Supplemental Information

PIWI Slicing and EXD1 Drive Biogenesis of Nuclear piRNAs from Cytosolic Targets of the Mouse piRNA Pathway

Zhaolin Yang, Kuan-Ming Chen, Radha Raman Pandey, David Homolka, Michael Reuter, Bruno Kotska Rodino Janeiro, Ravi Sachidanandam, Marie-Odile Fauvarque, Andrew A. McCarthy, and Ramesh S. Pillai
UCSC Genome Browser view of Gl(ROSA)26 locus Chr 6 (mm10 genome assembly).

The gene is on the bottom strand.

chr6:113067429-113077244 (9.8 kb)

Wildtype Rosa26 allele

Targeting cassette

Targeted Rosa26 allele

Putative transcripts from targeted allele

Figure- S1
A. RNA substrate

1. RNA alone
2. RNA + buffer
3. RNA + buffer, protein
4. RNA + buffer, protein, Mg²⁺
5. RNA + buffer, protein, Mg²⁺, RNase inhibitor
6. RNA + buffer, protein, Mg²⁺, heat-inactivation

B. Limited proteolysis of Bombyx Exd1

- Tdrl2
- Dimersization
- FL

- Band A (structure solved)
- Band B

Time of incubation with Trypsin

C. Coomassie

D. 'Open' conformation (Native crystal)

'Open' conformation (Gd²⁺-incorporated crystal)

Gd²⁺

90° > 30Å

E. 'Closed' conformation

Q273 (Hinge region)

F. 'Open' conformation (Native crystal)

'Dimerization'

E266

H263

H262

D113

D138

G. 'Closed' conformation (Gd²⁺-incorporated crystal)

'Tdrl2'

Dimerization

FL (1-315 aa)

Dimersization

A (1-228 aa) -
B (1-274 aa) +
C (32-315 aa) +
D (51-315 aa) +
E (79-315 aa) +

Input

HA-BmExd1 (FL)

HA-IP

Input

HA-BmExd1 (FL)

Input

Myc

Myc-BmExd1

Figure- S3
**Figure S4**

A. Western blot analysis of Lsm ring of snRNP.

B. Immunoprecipitation of BmExd1 with modelled ssRNA and yeast Rrp6 with bound ssRNA.

C. Diagram showing the lam ring of snRNP.

D. Schematic representation of Tdrd12 interaction and dimerization.

E. Nickel beads pull-down assay showing input and HA-IP (Exd1) for BmExd1 and BmTdrd12.

F. Diagram showing Myc-tagged proteins and HA-mTdrd12.

G. Localization of BmExd1 and BmTdrd12 in BmN4 cells.

H. Western blot analysis of BmExd1 and BmTdrd12 in BmN4 cells.

I. Diagram showing Coomassie blue staining of HA-mTdrd12 and Myc-BmExd1.

J. Diagram showing HA-BmTdrd12 (FL) and localization in BmN4 cells.

K. Diagram showing BmExd1 with modelled ssRNA.
Exd1 gene

Targeting construct
Exons 8-10 deleted
and replaced by
STOP-REN-PGK-Keo-polyA-REN
immediately after 1st codon
(ATG) of Exon 8

Targeted allele

Exd1 WT allele

Genomic DNA from targeted A9 ES cells
digested with HincII

Clone 7A used for generation of mice

Genomic PCR with
mouse tail DNA

MicroRNA

Total small RNAs

Fraction of piRNA pairs

Exd1 wild-type

Exd1 knockout

PIWI2

MILI

E18.5 Exd1 (-/-)

P0 testes
SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Generation of Rosa26-pi reporter mouse, Related to Figure 1

(A) Targeting strategy for generating a DsRed reporter with binding sites for endogenous Mili piRNAs. The 3′-UTR is based on non-coding LacZ and has perfectly complementary binding sites for 35 independent Mili-bound piRNAs. These targeting piRNAs are those expressed in embryonic or perinatal male germline. The reporter is knocked-into the ubiquitous Rosa26 locus. The binding sites for the A/I restriction enzyme (used for Southern probing) on the endogenous and targeted locus are shown. (B) Southern blotting to detect the targeted Rosa26 allele using ES cell genomic DNA. DNA ladders are 1kb Extension Ladder (Invitrogen) or 1kb ladder (NEB). (C) Design of the mouse reporter transcript with binding sites for endogenous Mili piRNAs (shaded areas). 5′ ends of the targeting Mili piRNAs complementary to the transcript were mapped and those exceeding the abundance of 1 rpm (reads per million) were plotted. (D) Mapping of the 5′ and 3′ ends of reporter-derived artificial piRNAs relative to the 5′ end (nucleotide position -1) of the targeting Mili piRNA. The secondary piRNAs with their 5′ end created by Mili slicing (position -10) are most abundant. Additional piRNAs (inchworm piRNAs) are produced downstream (small peaks at +18 to +33). (E) Mili and Miwi2 piRNAs were divided into secondary piRNAs (with 5′ end created by Mili slicing ~ at position -10) and inchworm piRNAs (produced downstream with 5′ ends at position +18 to +33). The graph shows the nucleotide composition of first 10 and last 10 nucleotides of the piRNAs. Notice that Mili- and Miwi2-bound secondary piRNAs do not display any strong bias for a particular nucleotide, as influenced by the reporter design and Mili slicing at a fixed position. However, inchworm piRNAs start preferentially with U indicating the nucleotide preference of the unknown nuclease generating it. The graph also shows the composition of the single nucleotide on the reporter sequence (DNA nt) after the 3′ end of the piRNAs. This reveals that the secondary piRNAs are not followed by a U in the reporter (DNA nt), suggesting that the 3′ ends of piRNAs are matured by 3′-5′ trimming after the transcript cleavage downstream that generates the 5′ end (U-biased) of the inchworm piRNA. The strong A10 of secondary piRNAs is influenced by the reporter design, while the high A10 of inchworm piRNAs was unexpected. (F) Endogenous targeting piRNAs and reporter-derived piRNAs produced from the DsRed reporter expressed in the Rosa26-pi mouse. This plot is intended to show that the piRNAs produced from the upstream DsRed region are lower in proportion to that generated from the regions designed to targeted by Mili slicing.

Figure S2. Characterization of Exd1 antibodies, Related to Figure 2

(A) Demonstration of specificity of mExd1 antibodies. Domain architecture of mouse Exd1 (mExd1) and the antigens (N-term or C-term) used for raising antibodies are indicated. The deletion versions of tagged mExd1 used in this study are also shown. (Left) Constructs expressing the Myc-tagged fusions were transfected into HEK293T mammalian cell cultures and examined by Western analysis with indicated anti-mExd1 antibodies. Endogenous mExd1 immunoprecipitated (IP) with the antibodies from mouse testes is shown alongside. The Myc-tagged mExd1 (in transfected HEK293T cells) migrates slower than the endogenous mExd1 protein. (Right) Note that the antibodies raised with the C-term antigen do not recognize the Myc-mExd1-C deletion construct. (B) Presence of mExd1 in mTdrd12 complexes isolated from mouse testes. This is a replicate of the experiment shown in Figure 1A. (C) Demonstration of specificity of BmExd1 antibodies. Lysates from Bombyx BmN4 cells expressing indicated tagged proteins were tested with both anti-HA and anti-BmExd1 antibodies. Note the detection of a slow-migrating band (arrow) of HA-BmExd1 present in transfected BmN4 cell lysates in addition to the endogenous BmExd1 protein. The anti-BmExd1 antibodies did not cross-react with HA-DmExd1. (D) Co-immunoprecipitation of Myc-BmExd1 with tagged BmTdrd12 and not with PIWI proteins expressed in BmN4 cells.
Figure S3. Biochemical and Structural Characterization of BmExd1, Related to Figure 3

(A) BmExd1 or its deletion versions lacks ribonuclease activity against single-stranded RNA (5'-end labelled). Reactions were carried out in the indicated buffer and supplemented with divalent metal ions when required. (B) Limited proteolysis of BmExd1 with trypsin yielded several well-defined protease-resistant species, two of which (Band A and B) were identified by mass spectrometry. Boundaries identified in band A (73-315 aa) led to successful diffraction-quality crystals. (C) Quality of proteins used for crystallization and biophysical experiments. Fractions from a gel-filtration column are shown. (D) Ribbon representation of BmExd1 dimer (red and blue monomers) model obtained with native crystals. This is in the ‘open’ conformation with a large groove in between the two monomers. (E) Ribbon representation of BmExd1 dimer (blue and white monomers) model obtained from crystals containing the heavy metal Gadolinium (Gd³⁺, pink dots; two out of the three in total are visible in the top view of the model). Note the striking change in orientation (a 55° rotation corresponding to 30 Å movement) of the second (white) monomer after superimposition of one (blue monomer) endonuclease domain. This Gd³⁺ BmExd1 dimer is in a ‘closed’ conformation, with a significant narrowing of the groove between the monomers. (F) Details of residues coordinating the Gd³⁺ ion. Note that an additional interaction of the Gd³⁺ ion with D113 (yellow) from the second monomer allows the conformation change. (G) Co-immunoprecipitation of HA-BmExd1 with various Myc-tagged BmExd1 deletion versions. Note the absence of association when the C-terminal region is deleted (constructs A and B). This is a replicate of the experiment shown in Figure 3G.

Figure S4. Biochemical interactions of BmExd1 with RNA and BmTdrd12, Related to Figure 4

(A) Electrophoretic mobility shift assay (EMSA) with native gels to demonstrate RNA-binding capacity of full-length BmExd1 to ssRNAs (numbers indicate different sequences used). Bovine serum albumin (BSA) is used as a negative control. Ctrl., RNA without protein. (B) Modelled ssRNA (green) placed in the catalytic pocket of one of the monomers in the BmExd1 dimer. Crystal structure of a ssRNA-yeastRrp6 complex, which provided the basis for the ssRNA modelled into BmExd1 is also shown. (C) Structure of a heptameric canonical Lsm core ring. One single unit is highlighted in orange. (D) Co-immunoprecipitation of endogenous BmTdrd12 and the PIWI protein Siwi with various HA-tagged BmExd1 deletion versions from transfected BmN4 cell lysates. Note that construct D which has a deletion of the N-terminal Lsm domain abolishes interaction with BmTdrd12, and as a consequence, with Siwi also. (E) Recombinant Lsm domain from BmExd1 pulls out HA-BmTdrd12 present in transfected BmN 4 cell lysates, while a control protein (6X-His-Thiredoxin) reveals only background levels. (F) N-terminal Lsm domain of mouse Exd1 is essential for interaction with mTdrd12 in transfected HEK293T cells. (G) Association of Myc-BmExd1 with various deletion versions of HA-tagged BmTdrd12. Deletions removing helicase domain of BmTdrd12 affects this association. (H) The helicase domain of BmTdrd12 is critical for this interaction. HA-tagged BmTdrd12 (construct F) carrying only the 1st Tudor domain and the entire helicase domain interacts with Myc-BmExd1. (I) MultiBac expression and purification of tagged BmTdrd12 and untagged BmExd1 in Sf21 insect cells; different independent preparations are shown.

