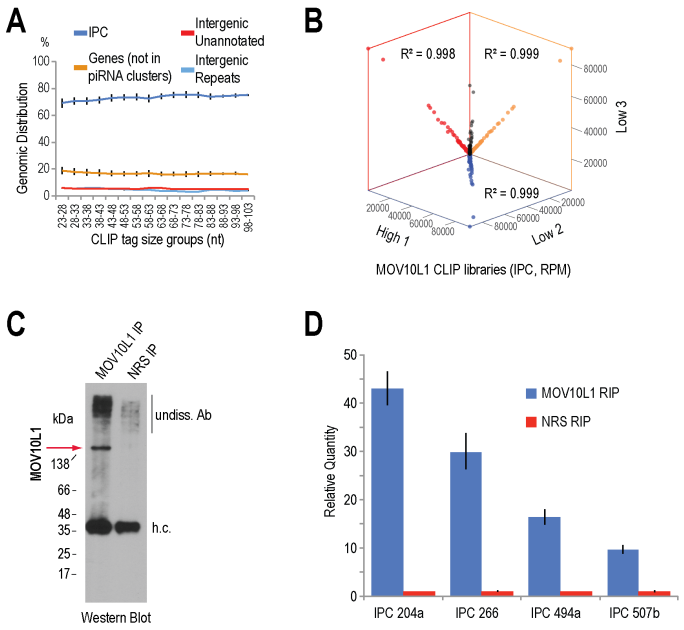
MOV10L1 RNA helicase binds piRNA precursors to initiate piRNA processing

Anastassios Vourekas, Ke Zheng, Qi Fu, Manolis Maragkakis, Panagiotis Alexiou, Jing Ma, Ramesh S. Pillai, Zissimos Mourelatos and P. Jeremy Wang

SUPPLEMENTAL INFORMATION

- Five Supplemental Figures with Legends
- Supplemental Text
- Three Supplementary Tables with Legends
- Supplemental Materials and Methods
- Supplemental References

Supplemental Fig. S1.



Supplemental Fig. S1. Genomic distribution and reproducibility of MOV10L1 CLIP libraries, and specificity of MOV10L1 binding of piRNA precursor clusters.

