTACCGGCTTTCGTC TGCAATAGCGAT                      |
| Primers used for cloning RNAi constructs into pPR-T4P vector |
| * part of primer corresponding to vector overhang is in bold |
| SMESG000043474.1 | dapk1   | FWD  | GGCTAGCGGCCATTACCATCCCGGAACACGTA TGGGACATACAG                       |
|                  |         | REV  | TACCGGGCCGCGCAATTCTACCGGTGCACAT TGGGATTTCATTATC                     |
| SMESG000052758.1 | histone | FWD  | GGCTAGCGGCCATTACCATCCCGGATGGCTTCA                                |
| Primer Set                        | Gene    | Part | Sequence                        |
|----------------------------------|---------|------|---------------------------------|
| **Primers used for SHAPE-MaP reverse transcription** |         |      |                                 |
| SMESG000052758.1                 | histone |      | TACCGGGGCCGGCCAATTCTACCCGGGTAAC |
|                                  | H2B     | REV  | AGCTTTGGTTCC                    |
| SMESG000000294.1                 | traf6   | FWD  | GGCTAGCGGCGATTACCATCCCGGATGCTGAA |
|                                  |         |      | AGCGTTTGG                       |
| SMESG000000294.1                 | traf6   | REV  | TACCGGGGCCGGCCAATTCTACCCGGGAAAC |
|                                  |         |      | GAACCGAAATTG                     |
| SMESG000076223.1                 | ank1    | FWD  | GGCTAGCGGCGATTACCATCCCGGCTGATCGG |
|                                  |         |      | ATCATTCAA                        |
| SMESG000076223.1                 | ank1    | REV  | TACCGGGGCCGGCCAATTCTACCCGGTCAACA |
|                                  |         |      | GGGTTTCTCC                       |

**Primers used for SHAPE-MaP amplicon PCR**

| Primer Set                        | Gene    | Part   | Sequence                        |
|-----------------------------------|---------|--------|---------------------------------|
| SMESG000081970.1                  | smedwi-3| Part1  | TAGTCATATGTCATACACACTATTC       |
| SMESG000081970.1                  | smedwi-3| Part2  | AATACCCACTTGACTTGGAACAGGGC      |
| SMESG000081970.1                  | smedwi-3| Part3  | TCTGTCACATAAGGTACTAGGGG         |
| SMESG000052758.1                  | histone | Part1  | TTGAACCAGTATATTTGGAACAGGC       |
| SMESG000056451.1                  | MYH15   | Part1  | AATACCTTCCCAATACACGGTTC         |
| SMESG000056451.1                  | MYH15   | Part2  | GTTCTCGACCAAGATTGTTCC          |
| SMESG000056451.1                  | MYH15   | Part3  | GAACATCGGCTTTGCGCTTCTTG        |
| SMESG000056451.1                  | MYH15   | Part4  | CGGCTGTATACACGCAAGCTGTC        |
| SMESG000056451.1                  | MYH15   | Part5  | TTAGTCTACAGGCTGGATCTG          |