Figure S5. Generation of Exd1 knock-out mouse, Related to Figure 5

(A) Genome targeting strategy for creating Exd1 knock-out mouse by insertion of a triple-stop codon cassette into exon 8 of the Exd1 coding sequence in the mouse genome. (B) Southern blotting of genomic DNA from mouse embryonic stem (ES) cells to confirm successful gene targeting. (C) Genomic PCR using mouse tail DNA to identify presence of targeted allele. (D) Read length profiles are shown for libraries prepared from testicular total small RNA (19-40 nt). Separate plots were created for reads which mapped uniquely to the genome and those mapped to multiple sites. Both
groups exhibit distinct abundance profiles in the Exd1−/− mutant. The peak corresponding to endogenous miRNAs is indicated. The 24-32 nt reads correspond to piRNAs in the total RNA libraries. There is an increase of uniquely mapping reads and a decrease in multi-mapping reads, in the libraries prepared from the the Exd1−/− mutant. (E) Western blotting to show expression of L1ORF1p in the testes of Exd1−/− mutant. Age of donor animals is indicated as days after birth (P). P0 indicates new-born pups. (F) Immunoprecipitation of Mili and Miwi2 from new-born pups (P0) and labelling of associated piRNAs by 5′-end labelling. Note the reduction in Miwi2 piRNAs in the Exd1−/− mutant. This is a replicate of the experiment shown in Figure 5I. (G) Mislocalization of nuclear Miwi2 to the cytoplasm in the Exd1−/− mutant embryonic (E18.5) testicular germ cells. (H) Analysis of 5′-to-5′ end distances of piRNAs present in the total testicular small RNA libraries prepared from P0 animals of the indicated genotypes. Reads that map to both strands of the LINE1 consensus were considered. A prominent peak (indicated with arrowhead) corresponding to a 10 nt overlap of 5′ ends is reduced in the Exd1−/− mutant. (I) Plot showing the fertility (number of pups/cross) of Exd1−/− mutant mice.

**Figure S6. Analysis of Exd1 (CG11263) mutant Drosophila, Related to Figure 6**

(A) Predicted length of Exd1 proteins in various indicated species. Sequence analysis indicates that Drosophilids lack the N-terminal region corresponding to the Lsm domain. (B) Cartoon showing domain architecture of Drosophila Exd1 (DmExd1) and the antigen used to prepare antibodies. Left: Purified rabbit polyclonal antibodies detect transiently expressed HA-DmExd1 in BmN4 cell lysates, but not the HA-BmExd1. Right: Anti-DmExd1 antibodies are capable of immunoprecipitating the fly protein transiently expressed in BmN4 cells. (C) Beta galactosidase (LacZ) staining of ovaries from flies with soma-specific expression of indicated knock-down hairpin. Note that reduction of armi and yb, two established piRNA biogenesis factors, results in activation of the LacZ transgene that is under control of endogenous piRNAs. In contrast, knock-down of DmExd1 with two independent hairpins did not elicit such an effect. (D) Transposons (blood and Het-A) are not de-repressed in fly ovaries expressing RNAi hairpins against exd1. RNAi hairpins were specifically expressed in the fly germline and total ovarian RNA collected for RT-PCR analysis of indicated transposons. (E) Left: Analysis of a minos-insertion line that we confirmed to be a knock-out for DmExd1 by Western analysis. Lysates (from a single ovary) were resolved in denaturing protein gel and probed with indicated antibodies. Note the absence of DmExd1 protein in the lane marked exd1 KO (knock-out). Right: RT-PCR analysis of total ovarian RNA indicates that transposons are not de-repressed in the knock-out line. Error bars represent s.d. from three independent experiments.

**SUPPLEMENTAL TABLE LEGENDS**

**Table S1: Deep Sequencing Data Used in this Study, Related to Experimental Procedures.**

All deep sequencing data generated in this study and deposited with the Gene Expression Omnibus (GEO) under the accession number GSE74423.
### Table S1. List of Deep-Sequencing Data Sets Generated or Used in this Study, Related to Experimental Procedures

| Code  | Library  | Organism   | Genome mapping reads (mm9) | GEO     |
|-------|----------|------------|-----------------------------|---------|
| RR214 | Exd1    | Mus Musculus | 10240106                    | GSE74423 |
| RR215 | Exd1    | Mus Musculus | 9384848                     | GSE74423 |
| RR216 | Exd1    | Mus Musculus | 7630814                     | GSE74423 |
| RR217 | Exd1    | Mus Musculus | 7568945                     | GSE74423 |
| RR218 | Exd1    | Mus Musculus | 7016516                     | GSE74423 |
| RR219 | Exd1    | Mus Musculus | 6298562                     | GSE74423 |
| RR220 | Exd1    | Mus Musculus | 5099241                     | GSE74423 |
| RR221 | Exd1    | Mus Musculus | 5348043                     | GSE74423 |
| RR222 | Exd1    | Mus Musculus | 15109900                    | GSE74423 |
| RR223 | Exd1    | Mus Musculus | 21220585                    | GSE74423 |
| RR224 | Exd1    | Mus Musculus | 20044578                    | GSE74423 |

| Code  | Library  | Organism   | all reads   | reporter sense mapping reads | GEO     |
|-------|----------|------------|-------------|------------------------------|---------|
| RR254 | Rosa26-pi reporter mouse; MILI IP from P0 testes | Mus Musculus | 51608864 | 45068 | GSE74423 |
| RR255 | Rosa26-pi reporter mouse; MIWI2 IP from P0 testes | Mus Musculus | 50863469 | 99357 | GSE74423 |

Reporter sense mapping reads refer only to the reads uniquely mapped to the reporter sequence of 25-31nt length.
SUPPLEMENTAL INFORMATION

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Clones and constructs

Constructs for BmN4 cell expression

For expression in Bombyx mori BmN4 ovarian cell cultures, the required coding sequences were cloned into pBEMBL vectors that drives expression from a OpIE2 promoter (immediate early promoter from Orgyia pseudotsugata multicapsid nucleopolyhedrosis virus) with an N-terminal HA or Myc tags (Xiol et al., 2012). Coding sequences for full-length Bombyx mori piRNA pathway factors were identified in our BmN4 polyA+ transcriptome data (Xiol et al., 2014), and cloned by reverse-transcription PCR (RT-PCR) from BmN4 total cellular RNA. Complete protein sequences are provided at the end of this document.

Full-length coding sequence for Bombyx Exd1 (BmExd1; 1-315 aa), Vreteno (Vret; 1-1362 aa), Papi (1-545 aa) and Spindle-E (Spn-E; 1-1362 aa) were cloned into pBEMBL-HA vectors. Constructs expressing full-length HA-tagged Siwi (1-899 aa) and Ago3 (1-900 aa) (Xiol et al., 2012), BmTdrd12 (Pandey et al., 2013) and Vasa (1-601 aa) (Xiol et al., 2014) are described. Note: The cDNA for BmTdrd12 (1-1759 aa) present in the transcriptome dataset encodes a protein of ~200 kDa. This is confirmed by Western blot analysis that detects the endogenous BmTdrd12 protein in BmN4 cell lysate as a band appearing between the molecular weight markers 180 and 250 kDa (Figure 2E). Our attempts to express a tagged version of this sequence in BmN4 cells were unsuccessful. So we expressed (Figure 2D) an N-terminal truncated version (260-1759 aa) and refer to this protein as full length (FL) in our experiments, as also described in (Pandey et al., 2013).

Deletion versions of Bombyx Tdrd12 (1-1759 aa) expressed in BmN4 cells:

Tdrd12-A (260-1619 aa); considered full-length in this study
Tdrd12-B (260-1282 aa)
Tdrd12-C (260-778 aa)
Tdrd12-D (260-570 aa)
Tdrd12-E (1407-1759 aa)
Tdrd12-F (260-1007 aa)

Deletion versions of Bombyx Exd1 (1-315 aa) expressed in BmN4 cells:

Exd1-A (1-228 aa)
Exd1-B (1-274 aa)
Exd1-C (32-315 aa)
Exd1-D (51-315 aa)
Exd1-E (79-315 aa)

Constructs for mammalian cell expression

Coding sequence for full-length (FL) Mus musculus (house mouse) MEXD1 (1-570 aa) was isolated by RT-PCR from mouse testis total RNA. Constructs expressing HA-tagged mouse TDRD12 (Pandey et al., 2013), and MILI and MIWI (Reuter et al., 2009) are described.

The following mouse sequences were cloned into the pcDNA3-Myc vector for expression in HEK293T cells:

mEXD1-FL (1-570 aa)
mEXD1-A (56-570 aa), lacks N-term Lsm domain
mEXD1-B (94-570 aa), lacks N-term Lsm domain
mEXD1-C (1-430 aa), lacks C-term
mEXD1-D (1-358 aa), lacks C-term

Constructs for bacterial expression
The coding sequence for Drosophila Exd1 (CG11263; 1-265 aa) was isolated by RT-PCR from fly ovarian total RNA. For expression in E. coli BL21 strain for production of antigens or recombinant proteins for other experiments, vectors (EMBL Protein Expression and Purification Core Facility) providing suitable affinity or solubility tags (SUMO, Thioredoxin etc) were used.

The following proteins were produced in E. coli for use as antigens in antibody production.