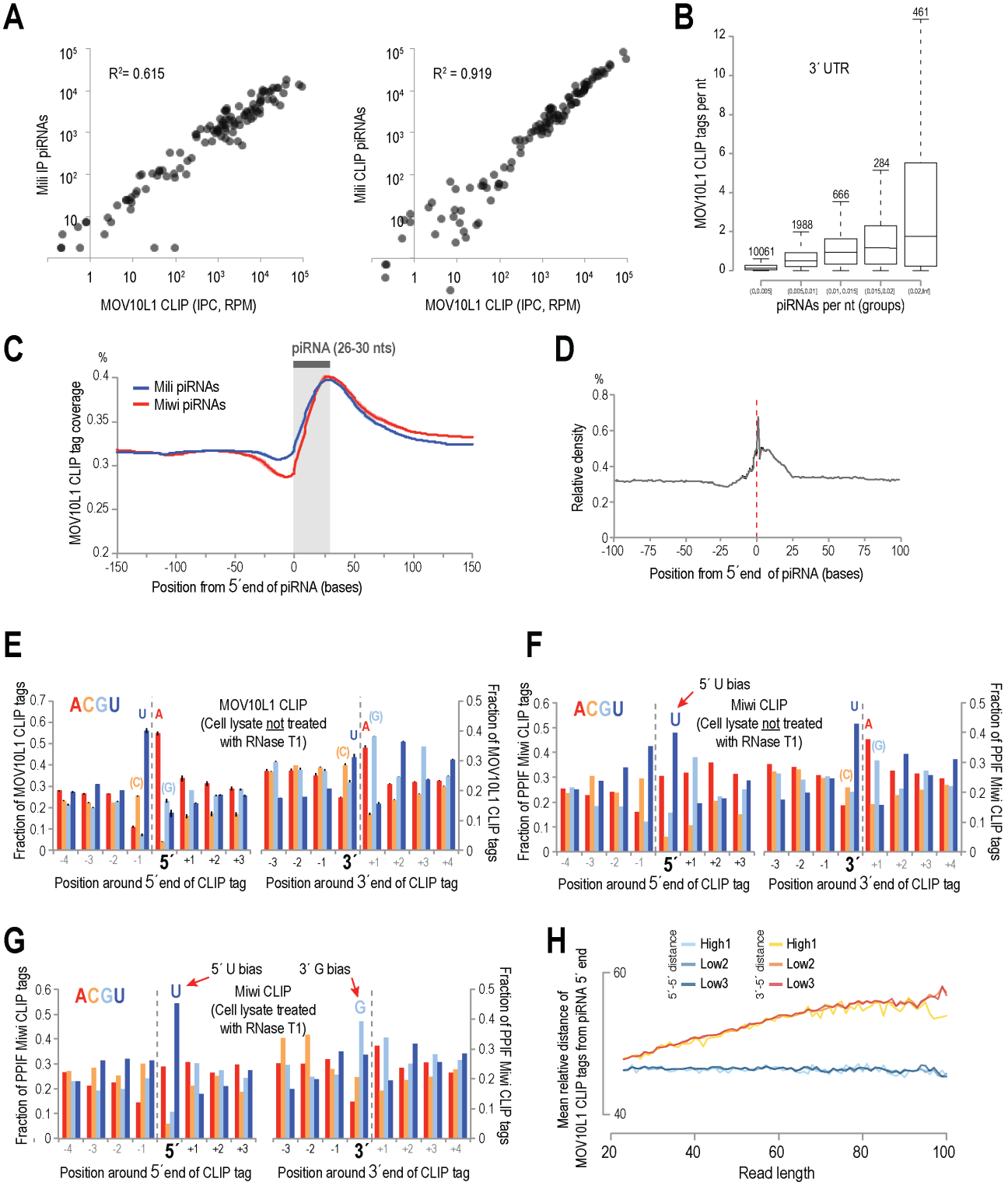
A. Genomic distribution of MOV10L1 CLIP tags with CLIP tag size, average of three libraries. The percentage of tags mapping within the indicated genomic categories is plotted. Mapped tags are separated in 16 size bins. Error bars represent one SD (n=3).

B. Three-way scatter plot of three MOV10L1 CLIP libraries, using RPM values of intergenic piRNA clusters (IPC), which comprise >70% of the CLIP tags (**Supplemental Table S1**). Pearson correlation (R^2) is shown for pairwise comparisons.

C. Western blot detection of MOV10L1 in immunoprecipitated samples from WT testes with anti-MOV10L1 antibodies and non-immune rabbit serum (NRS) respectively. H.c.: heavy antibody chain; undiss. Ab: undissociated antibody.

D. Quantitation of individual piRNA precursor transcripts in MOV10L1 IP and control IP samples. After spiking with equal amounts of *in vitro* transcribed Renilla luciferase mRNA, RNA was extracted from MOV10L1 and NRS RNA IP (RIP) beads using Trizol and was treated with DNase I. Pachytene piRNA precursor transcripts were detected by qRT-PCR using Renilla luciferase as a reference.

Supplemental Fig. S2.



Supplemental Fig. S2. Characteristics and comparative analysis of MOV10L1 and Piwi protein CLIP tags.

A. Scatter plot using RPM values of intergenic piRNA clusters between the average of MOV10L1 CLIP libraries, and Mili IP and Mili CLIP piRNA libraries (**Supplemental Table S1**). Pearson correlation (R^2) is shown for every comparison.