| Primer Set                        | Gene    | Part   | Sequence                        |
|-----------------------------------|---------|--------|---------------------------------|
| SMESG000081970.1                  | smedwi-3| Part1  | TAGTCATATGTCATACACACTATTC       |
| SMESG000081970.1                  | smedwi-3| Part1  | ACACATCGGACGGTAAACAGGTTC        |
| SMESG000081970.1                  | smedwi-3| Part2  | CAAACCGACTGATAAGACTGTC          |
| Accession | Gene | Part   | Orientation | Sequence               |
|-----------|------|--------|-------------|------------------------|
| SMESG000081970.1 | smedwi-3 | Part2 | REV         | TGTACTACTAACTGCACACCAG |
| SMESG000081970.1 | smedwi-3 | Part3 | FWD         | GCCCTGGCACATAATAAGTTATAG |
| SMESG000081970.1 | smedwi-3 | Part3 | REV         | ATGAGCATAACATACAGGTGCTG |
| SMESG000052758.1 | histone H2B | FWD | REV         | ATGGCTCCAAAAGCAAAGTTGC |
| SMESG000052758.1 | histone H2B | REV | FWD         | TTGAACCAGTATATTTGGTAAACAGC |
| SMESG000056451.1 | MYH15 | Part1 | FWD         | CAGAAAAAGGAGAAAGAAATTACC |
| SMESG000056451.1 | MYH15 | Part1 | REV         | TGTGTGATTCACTTGGGACAATG |
| SMESG000056451.1 | MYH15 | Part2 | FWD         | AATAAGGATCCATTGAACGAAAGTG |
| SMESG000056451.1 | MYH15 | Part2 | REV         | AGCTTTATTGAGGTTACTGACTTTC |
| SMESG000056451.1 | MYH15 | Part3 | FWD         | GAAACTCAATGCTGACATCGATG |
| SMESG000056451.1 | MYH15 | Part3 | REV         | ACTCATCGACATTTCTAAATCTCC |
| SMESG000056451.1 | MYH15 | Part4 | FWD         | CAGATCTTTCAAAACAAACTCGATG |
| SMESG000056451.1 | MYH15 | Part4 | REV         | AGCTTGCAATTCTGCTAATTGTTG |
| SMESG000056451.1 | MYH15 | Part5 | FWD         | CCGTCAAGATCTGAAAGAAGAC |
| SMESG000056451.1 | MYH15 | Part5 | REV         | CAGTTATAGAAATTTGGTTCAGG |
Supplemental References

Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22: 1600–1607.

Alexiou P, Maragkakis M, Mourelatos Z, Vourekas A. 2018. cCLIP-Seq: Retrieval of Chimeric Reads from HITS-CLIP (CLIP-Seq) Libraries. Methods Mol Biol 1680: 87–100.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6: 11.

Brandl H, Moon H, Vila-Farré M, Liu S-Y, Henry I, Rink JC. 2016. PlanMine—a mineable resource of planarian biology and biodiversity. Nucleic Acids Research 44: D764–73.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34: 525–527.

Busan S, Weeks KM. 2018. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. RNA 24: 143–148.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21.

Duncan EM, Chitsazan AD, Seidel CW, Sánchez Alvarado A. 2015. Set1 and MLL1/2 Target Distinct Sets of Functionally Different Genomic Loci In Vivo. CellReports 13: 2741–2755.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.

Grohme MA, Schloissnig S, Rozanski A, Pippel M, Young GR, Winkler S, Brandl H, Henry I, Dahl A, Powell S, et al. 2018. The genome of Schmidtea mediterranea and the evolution of core cellular mechanisms. Nature Publishing Group 554: 56–61.
Han BW, Wang W, Zamore PD, Weng Z. 2014. piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. Bioinformatics 31: 593–595.

Hauptmann J, Schraivogel D, Bruckmann A, Manickavel S, Jakob L, Eichner N, Pfaff J, Urban M, Sprunck S, Hafner M, et al. 2015. Biochemical isolation of Argonaute protein complexes by Ago-APP. Proc Natl Acad Sci USA 112: 11841–11845.

Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28: 3211–3217.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589–595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. 2009. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 10: 161.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal 17: 10–12.

Mathews DH. 2004. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. RNA 10: 1178–1190.

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Research 40: 4288–4297.
Kim I.V., Duncan E.M., Ross E.J., Gorbovyskva V., Nowotarski S.H., Elliott S.A., Sánchez 14 Alvarado A., Kuhn C.D.

Mustoe AM, Busan S, Rice GM, Hajdin CE, Peterson BK, Ruda VM, Kubica N, Nutiu R, Baryza JL, Weeks KM. 2018. Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell* **173**: 181–195.

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Meth* **14**: 687–690.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**: 129.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Rosenkranz D, Zischler H. 2012. proTRAC—a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics* **13**: 5.