- pETM-22 (6xHis-Thioredoxin-3C-fusion)-Bombyx Exd1 FL (1-315 aa)
- pETM-11 (6x-His-TEV)-Drosophila Exd1 (1-56 aa)
- pETM-11-(6x-His-TEV)-Mouse mEXD1N-term antigen (1-126 aa) and C-term antigen (420-570 aa)

Recombinant Bombyx Exd1 proteins with different deletions were prepared in following bacterial expression vectors:

- pETM-22 (6xHis-Thioredoxin-3C-fusion)- BmExd1 FL (1-315 aa)
- pETM-28 (6xHis-SUMO-TEV-fusion)- BmExd1 (1-274 aa), monomeric form
- pETM-28-BmExd1 (73-315 aa), lacks N-term Lsm domain, dimeric form
- pETM-28-BmExd1 (73-274 aa), lacks N-term Lsm domain, monomeric form
- pETM-28-BmExd1 (73-291 aa), lacks N-term Lsm domain, dimeric form
- pETM-28-BmExd1 (73-301 aa), lacks N-term Lsm domain, dimeric form
- pETM-11-BmExd1 (1-74 aa) N-term Lsm domain alone

**Constructs for Bactulovirus-mediated expression in Sf21 or High Five (Hi5) cells**

For production of recombinant proteins in the insect cells the following ovary-derived cells were used: Sf21 or Sf9 from Fall Army worm Spodoptera frugiperda or High Five (Hi5) from the cabbage looper Trichoplusia ni. Expression of desired coding sequences was carried out with the use of recombinant Baculoviruses. Either single or multiple coding sequences were integrated into the Baculovirus genome using the MultiBac protein expression system (Bieniossek et al., 2012).

The following proteins were expressed in insect cells

- 6xHis-SUMO-Strep-TEV-BmTdrd12 (260-1759 aa) + untagged Siwi and untagged BmExd1
- 6xHis-SUMO-Strep-TEV-BmTdrd12 (260-1759 aa) + untagged BmExd1

Briefly, Tdrd12 (260-1759) coding sequence was cloned into pACEBac2Sumo acceptor vector and Siwi (1-899) was inserted into pIDC donor vector, while Exd1 (1-315) was placed in the pIDK donor vector. Two or three vectors were recombined to generate multi-gene expression plasmid using the Cre-Lox recombination strategy.

**Antibodies**

Antigens were produced in the E. coli BL21 strain and soluble proteins were purified using the relevant affinity-tag. If insoluble, the inclusion bodies were purified and then solubilized with denaturing agents like urea.

**Purification of soluble antigens:**

Refer to the prokaryotic protein production section for procedures used to prepare soluble proteins.

**Purification of insoluble antigens**

The nonconserved regions from the N-terminal or C-terminal of the proteins were chosen and were expressed in E.coli (BL21) as antigens. Usually the recombinant protein was produced as inclusion bodies. The bacteria cell pellet was resuspended in 1 x PBS (containing 1 mM PMSF), and was sonicated with MISONIX Sonicator S-4000. Cell lysate was centrifuged at 15000 x g for 40 min at 4°C. Pellet was then resuspended in buffer (100 mM Tris-HCl pH 7.0, 20 mM EDTA, 5 mM DTT), and was further centrifuged at 15000 x g for 40 min at 4°C. High salt wash buffer (50 mM Tris-HCl pH8.0, 60 mM EDTA, 6% (v/v) Triton X-100, 1.5 M NaCl and 5 mM DTT) was used to resuspend the pellet. After spinning down, High salt wash buffer without Triton X-100 was added to resuspend the cell pellet to remove the remaining Triton. The pellet was finally dissolved in 8 M urea with agitation overnight.
Affinity purification of antibodies
When indicated, the antibodies were affinity purified from the crude immune serum using an affinity column prepared with the antigen. For insoluble antigens, large amounts of purified antigens were resolved by SDS-PAGE and blotted on nitrocellulose membrane by semi-dry western transfer. After reversible staining with Ponceau S (P3504, Sigma), part of the membrane containing the antigen was cut out and incubated overnight with crude immune sera. After washes (1xPBS), bound antibodies were eluted with low pH solution (500 µL 0.1 M Glycine pH 2.5, 150 mM NaCl) and neutralized with neutralization buffer (150 µL 0.5 M Tris-HCl pH 8.0, 150 mM NaCl). Antibodies were stored in 50% glycerol at -20°C and used over a period of several months.

For soluble antigens, CnBr-activated bead (Roche; Cat. no 17-0430-01) was used to purify the antibody. The antigen was first incubated with CnBr-activated beads in PBS (0.1% Tween) overnight at 4 °C. The unbound antigen was washed away by PBS. The antigen matrix was then incubated with crude serum in PBS (0.1% Tween) overnight at 4°C. The bead-slurry (bound with antigen and antibody) was loaded to column and the solution was allowed to drain by gravity flow. The bead was further washed with PBS several times. The antibody was finally eluted and neutralized as described above.

Antigens for antibodies prepared in this study
Bombyx Exd1 (1-315 aa): soluble 6xHis-Thiredoxin fusion was purified over Ni-NTA column and tag was removed with the 3C protease. Untagged soluble protein was used as an antigen.
Drosophila Exd1 (1-56 aa): insoluble
Mouse MEXD1: N-term antigen (1-126 aa) and C-term antigen (420-570 aa): insoluble

Antibodies used in this study from previous work
Antibodies for Bombyx Tdrd12 (rabbit polyclonal) (Pandey et al., 2013), Bombyx Siwi and Ago3 (rabbit or rat polyclonal) (Xiol et al., 2012), mouse MILI (mouse monoclonal) (Reuter et al., 2009) and MIWI2 (rabbit polyclonal) (Pandey et al., 2013) are described. Mouse anti-Myc monoclonal antibody (clone 9E10) was produced at the EMBL Monoclonal Antibodies Core Facility (MACF), Monterotondo, Rome, Italy. Mouse anti-HA monoclonal was a kind gift of Marc Buhler, FMI, Basel, Switzerland. Mouse anti-HA monoclonal was a kind gift of Marc Buhler, FMI, Basel, Switzerland.

Commercial antibodies
Anti-His (Amersham; 27-4710-01), rabbit anti-HA (Santa Cruz; sc-805), anti-ß-Tubulin (Abcam; ab6046) for detecting Drosophila Tubulin, anti-HA affinity matrix (Roche; cat. No. 11815016001) for immunoprecipitations, were purchased. For immunofluorescence studies the following secondary antibodies coupled to fluorescent dye were used: anti-rabbit (Alexa Fluor 488, cat. No. A11034), anti-rabbit (Alexa Fluor 594, cat. No. A11037) and anti-mouse (Alexa Fluor 594, cat. No. A11005). For Western blot analyses the following secondary antibodies conjugated to Horse Radish Peroxidase were used: anti-rabbit IgG HRP-linked antibody (GE Healthcare; NA934), anti-mouse IgG HRP-linked (GE Healthcare; NA931) and anti-rat IgG HRP-linked (Santa Cruz Biotechnology, cat. No. sc-2032).

Cell culture, immunoprecipitations and immunofluorescence
BmN4 cell culture
The Bombyx mori (Silkworm) ovarian cell line BmN4 (Kawaoka et al., 2009) was maintained in IPL-41 media (Invitrogen catalogue no. 11405057) supplemented with 10% fetal bovine serum (Invitrogen; Catalogue no. 16000044) and 0.2% Penicillin:Streptomycin (Life technology, cat. No. 15140122). The adherent cells were cultured in a temperature-controlled incubator at 27°C. Initially, cells grow as a monolayer with well-defined, elongated morphology that attach to the plate surface. After reaching confluence, a second layer of rounded cells grow on top of the monolayer. Upon reaching the second layer, cells were passaged and monolayer cells at 80-90% confluency were used for transfections. For immunoprecipitations, cells were grown in 6-cm petri dishes and
transfected (X-tremeGENE HP DNA Transfection Reagent, Roche, cat. No. 6366236001) with ~2 μg of expression plasmids. For immunofluorescence analysis, glass coverslips are placed in 6-well or 12-well plates before seeding cells.

For immunoprecipitations, cells were washed once with PBS, and 1 mL of lysis buffer [50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5% Triton X-100, 0.1% NP-40, protease inhibitor (Complete Protease Inhibitor Cocktail Tablet, Roche, cat. No. 5056489001)] was added to the petri dish, and cells were removed from their growth surface using a cell scraper (Cell scraper, Costar; catalogue no. 3010). Cells were transferred to a glass tissue homogenizer and were further lysed by douncing (about thirty strokes). The total cell lysate was spun at 10,000 rpm for 10 min at 4°C. Cleared lysates were incubated with 15 μl of anti-HA affinity matrix for 3 hours at 4°C. Beads were collected by gentle centrifugation, 600 x g for 1 min at 4°C. After five washes with wash buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% NP-40), bead-bound complexes were used for protein or RNA analyses. When required, immunoprecipitated complexes were treated with 1-2 μl RNase A (Life technology, cat. No. EN0531) and 1-2 μl RNase T1 (Life technology, cat. No. EN0541) for 30 min at 4°C (for harsh treatment, leave at room temperature) prior to analyses. Complexes were separated by 10% SDS-PAGE and examined by Western blotting using specific antibodies.

To isolate associated RNAs, proteins in bead-bound complexes were digested with Proteinase K (50°C, 30 min) and RNAs was extracted with phenol-chloroform, and precipitated in ethanol. To visualize RNAs, they were dephosphorylated with rAPid alkaline phosphatase (recombinant bovine phosphatase; Roche, cat no. 4898141001) and 5′-end labeled with γ-[32P]-ATP with T4 polynucleotide kinase (Thermo Scientific; EK0031). The labeled RNAs were resolved by 15-20% (w/v) urea-PAGE. Gels were exposed to Phosphor Storage screens (GE Health) and scanned (Typhoon scanner; GE Health).

For immunofluorescence studies, BmN4 cells were grown on coverslips. Cells were washed once with 1 x PBS, and then fixed with 4% paraformaldehyde for 10 min. Cells were washed twice in 1 x PBS and incubated in 0.1 M Glycine for 4 min. After washing twice in 1 x PBS, cells were permeabilized in 0.1% Triton X-100 for 4 min, and followed by incubation in blocking solution (0.2% BSA, 1% goat serum, 0.1% Triton X-100 in PBS) for another 20 min. Proteins were detected by incubating with primary antibodies to specific proteins or anti-tag antibodies for 1hr. The secondary antibodies coupled to fluorescent dye were used and incubated for another 1 hr. DAPI (0.5 μg/ml, Bio-Rad, cat. No. 10043282) were used for nuclei detection. The cells were mounted with a drop of Slowfade Gold Antifade Reagent (Life technology, cat. No. S36942), and were examined using a Leica TCS SP2 AOBS, inverted confocal microscope.