B. Box plot showing per nucleotide (nt) abundance of MOV10L1 CLIP tags within mRNA 3′ UTRs (y-axis) with relation to piRNA abundance on the same locations (sorted in five groups as shown, x-axis). The number of genes in every group is marked above each box. Note the strong positive correlation of the two abundances, signifying strong binding by MOV10L1 of the mRNA 3′ UTRs that are processed into piRNAs.

C. MOV10L1 CLIP tag coverage relative to Mili and Miwi piRNA 5′ ends. The position occupied by a typical piRNA molecule is shaded. Shaded areas represent one SD (n=3).

D. Density plot of 5′ end to 5′ end distances of MOV10L1 CLIP tags from piRNAs. The majority of MOV10L1 CLIP tag 5′ ends are positioned downstream of piRNA 5′ ends, within a distance of up to 25 nucleotides. A peak at position +1 is likely the result of the artifactual coincidence of the nucleotide bias (U) one nucleotide upstream of the 5′ of the MOV10L1 CLIP tags (generated by an endogenous nuclease during the IP process, see **Supplemental Text** and panels **S2E-H**), in combination with the piRNA 5′ end preference (also U). Error bars represent one SD (n=3)

E. Nucleotide preference around the 5′ and 3′ ends of MOV10L1 CLIP tags. The cell lysates used for the IP step were not treated with exogenous nucleases. A dashed line separates upstream (-) and downstream (+) positions for the 5′ and 3′ ends, and it also practically represents the cleavage site that generates the ends of the crosslinked RNA fragment. The U(C)-A(G) bias around the ends is most likely generated by an endogenous nuclease unrelated to piRNA biogenesis. Error bars represent one SD (n=3).

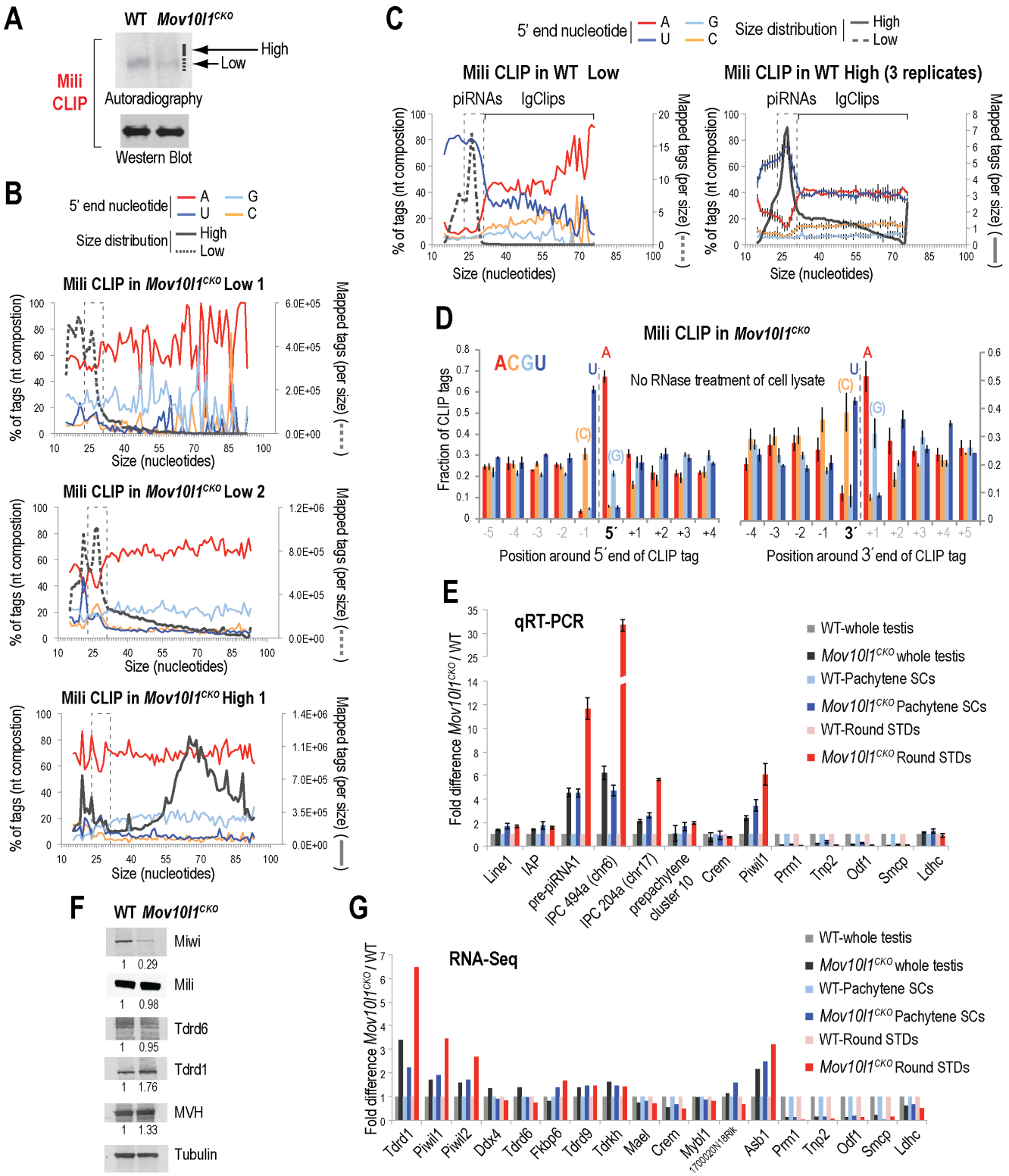
F. Nucleotide preference around the 5′ and 3′ ends of Miwi lgClip tags from one CLIP library generated using cell lysate not treated with exogenous RNase. Miwi lgClip tags that are mapped within intergenic piRNA clusters (IPCs) are plotted. These CLIP tags represent intermediate piRNA-processing fragments and bear a distinct 5′-U preference due to stabilization of their 5′ ends by Miwi. In contrast, the same 3′-end bias of the Miwi-bound intermediate fragment, U(C)-A(G), as