Rouhana L, Weiss JA, Forsthöfel DJ, Lee H, King RS, Inoue T, Shibata N, Agata K, Newmark PA. 2013. RNA interference by feeding in vitro-synthesized double-stranded RNA to planarians: methodology and dynamics. *Dev Dyn* **242**: 718–730.

Rozanski A, Moon H, Brandl H, Martín-Durán JM, Grohme MA, Hüttner K, Bartscherer K, Henry I, Rink JC. 2019. PlanMine 3.0—improvements to a mineable resource of flatworm biology and biodiversity. *Nucleic Acids Research* **47**: D812–D820.

Shah A, Qian Y, Weyn-Vanhentenryck SM, Zhang C. 2017. CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. *Bioinformatics* **33**: 566–567.

Shen E-Z, Chen H, Ozturk AR, Tu S, Shirayama M, Tang W, Ding Y-H, Dai S-Y, Weng Z, Mello CC. 2018. Identification of piRNA Binding Sites Reveals the Argonaute Regulatory Landscape of the C. elegans Germline. *Cell* **172**: 937–951.
Kim I.V., Duncan E.M., Ross E.J., Gorbovytska V., Nowotarski S.H., Elliott S.A., Sánchez 15 Alvarado A., Kuhn C.D.

Vourekas A, Alexiou P, Vrettos N, Maragkakis M, Mourelatos Z. 2016. Sequence-dependent but not sequence-specific piRNA adhesion traps mRNAs to the germ plasm. *Nature Publishing Group* **531**: 390–394.

Zhang C, Darnell RB. 2011. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* **29**: 607–614.

Zhang D, Tu S, Stubna M, Wu W-S, Huang W-C, Weng Z, Lee H-C. 2018. The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. *Science* **359**: 587–592.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Zhang Z, Theurkauf WE, Weng Z, Zamore PD. 2012. Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence* **3**: 9.
Figure S1

A

| Antibody       | Insect cell lysate | Planarian lysate |
|----------------|--------------------|------------------|
| Anti-Strep     | 1:2000             | 1:2000           |
| Anti-SMEDWI-1  | 1:2000             | 1:500            |
| Anti-SMEDWI-2  | 1:2000             | 1:20,000         |
| Anti-SMEDWI-3  | 1:4000             | 1:4000           |