HEK293T mammalian cell culture

Human embryonic kidney cell line 293 (HEK293) transformed with the SV40 large T antigen (HEK293T) were grown in Dulbecco’s modified Eagle Medium (DMEM; Invitrogen, cat. No. 21969-035) supplemented with 10% fetal calf serum (Invitrogen, cat. No. 16000044) and 1% Penicilline/Streptomycin (Life technology, cat. No. 15140122), and maintained in an environment with 5% CO2 at 37°C. For transfection, cells growing in a 75 cm² flask were washed with warm (37°C) 1X PBS and incubated with 1 ml of Trypsin (Life technology, cat. No. 25300-054) for 5-10 seconds to promote removal of cells from the growth surface. Trypsin was aspirated and the cells were incubated for another 1 min at room temperature. 10 mL warm DMEM media was added to remove and passage the cells by pipetting and approximately 500 μL cells were seeded into a 6 cm dish and grown overnight till they are 50-60% confluent. Plasmid was transfected with Lipofectamin (Invitrogen, cat. No. 18324012) and Plus™ reagent (Invitrogen, cat. No. 11514015). About 2 μg plasmids was mixed with 30 μL serum-free media and 1.5 μL Plus reagent, and incubated at room temperature for 15 min. 1.5 μL Lipofectamin was mixed with 30 μL serum-free media. The two mixture systems were combined together, and incubated at room temperature for another 15 min. Media in the petri dish was replaced with pre-warmed serum-free media, and the mixture was added to the petri dish. The serum-free media was replaced with complete media (with serum) after 5 hours of incubation with transfection mix. Cells in each 6 cm dish were harvested 48 hours post-transfection. The immunoprecipitation was performed as described above.
Mass spectrometry

Complexes containing HA-BmTdrd12 were isolated from transfected BmN4 cells using the anti-HA affinity matrix. Proteins were eluted by adding 30 μL 5 x SDS-loading buffer (250 mM Tris-HCl pH 6.8, 10% SDS (W/V), 50% (V/V) glycerol, 0.5% (W/V) bromophenol blue and 5% 2-mercaptoethanol) and heating to 95°C for 3 min. Proteins present in the complexes were identified by mass spectrometry (EMBL Proteomics Core, EMBL Heidelberg).

Recombinant protein production

Prokaryotic expression and purification

The full-length *Bombyx* Exd1 (1-315 aa) was expressed from a pETM-22 vector as a His-Thioredoxin-fusion protein using the *E. coli* BL 21 strain. Protein expression was induced by addition of 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) when the bacterial culture density reached ~0.6 (OD600). The proteins were expressed overnight at 18°C following induction. Cells were collected by centrifugation and lysed by sonication in lysis buffer [25 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM 2-mercaptoethanol, 20 mM Imidazole and protease inhibitor (Roche complete EDTA-free)]. Soluble proteins were isolated from the supernatant fraction via Ni2+-affinity chromatography. The 6 x His-Thioredoxin tag was removed with the His-3C protease (EMBL Protein expression and purification facility). The cleaved tag and protease was removed by passing through the second time of Ni2+-affinity column. The protein was further purified on an ion-exchange column (HiTrap™ Q HP, GE healthcare, 17-1154-01) and mono-dispersed fractions collected by gel filtration chromatography (Superdex S200 10/300GL, GE healthcare, cat. No. 17-5175-01). Analytical ultra-centrifugation experiments indicated the protein to exist as a homodimer (Figure S3F). A similar protocol was used for purification of all other proteins used for biochemical, biophysical and structural studies.

To obtain well-behaved core structural domains, the full-length *Bombyx* Exd1 protein was subjected to limited proteolysis with Trypsin (Figure S3B). Approximately 120 μg of protein was incubated with 120 ng Trypsin proteases (Sigma, cat. No. T4799) at 22°C, and an aliquot (20 μg) was removed at different time points: 0, 20, 40, 60, 80, 120 min. The reactions were resolved by 10% SDS-PAGE, stained with Coomassie Blue, and the resulting protein fragments were identified by acid hydrolysis and mass spectrometry (EMBL Proteomics Core Facility, EMBL Heidelberg). Several fragments were expressed in *E. coli* BL21 strain from the pETM-28 (N-terminal 6XHis-SUMO-tag) (Figure S3B). One of these, one fragment (73-315 aa) was robustly expressed and yielded crystals of high quality.

The fragment of BmExd1 (73-315 aa) that was crystallized was purified by Ni2+-affinity chromatography with lysis buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM 2-Mercaptoethanol, 20 mM Imidazole and protease inhibitor), followed by removal of the His-SUMO tag with the TEV protease (EMBL Protein expression and purification facility). The protein was further purified on a SP ion-exchange column (HiTrap™ SP HP, GE Healthcare, cat. No. 17-1152-01) equilibrated with 25 mM MES pH 6.0, 50 mM NaCl, 5 mM 2-Mercaptoethanol. Mono-dispersed fractions were collected by gel filtration chromatography (Superdex 75 10/300GL, GE healthcare, cat. No. 17-5174-01) with 25 mM MES pH 6.0, 150 mM NaCl and 5 mM DTT.

Insect cell expression and purification

*Bombyx* Tdrd12 (260-1619 aa) when expressed on its own in insect cells was poorly expressed and usually found in complex with cellular heat shock proteins, indicating problems in protein folding. However, when co-expressed with BmExd1 using the MultiBac insect cell expression system (Bieniossek et al., 2012), reasonable amounts of the complex were obtained. Briefly, we first inserted the BmTdrd12 cDNA into the pACEBac2-Sumo acceptor vector (His-Strep-Sumo tag), and BmExd1 (1-315 aa) into the pIDK donor vector. These two vectors were then fused together by Cre-Lox recombination.

After transformation of DH10EMBacY competent cells, recombination with the baculovirus genomic DNA (bacmid) resulted in recombinant bacmid DNA formation. The bacmid DNA was extracted and transfected into Sf21 insect cells. The supernatant (V0) containing the recombinant baculovirus was collected 48 hours post transfection. To monitor expression of the protein, cells were washed and collected in 1X PBS around 96 hours post-transfection.
After sonication (to break up genomic DNA), lysates were resolved by 10% SDS-PAGE and proteins were detected by Western analysis (anti-His for BmTdrd12 and anti-BmExd1). To expand the virus pool, 3.0 ml of the Vₜ virus stock was added into 25 ml of S21 (0.5 x 10⁶/ml) cells. The resulting cell culture supernatant (Vₜ) was collected 24 hours post-proliferation arrest. For large scale expression, High Five cells were infected with virus (Vₜ) at an appropriate ratio (1:400-1:800). Cells were harvested 72 hours post-proliferation arrest. The cells were lysed (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM 2-Mercaptoethanol, 20 mM Imidazole, 0.5% Tween-20, 10% Glycerol and protease inhibitor), and the His-Strep tagged protein were purified over Ni²⁺-affinity and StrepTrap HP column (GE Healthcare, cat. No. 28-9075-46). The tagged protein was further purified over the ion exchange column (HiTrap™ Q HP, GE healthcare, 17-1154-01). Fractions containing the recombinant proteins were further purified by gel filtration chromatography (Superdex S200 10/30GL, GE healthcare, cat. No. 17-5175-01). The pure fractions of BmTdrd12 were used for biophysical measurements and ATPase assays.

Crystallization, structure determination, and refinement of Bombyx Exd1 core nuclease domain

Full-length Bombyx Exd1 (1-315 aa) did not yield crystals. However, a version identified by limited proteolysis and encompassing the core nuclease domain (73-315 aa) gave initial needle shaped crystals in multiple conditions at 4°C by robotic screening. Most diffracted poorly but the best diffracting conditions included divalent Mg²⁺ or Ca²⁺ ions (eg. 10 mM CaCl₂, 200 mM KCl, 50 mM Na cacodylate, pH=6, and 10% (w/v) PEG 2K). Subsequent refinement searches around these initial crystallization conditions, specifically the substitution of Ca²⁺ for larger divalent metal ions, lead to large block shaped crystals grown in the presence of 1-5 mM Gd³⁺ or Sm³⁺. These crystals were flash frozen at 100 K after transferring them to identical crystallisation conditions containing 25% glycerol, and routinely diffracted to high resolution. A highly redundant single anomalous dataset of a Gd³⁺ co-crystal was collected to 1.6 Å resolution. The Gd³⁺ co-crystal crystallised in space group P2₁2₁2 with two molecules in the asymmetric unit. The Exd1 core nuclease domain was observed to bind RNA. In an effort to determine an Exd1-RNA complex structure a co-crystallisation trial with 10U-RNA was performed using robotic screening at 4°C. Several crystallization conditions were observed, some similar to the previous conditions above. One new condition was identified and refined and was composed of 50 mM Tris-HCl pH 8.5, 5 mM MgSO₄, and 35% (w/v) 1,6-hexanediol. These crystals were subsequently flash frozen at 100 K in identical conditions containing 40% (w/v) 1,6-hexanediol and a complete dataset to 2.4 Å was obtained. All data was collected on ID29 (de Sanctis et al., 2012) at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Integration and scaling was carried out with the XDS suite (Kabsch, 2010) and AIMLESS (Evans, 2006), respectively.

The large anomalous signal observed in the Gd³⁺ co-crystals allowed for the experimental phasing of the Exd1 core nuclease domain using this dataset. For the structural solution three Gd³⁺ sites were located on the basis of their anomalous differences using SHELXC/D/E (Sheldrick, 2010). These sites were refined and experimental phases to 1.6 Å were calculated using the single anomalous dispersion (SAD) procedure in SHARP (de La Fortelle, 1997). These phases were further improved by density modification followed by model building with Buccaneer (Cowtan, 2006). The U10 RNA co-crystal structure was solved using a single endonuclease core domain from the Gd³⁺ co-crystal (73-274aa from molecule 1 and 275-315 from molecule 2) with Phaser (McCoy et al., 2007). No RNA was observed in these co-crystals so this dataset was treated as a native structure. All subsequent refinement cycles were performed using REFMAC5v (Murshudov et al., 2011) with NCS restraints and TLS refinement options enabled. Model building was carried out with Coot (Emsley et al., 2010) and the stereochemical quality of the protein molecules were validated with Molprobity (Chen et al., 2010). All crystallographic coordinates have been deposited with in the Protein Data Bank (PDB code for native and Gd²⁺ crystals are 5FIQ and 5FIS respectively).