in MOV10L1 CLIP tags (panel E), indicates that both Miwi and MOV10L1 CLIP tags are exposed to the same endogenous nucleolytic activity. This finding is in agreement with the piRNA biogenesis model, where the 3′ ends of the intermediate fragments are accessible *in vivo* for (exo)nucleolytic processing.

G. Nucleotide preference around the 5′ and 3′ ends of Miwi lgClip tags mapping to IPCs. This CLIP-seq library was prepared from a lysate mildly treated with exogenously provided RNase T1. The 5′ ends are protected by Miwi binding, and therefore inaccessible to the nuclease; thus the *in vivo* 5′ U bias is maintained. The 3′ ends, being accessible, have acquired the bias of RNase T1 cleavage (Guanosine).

H. Mean relative distance of 5′ ends and 3′ ends of MOV10L1 CLIP tags from 5′ ends of piRNAs. The relative 5′-5′ distance remains constant among CLIP tags of varying lengths. In contrast, 3′-5′ distance increases with increasing MOV10L1 CLIP tag length.

Supplemental Fig. S3.



Supplemental Fig. S3. Analysis of *Mov10l1*^{CKO} Mili CLIP and RNA-seq libraries.

A. Mili CLIP using testes from WT or *Mov10l1*$^{fl/-}$ Neurog3-Cre (*Mov10l1*^{CKO}) mice, in which *Mov10l1* is disrupted after postnatal day 7 (Zheng and Wang, 2012). The area of the nitrocellulose membrane that was excised to extract Mili bound RNA for CLIP library preparation is marked with a dashed line (main signal), and a solid line (higher complexes).