Antibodies:
- Anti-Strep
- Anti-SMEDWI-1
- Anti-SMEDWI-2
- Anti-SMEDWI-3

Insect cell lysate:
1 2 3
Insect cell lysate:
- - -

Planarian lysate:
1 2 3
Planarian lysate:
- - -

Strep-SMEDWI-1
Strep-SMEDWI-2
Strep-SMEDWI-3

α-tubulin

100 µm

50 µm

B

SMEDWI-2  SMEDWI-1  DAPI  Merge  SMEDWI-3  Y12  DAPI  Merge

10 µm

C

SMEDWI-2  SMEDWI-1  DAPI  Merge  SMEDWI-2  SMEDWI-1  DAPI  Merge  SMEDWI-3  SMEDWI-1  DAPI  Merge

100 µm

100 µm

100 µm

50 µm
Figure S1. Characterization of planarian PIWI proteins. (A) Western blot analysis of SMEDWI-1, -2 and -3 antibodies confirms their specificity. Test samples were prepared from insect cell lysates expressing Strep-SUMO-tagged SMEDWI proteins or directly from planarian lysate. α-Tubulin was used as a loading control. (B) Subcellular localization analysis of SMEDWI-1, -2, and -3 using immunofluorescence. SMEDWI-3 signal (in red) is enriched in the cytoplasm and overlaps with chromatoid bodies stained by anti-Y12 antibodies (in green). SMEDWI-1 (in green) is cytoplasmic, while SMEDWI-2 (in red) shows a nuclear localization. Nuclei stained with DAPI are in blue. (C) Co-immunostaining of SMEDWI-1 and SMEDWI-2 and of SMEDWI-1 and SMEDWI-3 on cross-sections through the planarian brain and the pharynx. SMEDWI-1 is in green, SMEDWI-2 and SMEDWI-3 are in red. Nuclear DAPI staining is shown in blue. SMEDWI-2 is present in both neoblasts and differentiated cells, while the signal for SMEDWI-1 and SMEDWI-3 is predominantly cytoplasmic and overall restricted to neoblasts. Both SMEDWI-2 and SMEDWI-3 are present in cephalic ganglia on the cross-sections through the planarian brain.
Figure S2. Immunoprecipitation of planarian PIWI proteins and detection of planarian piRNA clusters. (A) Radioactive labeling of total RNA from planarians isolated from whole animals. piRNAs are predominantly 30-35 nt long, while miRNAs are less abundant and have a length of ~22 nt. (B) Western blot analysis of SMEDWI-2 and SMEDWI-3 in the presence of HEPES hemisodium or NaOH-buffered HEPES acid confirms the sensitivity of SMEDWI proteins to ionic strength. α-Tubulin was used as a loading control. (C) Percentage of unique sequences bound by different SMEDWI proteins. Only genome-mapped piRNAs that appeared in at least two replicates were counted. (D) Coverage profile of H3K4me3 CHIP-seq, RNA-Seq and total piRNAs on unidirectional piRNA cluster 12. (E) Evaluation of SMEDWI-1, -2 and -3 protein levels by Western blot upon unc-22(RNAi), smedwi-2(RNAi) or smedwi-3(RNAi) at different days post feeding (dpf) with dsRNA. α-Tubulin was used as a loading control. (F) Gene expression levels in TPMs (transcript per million) of smedwi-1, -2 and -3 transcripts upon unc-22(RNAi) and smedwi-2(RNAi) or smedwi-3(RNAi) in neoblasts (X1) and differentiated (Xins) cells. (G) piRNAs co-immunoprecipitated from whole-worm lysates of control (unc-22 RNAi) or knockdown (smedwi-1,-2,-3 RNAi) animals using antibodies against SMEDWI-1. Below, western blot showing SMEDWI-1 protein levels in different RNAi knockdown backgrounds.
Kim I.V., Duncan E.M., Ross E.J., Gorbovytska V., Nowotarski S.H., Elliott S.A., Sánchez Alvarado A., Kuhn C.-D.

**Figure S3.** Immunostaining of SMEDWI-1, -2 and -3 on cross-sections through the planarian pharynx in control (**unc-22 RNAi**) or knockdown (**smedwi-1,-2,-3 RNAi**) animals. SMEDWI-1 is in green, SMEDWI-2 is in red, SMEDWI-3 is in yellow. Nuclear staining with DAPI is shown in blue.
Figure S4

A) Total planarian RNA

B) Depleted planarian RNA

C) hAT−12_SM_hAT

D) smedwi−2 (RNAi)  smedwi−3 (RNAi)

Number of H3K4me3 peaks in TE regions changed over control.
Kim I.V., Duncan E.M., Ross E.J., Gorbovyska V., Nowotarski S.H., Elliott S.A., Sánchez Alvarado A., Kuhn C.-D.

**Figure S4. Differential expression of planarian transposons upon smedwi-2 and smedwi-3 knockdown (A)** Bioanalyzer plot of planarian total RNA and planarian RNA after rRNA depletion. Due to a cleavage event planarian 28S and 18S rRNAs are detected as a single peak on electropherograms. **(B)** Efficiency of the custom rRNA depletion procedure. **(C)** Heatmap showing transcript abundance (log2CPM) and differential expression (log2FC) for transposable element families upon smedwi-2 or smedwi-3 knockdown. **(D)** Number of differentially expressed H3K4me3 peaks overlapping with transposable elements upon smedwi-2(RNAi) or smedwi-3(RNAi) in neoblasts.
Figure S5

A) Immunoprecipitated piRNAs mapped to 5'–UTRs of coding genes

B) Immunoprecipitated piRNAs mapped to 3'–UTRs of coding genes

C) SMEDWI-3 bound piRNAs

D) RNA-Seq analysis

- Genic piRNAs
- Sense to TE
- Antisense to TE

RPM (reads per million)
Kim I.V., Duncan E.M., Ross E.J., Gorbovytska V., Nowotarski S.H., Elliott S.A., Sánchez Alvarado A., Kuhn C.-D.