In vitro pull-down Lsm domain (BmExd1) with HA-BmTdrd12

To demonstrate interaction between HA-BmTdrd12 and the recombinant Lsm domain of BmExd1, we first bound the 6xHis-BmExd1 Lsm domain (1-74 aa) or a control protein (BSA) to Ni-NTA beads (Qiagen, cat. No. 30230). Beads were washed once with wash buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% NP-40) to remove the unbound proteins. The beads with bound proteins were then incubated with cleared cell lysates from BmN4 cells expressing
HA-BmTdrd12. The mixture was incubated at 4°C for 3 hours on a rotating wheel, after which the beads were washed with buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% NP-40) five times. The beads-bound proteins were resolved by 10% SDS-PAGE and proteins detected by Western blotting (for HA-BmTdrd12) or by staining with Coomassie blue (for detecting bait proteins) (Figure S4E).

**Nuclease assay with Exd1**

The single-stranded RNAs (ssRNAs) (Microsynth, Switzerland) used in the nuclease assays are:

- RNA1: UGACAUGGAACAGGUGCUCAGAUAGCUUU (30 nt)
- RNA2: GGGCGAGAAAGCUAUCUGAGCACCUGUGUUCAUGUCAGCAU (40 nt)
- RNA3: GGGCGAGAAUCCCUAGGAGAAAAACUAUGACCUAGCAUCAGCAU (40 nt)
- RNA4: GGGCGAGAAUCUAGGCCCAUUCACUAUAGCAUGCAU (39 nt)
- RNA5: GGGCGAGAAACCACUUCAUUCACAGUGCACUCAGCAU (39 nt)
- RNA6: GGGCGAGAAUCUGCCAGUCAUCUUCCAGCAU (38 nt)

The RNAs were labeled at the 5′ end with [γ-32P]ATP and T4 polynucleotide kinase (Thermo Scientific EK0031). The labeled RNAs were gel-purified to eliminate presence of contaminating degradation fragments. For each reaction, ~0.02 pmol of RNA substrate was mixed with ~50 pmol of purified recombinant BmExd1 proteins [FL (1-315 aa), ΔN (73-315 aa) and ΔC (1-274 aa)], and incubated at 27°C for 1 hour. Two different buffers were used in the assay: Tris buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM CaCl2, 2 mM DTT) or MES buffer (25 mM MES-NaOH pH 6.5, 150 mM NaCl, 2 mM CaCl2, 2 mM DTT). The reaction mixture was then treated with Proteinase K (0.1 mg/ml final concentration, 50°C, 30 min). RNA was extracted by phenol-chloroform and precipitated with ethanol, followed by migration in 15% denaturing urea-polyacrylamide gels. After electrophoresis, the gel was dried and exposed to Phosphor Storage screen (GE Healthcare).

**UV cross-linking assays**

Proteins (2 µg) were incubated for 20 min at room temperature with 0.02 pmol of synthetic 5′-[32P]-labeled 10-mer poly(U) RNA (Microsynth) in buffer (10 mM Tris-HCl pH 7.5, 50 mM KCl, 1 mM DTT, 10% Glycerol) in a final volume of 20 µl. The solution mixture was then carefully deposited on the inside of an eppendorf tube kept on ice, and irradiated for 5 min with a 254 nm UV lamp (Stratalinker 1800, Stratagene). The distance between the UV lamp and the sample is about 2 cm. After irradiation, samples were boiled for 5 min in SDS loading buffer and resolved by 10% SDS-polyacrylamide gel electrophoresis. After electrophoresis, the gels were first stained with Coomassie Blue (for visualizing proteins) and then exposed to Phosphor Storage screens (GE Health) and scanned (Typhoon scanner; GE Health).

**Electrophoretic Mobility Shift Assay (EMSA)**

The single stranded RNAs (Microsynth) used are listed below. Approximately 0.02 pmol of 5′-[32P]-labeled RNAs were incubated with ~50 µmol purified recombinant proteins for 30 min at room temperature, and 1 hour on ice in buffer 10 mM Tris-HCl pH 7.5, 10% Glycerol, 50 mM KCl, 1 mM DTT in a final volume of 20 µL. The samples were resolved by 10% native polyacrylamid gel electrophoresis. The electrophoresis was performed at room temperature, but run at 5 V (0.3 V/cm) to avoid heat-generation during the run. After electrophoresis, the gels were dried and exposed to Phosphor Storage screen (GE Healthcare).

- RNA1: UGACAUGGAACAGGUGCUCAGAUAGCUUU (30 nt)
- RNA2: GGGCGAGAAAGCUAUCUGAGCACCUGUGUUCAUGUCAGCAU (40 nt)
- RNA3: GGGCGAGAAUCCCUAGGAGAAAAACUAUGACCUAGCAUCAGCAU (40 nt)
- RNA4: GGGCGAGAAUCUAGGCCCAUUCACUAUAGCAUGCAU (39 nt)
- RNA5: GGGCGAGAAACCACUUCAUUCACAGUGCACUCAGCAU (39 nt)
- RNA6: GGGCGAGAAUCUGCCAGUCAUCUUCCAGCAU (38 nt)
**Isothermal Titratation Calorimetry (ITC)**

ITC experiments were performed using a Microcalorimeter ITC 200 (GE Healthcare) at 25°C. Different versions of BmExd1 proteins were prepared as described above. The RNA (5'-GGGGUAUCUUAAUUUUC-3') was dialyzed in the same buffer as that used for protein (25 mM Tris-HCl pH 7.2, 150 mM NaCl, 5 mM 2-mercaptoethanol). Approximately 200 µM RNA was titrated into the sample cell containing ~20 µM protein. Titrations were carried out by constant volume injections (39 injections of 1.0 µL) with 120 sec spacing. Data analysis was done using Origin 7 program. RNA was purchased from Microsynth AG (Switzerland).

**Sedimentation Velocity measurements**

Purified proteins were subject to analytical ultra-centrifugation (AUC) on an analytical ultracentrifuge XLI (Beckman Coulter, Palo Alto, USA) with a rotor speed of 42000-60000 rpm (depending on the protein being studied), at 10°C (Plateforme Biophysique, Partnership for Structural Biology, Grenoble, France). Acquisitions were made using absorbance at 280 nm and interference optics. The buffer used for dissolving the protein was used as reference: 25mM Tris-HCl pH 8.0, 150mM NaCl, 0.5mM DTT [for BmExd1 (1-74 aa)] or 20mM Tris-HCl pH 8.0, 300mM NaCl, 5% Glycerol, 5mM DTT [for BmExd1-FL or deletion versions]. The analyses were done with the SEDFIT software.

**Drosophila experiments**

**RNAi lines**

The following double-stranded RNA (dsRNA) hairpin constructs targeting genes of interest were obtained from VDRC (Vienna Drosophila Resource Center, Vienna, Austria) (Dietzl et al., 2007): zuc (48764GD), armi (16205GD), spn-E (21374GD), fs(1)Yb (23437GD), DmExd1/CGL1263 (31112GD and 31113GD), BoYb (18149GD), or from the NIG (National Institute of Genetics Fly Stock Center, Japan): DmExd1/CGL1263 (11263-R1 and 11263-R2), BoYb (11133-R1), SoYb (31755-R1).

**Monitoring of piRNA pathway activity in knock-down flies**

RNAi knock-down experiments in either the somatic follicle cells or the germline of Drosophila ovary was performed as previously described (Handler et al., 2011; Olivieri et al., 2010). Briefly, directed expression in the ovarian somatic follicle cells was achieved with the soma-specific driver line tj-GAL4 (traffic jam) (insertion P[w+wm.w.hs]=GavBjNP1624), carrying the reporter construct gypsy-lacZ (in which 5' UTR has sequences for the fly transposon gypsy) (Olivieri et al., 2010; Sarot et al., 2004). Specific expression in the fly ovarian germ-line was achieved with the germ-line-specific nos-GAL4 (nanos) (insertion P[GAL4-nos.NGT]) driver line carrying a UAS-Dcr (Dicer2) transgene (P[UAS-Dcr-2.D]). In this strain, the expression of Dicer2 double-stranded RNA ribonuclease promotes RNAi from hairpins in the germine, as described (Handler et al., 2011). The corresponding stocks were a kind gift from Julius Brennecke (IMBA, Vienna): tj-Gal4; gypsy-lacZ (Olivieri et al., 2010) and UAS-Dcr2; NGT[α].Gal4 (Handler et al., 2011).

Impact of gene silencing in the somatic follicle cells was monitored by X-Gal staining of isolated ovaries using classical methods to measure activity of β-gal encoded by the gypsy-lacZ. In wildtype cells, antisense piRNAs to gypsy transposon silence the reporter, hence no signal (blue colour) is detected in ovary staining. In contrast, a defective piRNA pathway allows for gene transcription and production of active β-gal, hence a blue signal is detected in ovary staining (Figure S6C). Stained ovaries were examined under an Axiovision ZI Zeiss wide-field microscope and images were recorded. In the case of the germline, integrity of the piRNA pathway was monitored by determination of transposon transcript levels by reverse-transcription quantitative PCR (RT-qPCR) analyses (Brilliant II SYBR Green qPCR master mix, Stratagene, cat. No. 600828) from total RNAs isolated from dissected ovaries. Ovaries (3 pairs) from different genotypes were collected and placed in Trizol for RNA extraction and subsequent analysis by qPCR. The list of primers used for RT-PCR analyses are given below.

**Fertility testes**
For female fertility assays, 10 females of each genotype were individually crossed with 5 wild type males. Silencing DmExd1 expression in the female germline alone (11263-R2 or 31112-GD) or in combination with either BoYb (31112-GD, 18149-GD recombinant) or SoYb (11263-R2, 31755-R1 recombinant) by using specific shRNA and the driver line UASDcr;NGTGal4, did not affect female fertility. As expected, silencing both BoYb and SoYb jointly (11133-R1, 31755-R1 recombinant) resulted in almost complete female sterility (9/10 female with no progeny, 1/10 female with less than 20 progeny). Therefore, Exd1 may be non-essential for fertility in normal conditions (ie, standard culture medium at 25°C).