B. Size distribution (absolute CLIP tag counts per size) and 5′ nucleotide composition (percent of mapped tags carrying each of the four nucleotides, per size) for three Mili CLIP libraries from *Mov10l1*^{CKO} mice. Note the absence of a distinct 23-31 nucleotide-sized population with a 5′ U bias (marked with a dashed box), consistent with lack of pachytene piRNAs. A peak at 26 nt (which however does not show a 5′ U bias, and therefore does not represent bona fide piRNAs) could be attributed to random cleavage of crosslinked RNA to such extent that the protein-protected footprint on the RNA is 26 nt. Upper panel: library (Low 1) generated by extracting RNA from the main radioactive signal (panel **S3A**, dashed line). Middle panel: replicate library of Low 1 (Low 2). Bottom panel: library prepared using higher molecular weight RNPs (panel **S3A**, solid line).

C. Size distribution and 5′ nucleotide composition for one representative Mili piRNA enriched library and the average of three replicate Mili CLIP libraries enriched in large CLIP tags from WT mice (this work and (Vourekas et al., 2012)). Note the presence of a distinct 23-31 nucleotide-sized population (marked with a dashed box) with an overwhelming 5′ U bias, which is characteristic of pachytene piRNAs. Left panel: library (Low 1) generated by extracting RNA from the main radioactive signal (Vourekas et al., 2012). Right panel: three replicate libraries (High 2, High 5, High 6) prepared using higher molecular weight RNPs (this work and (Vourekas et al., 2012)). Error bars represent one SD (n=3).

D. Nucleotide composition around the 5′ and 3′ ends of Mili large CLIP tags mapping within IPCs (3 libraries generated from *Mov10l1*^{CKO} mice). A dashed line separates upstream and downstream positions, and thus designates the cleavage site that gives rise to the ends of the crosslinked RNA fragment. Error bars represent one SD (n=3).

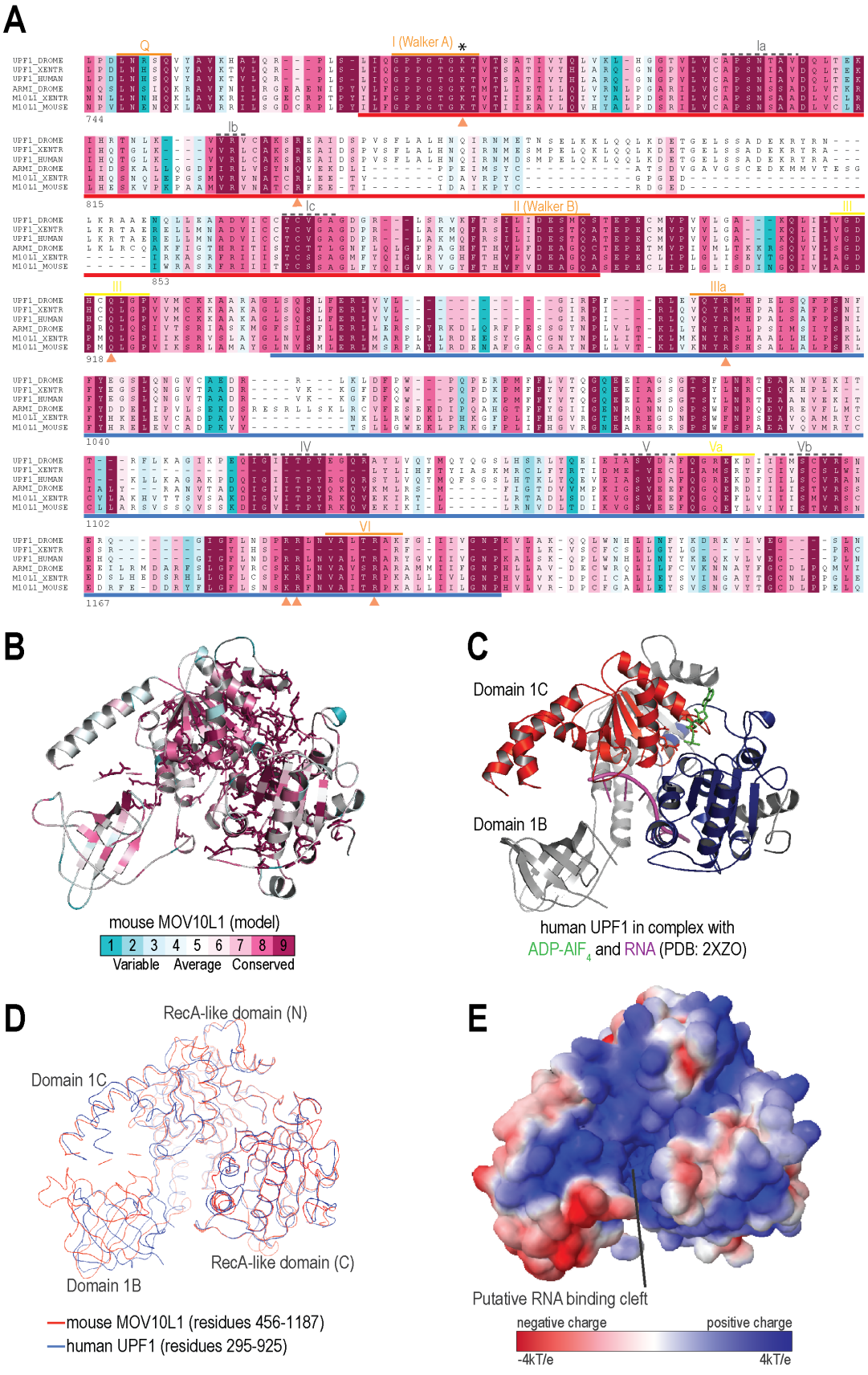
E. Quantitative qRT-PCR analysis of differential expression of various transcripts between *Mov10l1*^{CKO} and WT. RNAs were prepared from testis, isolated pachytene spermatocytes (SCs), and round spermatids (STDs). LINE1, IAP, IPC represent retrotransposon consensus sequences; Chr6 and Chr17 are long piRNA precursor transcripts; transcripts were detected as previously