**Figure S5. SMEDWI-3 binds genic piRNAs.**  
(A) piRNAs co-immunoprecipitated with SMEDWI proteins are mapped to 5'-UTRs of coding genes.  
(B) piRNAs co-immunoprecipitated with SMEDWI proteins are mapped to 3'-UTRs of coding genes.  
(C) Percentages of SMEDWI-3-bound genic piRNAs mapping in sense and antisense orientation to transposable elements.  
(D) Coverage profile of RNA-seq and immunoprecipitated piRNAs for a *npk1-like* gene (SMESG000017261.1).
Figure S6

A

B

C

D

| Sample               | Sequenced reads | Collapsed | rRNA    | Others     |
|----------------------|-----------------|-----------|---------|------------|
| CLASH rep1 Up        | 24,491,573      | 4,852,114 | 23,402  | 4,828,712  |
|                      |                 |           | (0.48%) | (99.52%)   |
| CLASH rep1 Down      | 26,355,748      | 7,165,640 | 10,563  | 7,155,077  |
|                      |                 |           | (0.15%) | (99.85%)   |
| CLASH rep2 Up        | 27,365,590      | 6,868,529 | 26,767  | 6,841,762  |
|                      |                 |           | (0.39%) | (99.61%)   |
| CLASH rep2 Down      | 25,765,796      | 7,680,635 | 10,827  | 7,669,808  |
|                      |                 |           | (0.14%) | (99.86%)   |
| CLASH rep3 Up        | 21,465,736      | 4,734,867 | 26,630  | 4,708,237  |
|                      |                 |           | (0.56%) | (99.44%)   |
| CLASH rep3 Down      | 24,209,886      | 7,526,984 | 11,138  | 7,515,846  |
|                      |                 |           | (0.15%) | (99.85%)   |
| CLASH rep4 Up        | 20,977,519      | 5,260,232 | 20,596  | 5,239,636  |
|                      |                 |           | (0.39%) | (99.61%)   |
| CLASH rep4 Down      | 25,279,223      | 9,016,828 | 12,621  | 9,004,207  |
|                      |                 |           | (0.14%) | (99.86%)   |
| CLASH rep5 Up        | 19,402,092      | 1,125,858 | 154,604 | 971,254    |
|                      |                 |           | (13.73%)| (86.27%)   |
| CLASH rep5 Down      | 28,465,247      | 7,418,47 | 6,663   | 735,184    |
|                      |                 |           | (0.90%) | (99.10%)   |
| CLASH rep6 Up        | 25,713,339      | 1,790,701 | 133,178 | 1,657,523  |
|                      |                 |           | (7.44%) | (92.56%)   |
| CLASH rep6 Down      | 28,902,373      | 9,407,61  | 12,647  | 928,114    |
|                      |                 |           | (1.34%) | (98.66%)   |
| HITS-CLIP rep1 Up    | 22,658,032      | 5,740,698 | 33,265  | 5,707,433  |
|                      |                 |           | (0.58%) | (99.42%)   |
| HITS-CLIP rep1 Down  | 25,275,256      | 8,037,363 | 12,600  | 8,024,763  |
|                      |                 |           | (0.16%) | (99.84%)   |
| HITS-CLIP rep2 Up    | 27,914,812      | 1,992,586 | 197,431 | 1,795,155  |
|                      |                 |           | (9.91%) | (90.09%)   |
| HITS-CLIP rep2 Down  | 29,016,966      | 1,554,661 | 14,472  | 1,540,189  |
|                      |                 |           | (0.93%) | (99.07%)   |
| HITS-CLIP rep3 Up    | 25,854,504      | 1,203,988 | 247,219 | 956,769    |
|                      |                 |           | (20.53%)| (79.47%)   |
| HITS-CLIP rep3 Down  | 26,441,640      | 1,426,598 | 5201    | 1,421,397  |
|                      |                 |           | (0.36%) | (99.64%)   |
Kim I.V., Duncan E.M., Ross E.J., Gorbovytska V., Nowotarski S.H., Elliott S.A., Sánchez Alvarado A., Kuhn C.-D.