Characterization of Drosophila Exd1 MI07943 null allele

The fly line Exd1MI07943 (y[1] w*; Mi[y+;Dint2]=MIC CG11263[M107943]) carrying an insertion in the Exd1 locus was obtained from a collection of Minos transposon-mediated integration cassettes (MiMIC) in Drosophila (Venken et al., 2011). The stock was obtained from BDSC (Bloomington Drosophila Stock Center) (BL#44902). Western blot analysis of mutant ovary lysates confirmed the complete absence of the protein (Figure S6E). In addition, the insertion MiCG11263[M107943] in the Exd1 locus is fully viable and fertile, either alone or in combination with the insertion P-EP-SoYb[419414], a P element inserted in SoYb coding sequence which likely disrupts its expression.

Total protein extracts were prepared from isolated ovaries (1 pair) and western blots were probed using rabbit polyclonal anti-DmExd1 antibodies. Tubulin was used as loading control. Ovaries from different genotypes were collected and placed in Trizol for RNA analysis. The list of primers used for RT-PCR analyses are given below.

Primers used for qRT-PCR in this study:

| Primer          | Sequence                      |
|-----------------|-------------------------------|
| HeT-A_forward   | CGCCGCGAACCATCTTCAGA          |
| HeT-A_reverse   | CGCCGCAGTCGTTTGGTGAGT         |
| blood_forward   | AACAATAGAAAGAACCCCGGAGC      |
| blood_reverse   | AGTCAATGAGAAGATTGGGTG        |
| actin 5C_forward| AAGTTGCTGCTCTGTTGTCG          |
| actin 5C_reverse| GCCACACGCGTCACTTAGG          |

Generation of Exd1 mouse mutant

The Exd1 gene locus is on chromosome 2 of the mouse genome and is composed of 13 exons, with the translated sequence being provided by exons 3-13. The N-terminal Lsm domain is contributed by exons 3-5, while the nuclease domain is encoded by exons 8-12. The Exd1 locus in the hybrid C57BL/6 × 129Sv embryonic stem (ES) cell line A9 was disrupted using a targeting construct that inserts an in-frame triple stop codon cassette immediately downstream of the first codon in the exon 8 (ATG TAG A TAG A TGA) (Figure S5A). This also inserts the PGK-Keo (Kanamycin-neomycin)-polyA cassette downstream of the stop codons. In the targeted allele, the translation is prematurely terminated at the beginning of exon 8, probably leading to nonsense-mediated decay (NMD) of the mutant transcript. In addition, the presence of a strong polyA site from the selection cassette will also truncate the transcript (deleting everything downstream of exon 8).

Electroporated A9 ES cells were selected (neomycin; G418) and 300 clones screened by Southern blotting with probes recognizing the 5’ (EcoRI-digested DNA) and 3’ (HindII-digested DNA) regions flanking the targeted site (Figure S5B). DNA ladder used for Southern probing is a 1kb ladder (NEB), as indicated in the figure. Only 7 ES cell clones that were detected by both probes were retained. These ES cell clones were karyotyped for genome integrity. One clone (#3F7A) was used for injection into C57BL/6N host embryos for mouse generation using 8-cell stage injection. Founder animals were identified by coat color and genotyped by PCR of tail genomic DNA to detect insertion of the 3xStop-Keo cassette as shown below. Three founder ES mice carrying the targeted allele were obtained. Backcrosses with C57BL/6j Rj (Janvier labs) wildtype females were performed to transfer the mutation to.
the C57BL/6 background and to obtain heterozygous mice carrying the mutated allele \textit{Exd1}_{Stop-neo} (hereafter referred to as \textit{Exd1}^{+/−}). Heterozygous males and female siblings were intercrossed to obtain homozygous null \textit{Exd1}^{−/−} mutants.

\textbf{Southern probe sequences}

\textbf{5′ probe (420 bp)}

\begin{verbatim}
TGAGAATGAAGGTCTAGGCAAAGAGGTGGGATTGGTAGTCACTGAAGAGTTTACACACCCTTAAAGGGATGAGGACTCTAAAGAGAGAAGGGAGGTGAAGCACTGGAGAGCTTTTTTGGGTTTTG
\end{verbatim}

\textbf{3′ probe (511 bp)}

\begin{verbatim}
GGCCAGGGTGGAGCCCAGGGCCTGGGGCATACTAATACCTCTGATGTGAGATCCTTATCTGTGTATATTACCT
AATCCTCTTTTGTTGTTGTGTGATTGTTTTCTGTTTCATACCTCAGGTAGCCGACGTCCTTCAATTTTCTATGGAAACGGGTGGCTTTCTTCCAAACTGTATCAGTACTTTGCAGGAGAGTTAATCAGACACCTTAAAGTTGCTCCCAGATACCTTTTCTTTTTAGAAGAGAGACAGAAACGTATTCAGGTAAGTATTGAAAGGTGTGTCACTACATTTCAGGTGTGTTCTGCACAGAACTGTTAGTCTGTTGGA
\end{verbatim}

\textbf{Tail genomic DNA isolation and genotyping PCR conditions}

Tails were digested in 500 µl tail buffer (50 mM Tris-HCl pH 8.0, 100 mM EDTA, 100 mM NaCl, 1% SDS) and 2.5 µg Proteinase K at 55°C overnight. Spin at maximum speed for 10 min to remove hairs. Transfer supernatant by pouring into a new tube. Precipitate DNA by adding 500 µl Isopropanol (turning tube upside down until white precipitates appears). Spin at maximum speed for 10 min, remove supernatant and wash pellet in 1 ml 70% Ethanol. Pellet was allowed to be dried and resuspended in 100 to 150 µl 10 mM Tris-HCl pH 8.0 for at least 1 hour at 37°C. Use 1-1.5 µl for PCR.

\textbf{PCR primers used for genotyping \textit{Exd1} mice:}

\begin{verbatim}
CAGGATTTCTCTGTGTGTTCCTTC Exd1_forward
CTGAAGCCACGAGTCTCCATG Exd1_reverse
CCATCTGCACGAGACTAGTG Exd1_reverse_2
\end{verbatim}

\textbf{PCR reactions:}

Reaction volume is 25 µl [1X Taq buffer, 0.2 mM dNTP, 0.25 mM primers (primer mix stock is 20 mM each), 1-1.5 µl template genomic DNA and 0.5 µl Taq DNA polymerase.]

\textbf{PCR program:}

1. 94°C, 3 min
2. 58.5°C, 1 min
3. 72°C, 2 min
4. 94°C, 30 sec
5. 61°C, 40 sec
6. 72°C, 30 sec (Repeat steps 4. to 6 for 35 times)
7. 72°C, 3 min
8. 15°C, overnight.

Load 12 µl reaction in 1.5% Agarose gel (1X TBE) (Figure S5C)

\textbf{Sequenced PCR products from wildtype \textit{Exd1} allele (120 nt)}

\begin{verbatim}
CAGGATTTCTCTGTGTGTTCCTTC
CTGAAGCCACGAGTCTCCATG
CCATCTGCACGAGACTAGTG
\end{verbatim}

\textbf{Sequenced PCR products from \textit{Exd1} knock-out allele (329 nt)}

\begin{verbatim}
CAGGATTTCTCTGTGTGTTCCTTC
CTGAAGCCACGAGTCTCCATG
CCATCTGCACGAGACTAGTG
\end{verbatim}

\textbf{ATG: 1\textsuperscript{st} codon of Exon 8 from \textit{Exd1} ORF}

\textbf{TAAATAGATGA: STOP cassette insertion}

\textbf{PCR PRIMER POSITIONS}

\textbf{Generation of \textit{Rosa26-pi} reporter knock-in mouse}

For driving ubiquitous expression in the mouse system, we targeted exon 1 of the \textit{Rosa26} locus in the mouse embryonic stem cell (ES) A9 line with a cassette carrying a piRNA reporter (Figure S1A). The reporter contains the DsRed2 fluorescent protein coding sequence followed by a long 3′ UTR based on the LacZ backbone (rendered
noncoding by removing all ATGs) where we placed perfectly complementary sites for 35 independent MILI-bound piRNAs. The piRNAs we chose are those expressed abundantly in the embryonic/perinatal mouse male germline where PIWI proteins MILI and MIWI2 are co-expressed. The expectation is that the reporter would be expressed everywhere in the mouse, but silenced only in the embryonic/new-born testes (where complementary piRNAs are produced).

The targeting cassette had the following elements: First, the dsRed coding sequence (Kozak-ATG-dsRed-STOP). Second, immediately downstream of the dsRed reporter we placed perfectly complementary binding sites for 35 independent piRNAs abundantly present in MILI complexes isolated from embryonic/new-born testes. Each piRNA binding site was separated from the neighbouring ones by ~50 nt and the overall length of this region was 2750 bp. Third, flanking the region with complementarity to piRNAs, we placed LoxP sites to enable its eventual removal by the Cre recombinase (not relevant for this study). Fourth, downstream of the last LoxP, we inserted the SV40 polyA signal (131 bp) from pcDNA3.1. Fifth, downstream of the polyA signal, we inserted the FRT-flanked Keo selection marker for screening targeted ES cells. Sixth, additional restriction enzymes for Southern screening were introduced further downstream (EcoRI, EcoRV, KpnI, NsiI).