(Vourekas et al., 2012). Pre-piRNA1 and prepachytene cluster 10 were detected as previously (Zheng and Wang, 2012). Error bars represent SE (n=3).

F. Western blot analysis of proteins implicated in piRNA biogenesis in *Mov10l1*^{CKO} and WT testes. Tubulin serves for normalization.

G. Differential expression of selected transcripts between *Mov10l1*^{CKO} and WT. RNA-seq libraries were prepared from testes, isolated pachytene spermatocytes, and round spermatids. *Asb1* and *1700020N18Rik* are examples of a coding mRNA and a spliced non-coding RNA respectively, which are processed into piRNAs.

Supplemental Fig. S4.



Supplemental Fig. S4. Primary and tertiary structure analyses reveal similarity between MOV10L1 and UPF1 helicases.

A. Multiple sequence alignment of UPF1 and MOV10L1 homologues. 14 UPF1 and 13 MOV10L1 homologues were aligned. The helicase core sequences of only six are shown: UPF1 homologues from human, *Xenopus tropicalis* (XENTR) and *D. melanogaster* (DROME); MOV10L1 (M10L1) homologues from mouse, *X. tropicalis*, and *D. melanogaster* (ARMI). The red and blue lines mark the two helicase domains. The conserved sequence motifs are marked with a color-coded line for ATP binding (Walker A) and ATP hydrolysis (Walker B) (orange solid line), coordination of ATP hydrolysis with RNA binding (yellow solid line), and RNA binding (grey dashed line) (Fairman-Williams et al., 2010). Amino acid residues with experimentally verified role in ATP binding and hydrolysis are marked with orange triangles(Cheng et al., 2007; Weng et al., 1996). An asterisk marks the K778 residue that was mutated to an Alanine to generate the MOV10L1 knockin mouse. Residues are highlighted with a color scheme according to conservation (see also panel **S4B**).