**Figure S6. SMEDWI-3 CLIP library preparation and sequencing.** (A) Fragment Analyzer digital gel image of RNA fragments crosslinked and co-immunoprecipitated along with SMEDWI-3. Pre-immune serum was used as a negative control. The gel electrophoresis profile for total planarian RNA demonstrates the high level of RNase activity in planarian lysates. (B) Fragment Analyzer electropherogram showing the separation profile of crosslinked RNA fragments co-immunoprecipitated with the anti-SMEDWI-3 antibody or pre-immune serum. (C) SMEDWI-3 HITS-CLIP immunoprecipitated complexes. RNAs crosslinked to SMEDWI-3 were radiolabeled, separated on an 8% Bis-Tris gel and blotted onto a nitrocellulose membrane. (D) Sequenced HITS-CLIP and CLASH reads.
### Figure S7

#### A

| Sample         | Total reads (Up+Down) | Mapped to the genome | Unique mappers | Multimappers | Chimeric reads |
|----------------|-----------------------|----------------------|----------------|--------------|----------------|
| CLASH rep1     | 11 983 789            | 7 467 622 (62.3%)    | 5 123 574      | 2 344 048    | 3 836 (0.03%)  |
| CLASH rep2     | 14 511 570            | 9 685 201 (66.7%)    | 6 681 087      | 3 004 114    | 3 851 (0.026%) |
| CLASH rep3     | 12 224 083            | 7 785 957 (63.7%)    | 5 360 248      | 2 425 709    | 5 553 (0.045%) |
| CLASH rep4     | 14 243 843            | 9 336 396 (55.5%)    | 6 412 284      | 2 924 112    | 3 885 (0.027%) |
| CLASH rep5     | 1 706 438             | 632 053 (37%)        | 424 656        | 207 397      | 33 418 (1.95%) |
| CLASH rep6     | 2 585 637             | 1 110 606 (65.2%)    | 750 874        | 359 732      | 26 125 (1.01%) |
| HITS-CLIP rep1 | 13 732 196            | 9 058 729 (65.9%)    | 6 226 614      | 2 832 115    | 290 (0.002%)   |
| HITS-CLIP rep2 | 3 335 344             | 2 202 502 (66%)      | 1 539 741      | 662 761      | 9 979 (0.3%)   |
| HITS-CLIP rep3 | 2 378 166             | 1 324 806 (55.7%)    | 907 520        | 417 286      | 13 646 (0.57%) |

#### B

Mutations in CLIP tags (in % of total)

- No mutation
- Deletion
- Insertion
- Substitution

#### C

Substitutions in CLIP tags (in % of total)

#### D

Deletions, Substitutions, Insertions (in % of total mutations)

#### E

Length (nt)