The sequences encoding the following: DsRed2-LoxP-piRNAsites-inLacZ-LoxP-SV40polyA, were chemically synthesized (GeneArt) and inserted into the NurI-homology regions of a targeting construct for the Rosa26 locus. The FRT-flanked Keo reporter was cloned in later to complete the targeting construct. Given below is the final cassette inserted into the Rosa26 locus of the mouse genome. It consists of the DsRed2 coding sequence, the 3’ UTR based on the LacZ backbone with binding sites for 35 MILI-bound piRNAs, flanking LoxP sites, SV40 polyA signal, FRT-flanked Keo reporter cassette (with its own polyA signal).

```
DsRed2
LoxP
piRNA sites
LoxP
SV40 polyA:
FRT-Keo-RED
```

A9 ES cells were electroporated and 300 clones picked. After screening by Southern blotting (EcoRV was chosen for the 3’ arm and AfIII for the 5’ arm) for correct targeting of the Rosa26 locus, only clones scoring positive with both probes were retained for further study (Figure S1B). DNA ladders used for Southern probing are 1kb Extension Ladder (Invitrogen; cat. no. 10511-012) or 1kb ladder (NEB), as indicated in the figure. The clone (#1F1B) was retained for injection into C57BL/6N host embryos at the 8-cell stage, after which seven 100% agouti males were obtained. Backcrosses with C57BL/6j Rj (Janvier labs) wildtype females were performed to transfer the targeted allele to the C57BL/6 background to obtain heterozygous Rosa26-pi mice. Heterozygous males and female siblings were intercrossed to obtain Rosa26-pi homozygous animals. Note that the FRT-flanked Keo selection
cassette was not removed in the animals used for this study. However, the presence of the upstream polyA site from SV40 is supposed to generate a transcript that has the DsRed coding sequence and a 3’UTR with the piRNA complementary sequences (flanked by LoxP sites).

**Southern probe sequences**

5’ probe (1224 bp)

3’ probe (1028 bp)

**PCR primers used for genotyping Rosa26-pi mice:**

TCTCCCAAGATCGTCTGGAG  
AAGACCCGAAGAGTTTGTGTC  
DsRed_forward  
DsRed_reverse

Reaction volume is 25 µl [1X Taq buffer, 0.2 mM dNTP, 2 mM MgCl₂, 1.0 µl primers (primer mix stock is 10 mM each), 1-1.5 µl template genomic DNA and 0.5 µl Taq DNA polymerase].

**PCR program:**

1. 94°C for 3 minutes  
2. 94°C for 20 seconds  
3. 65°C for 30 seconds  
4. 72°C for 30 seconds  
5. Repeat step 2-4 for 35 cycles  
6. 72°C for 5 minutes  
Load 15 µl reaction in 2% Agarose gel (1X TAE)

**Sequenced PCR products from Rosa26-pi knock-in allele (311nt)**

**Histology and Immunofluorescence of mouse testes sections**

**Histology of mouse testes sections**

To prepare the paraffin sections, the mouse testes were washed in PBS, and fixed in 4% paraformaldehyde overnight at 4°C. After washing in PBS, testes were dehydrated in 70% ethanol (90 min minimum), testes tissue can be stored in 70% ethanol for 1-2 months at 4°C. Testes were then further dehydrated in 80%, 90%, 96% and 100% ethanol (90 min for each step), followed by incubation in xylene (30 min x 3). Xylene was removed and replaced with paraffin, and incubated at 56-58°C. Testes were then transferred into plastic molds (polysciences, mold S-22) filled with paraffin, and paraffin was allowed to become solid at room temperature. The testes sections (~7 µM thickness) were prepared using microtome and mounted on the Superfrost Plus slides with 10% ethanol. The sections were allowed to stretch at 42°C. For histology, the slides containing the paraffin sections were placed in a glass slide holder filled with xylene (5 min x2) to become deparaffinized. For rehydration, the slides were incubated in 100% ethanol, 96% ethanol, 70% ethanol, 50% ethanol and water (2 min for each step). Sections were stained with Hematoxylin solution, Harris modified (Sigma, cat. No. HHS16) for 1 min and rinsed in water. To destain faster, sections can be incubated in Acidic alcohol (1% HCl in 70% ethanol) for 10-30 sec and rinsed with water. Sections were stained with Eosin Y solution with phloxine (Sigma, cat. No. HT110332) for 1 min and washed with water. For
dehydration, the sections were incubated in 96% (30 sec), 100% ethanol (2 min) and xylene (5 min x 2). Permount (Fisher Scientific, cat. No. SP15-100) was dropped on the section and covered with coverslips. The sections were examined and pictures were taken using Zeiss Axio Imager Z1 fluorescence microscope.

**Immunofluorescence of mouse testes sections**
To prepare the cryosections, the mouse testes were washed in PBS, and then fixed in 2% paraformaldehyde for 3 hr at 4°C. Testes samples were washed in PBS and then dehydrated by soaking in 15% sucrose at 4°C till the testes sink to the bottom of the tube. Testes were further dehydrated by soaking in 30% sucrose till the sample sink to the bottom and followed by incubation in OCT compound (Fisher Scientific, cat. No. 14-373-65) for 1 hr at 4°C. Testes were then transferred into cryomold filled with OCT compound and was slowly frozen by placement on dry ice. The sections (~7 µM thickness) were prepared using microtome. For immunostaining of the testes sections, the slides with sections were dried at room temperature, and fixed in cold 4% paraformaldehyde for 10 min. After washing with PBS and distilled water, the sections were soaked in 10 mM Sodium Citrate buffer (pH 6.0) and boiled using microwave oven for 20 min. The Sodium Citrate buffer was cooled down to room temperature and permeabilized with 0.3% Triton X-100 for 100 min. After washing with TBS-T (25 mM Tris-HCl pH 7.5, 0.14 M NaCl, 0.027 mM KCl and 0.1% Tween-20), the sections were then pre-blocked in blocking solution (5% goat serum in TBS-T) for 30 min. Proteins were detected by incubating with primary antibodies to specific proteins for overnight at 4°C in humidified chamber. The secondary antibodies coupled to fluorescent dye were used and incubated for another 1 hr. DAPI (0.5 µg/ml, Bio-Rad, cat. No. 10043282) were used for nuclei detection. The sections were mounted with a drop of Slowfade Gold Antifade Reagent (Life technology, cat. No. S36942). Pictures were taken using Leica TCS SP2 AOBS, inverted confocal microscope.

**Mouse Piwi immunoprecipitations, 5′-end labelling of RNAs and Northern analysis**

**Mouse Piwi Immunoprecipitation**
Mouse Piwi antibodies were incubated with approximately 15 µL protein G-Sepharose beads (GE Healthcare, cat. No. 17-0618-01) for 3 hrs and followed by washing (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.05% NP-40) to remove the unbound antibody. The mouse testes was homogenized in a glass tissue homogenizer by douncing in lysis buffer [50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM MgCl2, 1 mM DTT, 0.5% sodium deoxycholate (Sigma, cat. No. 30968), 1% Triton X-100, protease inhibitor (Roche)] and spun down for 15 min at 4°C. The bead was incubated with cleared testes lysate supernatant for 3 hr and washed five times (10 mM Tris-HCl pH8.0, 150 mM NaCl, 0.05% NP-40). The immunoprecipitated complex was further subjected to western blot or small RNA libraries preparation.

**Northern blot analysis**
Total RNA was extracted from testes by using TriZol reagent (Life technology, cat. No. 15596026) according to the manufacture’s guide. Approximately 10 µg RNA was resolved on 1% formaldehyde denaturing agarose gel and transferred onto positively-charged Hybond-N+ Nylon membrane (GE Healthcare, cat. No. RPN203B) through capillary transfer for 18 hr in 20 X SSC solution (3 M NaCl, 300 mM Sodium Citrate). RNA was crosslinked to the membrane by UV-crosslinker (Stratalinker, 120 µJ, 254 nm). The membrane was pre-hybridized with pre-warmed Church buffer (65°C, 0.25 M Sodium Phosphate pH 7.2, 1 mM EDTA, 1% BSA, 7% SDS) by incubating for 1 hr in a rotating hybridization oven. Denatured radiolabeled probe (95°C, 1 min, cool down on ice) was added into the hybridization bottle filled with new pre-warmed Church buffer, and incubate overnight at 65°C in the hybridization oven. The membrane was washed twice with low-stringency wash buffer (2 x SSC containing 0.1% SDS, 10 min for each wash), and further washed once with high-stringency wash buffer (0.1 x SSC containing 0.1% SDS) for 5 min. The membrane was exposed to Phosphor Storage screens (GE Health) and scanned (Typhoon scanner; GE Health). LINE1 DNA fragment (515 bp- 1680 bp) and IAP was selected as probe. DNA probe was labeled using Random Primed DNA labeling kit (Roche, cat. No. 11004760001) with dCTP[α-32P] (Perkin Elmer, cat. No. NEG013H001MC), and was further purified with microspin G-25 column (GE Healthcare, cat. No. 27-5325-01).
Small RNA libraries and bioinformatics

Small RNA libraries

RNAs present in endogenous MILI or MIWI2 complexes from testes of new-born pups (P0) were isolated. Briefly, immunoprecipitations were treated with Proteinase K in 300 µL reaction at 42°C for 15 min (10 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5% SDS). RNAs present in the sample were purified by phenol-chloroform extraction and precipitation with ethanol. Approximately 20% of the sample was radioactively labeled at the 5’ end and mixed with the cold RNA. This mixture of radiolabeled and cold RNA was resolved by 15% urea-PAGE, and the RNAs present in the region (~24-30 nt) was excised and placed in nuclease-free tube. The gel was crushed in the tube and 400 µl of 0.4 M NaCl was added to it. The RNA was eluted from the gel overnight at 25°C in a thermomixer. Gel pieces were removed by filtering the RNA containing gel with SpinX column (Costar, cat. No. 8160). RNA was directly precipitated with glycogen (10-20 µg) and absolute ethanol (1.0 ml) at -20°C overnight. Next day, RNA samples were centrifuged at 16000 x g for 30 minutes. Pellet was washed once with 75% ethanol and finally dissolved in 6 µl RNase free water. RNA was directly used for library preparation. When total small RNA libraries were required, total RNA from testes of new-born pups (P0) was isolated with TRI-reagent (Life Tech, AM9738). RNAs were resolved by 15% urea-PAGE, and RNAs present in the region (~15-40 nt) was gel-eluted as mentioned above. After purification they were directly used for library preparation. Libraries were prepared (barcoded at 3’ end) using NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (NEB Catalogue No. E7300) following manufacture instructions. Multiple libraries with different barcodes (at 3’ end) were mixed in equimolar ratio and sequenced with the Illumina HiSeq 2000 platform (EMBL GeneCore facility) for 50 cycles. Deep sequencing datasets generated in this study are deposited with Gene Expression Omnibus (GEO) under the accession number (GSE74423).

Small RNA bioinformatics of Rosa26-pi reporter knock-in mouse

Reads were sorted into individual libraries based on the barcodes and the 3’ adapter sequences were clipped from the reads using cutadapt (DOI:http://dx.doi.org/10.14806/ej.17.1.200). Reads of at least 15 nucleotides were then aligned to the DsRed2-LoxP-piRNAsites-in-LacZ-LoxP-SV40pA reporter sequence using bowtie (Langmead et al., 2009) allowing no mismatches. Only the reads of 25nt-31nt, which mapped to unique reporter site, were used for subsequent analysis.

The 5’ ends of piRNAs were mapped and plotted along the reporter (for each nucleotide we plotted the number of piRNAs starting at that nucleotide normalized to millions of library reads (rpm) if their abundance exceeded 1 rpm) in logarithmic scale. Separate panels show the piRNAs targeting the reporter (Figure S1C) and those produced from the reporter (Figure 1A).

To follow the pattern of piRNA production in respect to the target sites, we aggregated the counts of produced piRNAs from all 35 MILI targeted sites and their downstream sequences and plotted the overall number of piRNAs (rpm) having 5’ or 3’ end at specific distance from the 5’ end of targeting piRNA (Figure 1B, Figure S1D). Produced piRNAs with 5’ end distance equal to 0 starts immediately downstream from 5’ end of targeting piRNA, the piRNAs with distance -10 share 10 nucleotide 5’ end overlap with the targeting piRNA. The y-axis showing the piRNA abundance was scaled to show both low and highly abundant piRNAs (Figure 1B). The distribution of the ends of produced piRNAs demonstrated that besides the dominant secondary piRNAs with 5’ end created by MILI slicing also additional piRNAs are produced downstream of the 3’ end of secondary piRNAs. These additional piRNAs which we call “inchworm” piRNAs associate with both MILI and MIWI2.

To better characterize the distance between 3’ end of secondary and 5’ end of inchworm piRNAs we plotted the fraction of produced piRNA pairs having specific 3’ to 5’ end distance Δ (Figure 1E). The score for specific distance Δ was calculated as: \( score(\Delta) = \sum \min(M(i), N(i+\Delta)) \) where \( M(i) \) is the count of produced secondary piRNAs (in rpm) with 3’ end on the plus strand at a particular position \( i \) and \( N(i+\Delta) \) is the count of piRNAs on the same strand which have their 5’ end position at \( i + \Delta \). Similarly, we calculated the scores for the distances of the 5’ ends (Figure 1F) and for 3’ to 5’ distances of piRNAs produced upstream from the first cleavage site (Figure 6D).
To visualize which individual piRNA sequences are produced we plotted the sequence level detail of the region encompassing 2nd and 3rd site of the reporter. piRNAs produced in an abundance of at least 1 rpm were shown (Figure 1E).

To investigate the nucleotide composition of produced piRNAs, we aligned their 5′ and 3′ ends and calculated the proportion of individual nucleotides in the first and last 10 nucleotides (Figure S1E). Strong preference for U was observed for the starting nucleotide of MIWI2 secondary piRNAs and both MIL1 and MIWI2 inchworm piRNAs. Nucleotide frequency was plotted also for the reporter (DNA) nucleotide immediately downstream of the 3′ piRNA end.

To visualize the individual piRNA production along the reporter, we plotted the individual piRNAs produced from the reporter as well as reporter targeting piRNAs in color based on the presence of 1U and 10A nucleotide which are hallmarks of primary and secondary piRNAs, respectively (Figure 6C, Figure S1F).

**Mouse small RNA bioinformatics of Exd1 mouse mutant**

Reads were sorted into individual libraries based on the barcodes, the 3′ adapter sequences were removed and mapped to the mouse genome (mm9). The software used for processing the data (genomic coordinates etc) from the raw data files are in-house tools developed by the Sachidanandam lab (Olson et al., 2008). Only reads perfectly matching the genome were kept for further analysis.

The reads from total small RNA libraries were divided into two groups based on whether they match to single or multiple locations in the genome. Read length distribution between Exd1+/− and Exd1+/− showed apparent enrichment of uniquely mapped piRNAs of length 28-32nt and lack of multi-mapped piRNAs of 24-30nt length (Figure 5C, S5D) in Exd1+/−. The enriched piRNAs started predominantly with 5′ U, mapped to unannotated parts of the genome and were not reliably detected in MIL1 or MIWI2 IPs (data not shown) giving us no clue about their nature. Therefore, we focused only on the reads mapped to multiple locations (24-30 nt) showing ~80% decrease in abundance in Exd1+/− (Figure 5D). Classification of reads based on genome annotation showed that mainly repeat antisense reads are missing. To find out whether specifically MIL1 or MIWI2 piRNAs were affected we classified the reads from total small RNA libraries into groups based on their presence in MIL1 and MIWI2 IP libraries. To consider the read as MIL1-specific we required its presence in both MIL1 IP libraries and its abundance to be at least 3 times higher than in MIWI2 IP libraries. The same criterion was applied for obtaining MIWI2-specific reads. If the reads were present in all IP libraries without the difference in abundance between MIL1 and MIWI2, we considered the read as the piRNA which associates with both MIL1 and MIWI2. The rest of the reads were annotated as not reliably associated with MIL1 or MIWI2. We found that MIWI2-specific reads were underrepresented in Exd1+/−. Annotation of reads to individual repeats showed that mainly LINE L1 repeat antisense reads are affected (Figure 5E) and that the reads having A at 10th nucleotide (potential secondary piRNAs) are depleted among LINE L1 piRNAs (24nt-30nt) in Exd1+/− (Figure 5L).

To visualize the difference in LINE1 piRNA abundance between Exd1+/− and Exd1+/− we mapped the 5′ ends of genome mapped 24-32 nt piRNAs to L1 repeat consensus sequence allowing at maximum of 3 mismatches (Figure 5K). Read counts were normalized to library sizes and divided by counts of mapped genomic sites. These reads were also used for ping-pong analysis (Figure 5M, S5H) where the product of the piRNA counts was used to calculate the score for the 5′ end distance Δ: score(Δ)=∑Mi(i)*N(i+Δ) where M(i) is the count of produced piRNAs (in rpm) with 5′ end on the plus strand at a particular position i and N(i+Δ) is the count of piRNAs which have their 5′ end position at minus strand at i + Δ. The distance equal to 0 refers to a situation where piRNAs share the 5′ end nucleotide and the distance 9 corresponds to 10nt overlap of piRNA 5′ ends.

**SEQUENCES OF PROTEINS INVESTIGATED IN THIS STUDY**

**Siwi (1-899 aa)**
**Ago3 (1-900 aa)**

**Bombyx mori Vreteno (1-1048 aa)**

**Bombyx mori Papi (1-545 aa)**

**Bombyx mori Vasa (1-601 aa)**

**Bombyx mori Spindle-E (1-1362 aa)**

**Bombyx mori Exd1 (1-315 aa)**

**Mus musculus Exd1 (1-570 aa)**

**Drosophila melanogaster Exd1 (CG11263) (1-265 aa)**

**SUPPLEMENTAL REFERENCES**

Bieniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). MultiBac: expanding the research toolbox for multiprotein complexes. Trends in biochemical sciences 37, 49-57.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta crystallographica Section D, Biological crystallography 66, 12-21.
Cowtan, K. (2006). The Buccaneer software for automated model building. Acta crystallographica Section D, Biological crystallography 62, 1002-1011.
de La Fortelle, E., Bricogne, Gerard. (1997). Heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In Methods in Enzymology, C.W. Carter Jr, Sweet, R. M., ed. (New York: Academic Press, Inc.), pp. 472-494.
de Sanctis, D., Beteva, A., Caserotto, H., Dobias, F., Gabadinho, J., Giraud, T., Gobbo, A., Guijarro, M., Lentini, M., Lavault, B., et al. (2012). ID29: a high-intensity highly automated ESRF beamline for macromolecular crystallography experiments exploiting anomalous scattering. Journal of synchrotron radiation 19, 455-461.
Di Giacomo, M., Comazzetto, S., Saini, H., De Fazio, S., Carrieri, C., Morgan, M., Vasiliauskaite, L., Benes, V., Enright, A.J., and O’Carroll, D. (2013). Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. Molecular cell 50, 601-608.
Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., et al. (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448, 151-156.
Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta crystallographica Section D, Biological crystallography 66, 486-501.
Evans, P. (2006). Scaling and assessment of data quality. Acta crystallographica Section D, Biological crystallography 62, 72-82.
Handler, D., Olivieri, D., Novatchkova, M., Gruber, F.S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and Brennecke, J. (2011). A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. The EMBO journal 30, 3977-3993.
Kabsch, W. (2010). Xds. Acta crystallographica Section D, Biological crystallography 66, 125-132.
Kawaoka, S., Hayashi, N., Suzuki, Y., Abe, H., Sugano, S., Tomari, Y., Shimada, T., and Katsuma, S. (2009). The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. Rna 15, 1258-1264.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10, R25.
McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser Crystallographic Software. J Appl Cryst 40, 658-674.
Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMACS for the refinement of macromolecular crystal structures. Acta crystallographica Section D, Biological crystallography 67, 355-367.
Olivieri, D., Sykora, M.M., Sachidanandam, R., Mechtler, K., and Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. The EMBO journal 29, 3301-3317.
Pandey, R.R., Tokuzawa, Y., Yang, Z., Hayashi, E., Ichisaka, T., Kajita, S., Asano, Y., Kunieda, T., Sachidanandam, R., Chuma, S., et al. (2013). Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice. Proceedings of the National Academy of Sciences of the United States of America 110, 16492-16497.
Reuter, M., Chuma, S., Tanaka, T., Franz, T., Stark, A., and Pillai, R.S. (2009). Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. Nature structural & molecular biology 16, 639-646.
Sarot, E., Payen-Groschen, G., Bucheton, A., and Pelisson, A. (2004). Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. Genetics 166, 1313-1321.
Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta crystallographica Section D, Biological crystallography 66, 479-485.
Venken, K.J., Schulze, K.L., Haelterman, N.A., Pan, H., He, Y., Evans-Holm, M., Carlson, J.W., Levis, R.W., Spradling, A.C., Hoskins, R.A., et al. (2011). MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nat Methods 8, 737-743.
Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z., Berninger, P., Palencia, A., et al. (2012). A role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing. Molecular cell 47, 970-979.
Xiol, J., Spinelli, P., Laussmann, M.A., Homolka, D., Yang, Z., Cora, E., Coute, Y., Conn, S., Kadlec, J., Sachidanandam, R., et al. (2014). RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. Cell 157, 1698-1711.