Number of unique piRNA sequences

#### F

Total number of chimeric reads

#### G

CLIP targets with CPM in the quantile

- CDS
- Cluster repDNA
- LINE
- LTR
- Unknown repeat
Figure S7. CIMS analysis and identification of chimeric reads. (A) HITS-CLIP and CLASH reads mapped to the planarian genome. (B) Percentage of mutations in CLIP tags. (C) Nucleotide substitution frequency observed in CLIP reads. (D) Distribution of deletions (in orange), substitutions (in grey) and insertions (in red) along CLIP reads. (E) Average length of the piRNA part of chimeric reads. (F) Cumulative sum of chimeric reads in nine replicates of CLIP targets. All CLIP targets were divided into bins based on their expression value (in CPMs) in the four quantiles (G) Percentage of chimeric reads mapped to different genomic features in sense and antisense orientation.
Figure S8. Analysis of SMEDWI-3-targeted transcripts (A) RNA-Seq, piRNA-seq and HITS-CLIP coverage profiles for two representative example genes, h2b (SMESG000067931.1) and h2afx (SMESG000067932.1). (B) The same as in (A) for smedwi-3 (SMESG000081970.1). (C) Genomic annotation of Degradome-Seq reads showing 5’ to 5’ overlap with piRNAs immunoprecipitated with anti-SMEDWI-1, -2, and -3 antibodies. (D) Scatter plot of normalized CPMs of Degradome-Seq and RNA-Seq data. Each black dot represents a transcript with ping-pong degradation signature. In red are SMEDWI-3 CLIP targets. Spearman’s correlation coefficient is indicated.
Kim I.V., Duncan E.M., Ross E.J., Gorbovytska V., Nowotarski S.H., Elliott S.A., Sánchez Alvarado A., Kuhn C.-D.

**Figure S9. Characterization of SMEDWI-3 ping-pong targets in the planarian epidermis. (A)** Boxplots showing the log2 fold change of up- and downregulated transcripts upon *smedwi-2*(RNAi) in neoblasts (X1) and differentiated cells (Xins). Statistical significance of differential expression of the gene sets was assessed using Generally Applicable Gene-set Enrichment (GAGE) analysis with a two-sample t-test (“ns” = not significant). **(B)** Histone H2B (SMESG000052758.1) mRNA (in green) localizes to chromatoid bodies stained with anti-Y12 (in magenta). Nuclei stained with Hoechst are in blue. **(C)** Nuclear localization of foci was calculated as a distance between each in situ signal (in orange) and the nearest nuclear center (x-axis = 0). The nuclear membrane data is plotted in blue. The analysis was performed on WISH signal for *traf6*. **(D)** Immunostaining of SMEDWI-1 (in green) and SMEDWI-2 (in red) in the epidermis on cross-sections through the pharynx of *S. mediterranea* upon differential RNAi knockdown. Nuclei are stained with DAPI and are shown in blue.
Kim I.V., Duncan E.M., Ross E.J., Gorbovyska V., Nowotarski S.H., Elliott S.A., Sánchez Alvarado A., Kuhn C.-D.

EXTENDED SUPPLEMENTAL DATA FIGURE LEGEND

Ex vivo SHAPE-MaP results for *smedwi-3* (A) and *MYH15* (B) mRNA

**Upper plot including exemplary structure cut-outs:** Median SHAPE reactivity (gray) and piRNA coverage prediction (orange) for regions of *smedwi-3* (A, SMESG000081970.1) and *MYH15* (B, SMESG000056451.1) mRNA. For *smedwi-3* mRNA nucleotides 1,000-2,000 are shown (A), for *MYH15* mRNA nucleotides 2,000-3,000 (B). Exemplary regions from the SHAPE reactivity-constrained secondary structure models of *smedwi-3* and *MYH15* mRNAs encompassing major predicted piRNA binding sites are shown above the plots. PiRNAs (green) are drawn along the structures.

**Lower plot:** Median SHAPE reactivity (gray) and median Shannon entropy (red) for a cut-out of *smedwi-3* mRNA and *MYH15* mRNA. Nucleotides below a threshold (in gray) defined by Shannon entropy < 0.04 and SHAPE reactivity < 0.3 were defined as structured (see exemplary secondary structure models above). The predicted piRNA binding sites predominantly fall into flexible regions. 3’ ends of predicted piRNAs at times extend into structured regions.