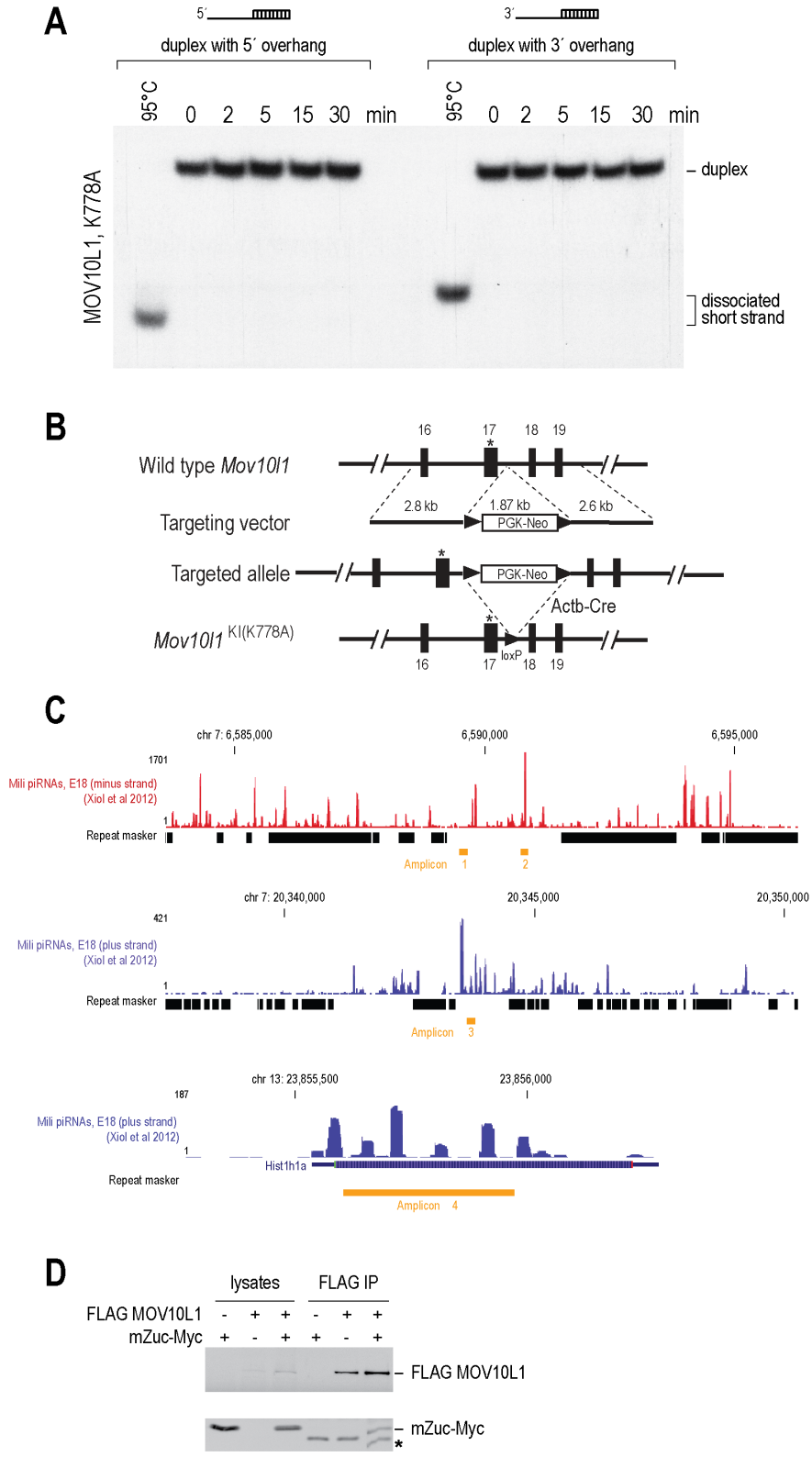
B. Modeling of MOV10L1 structure by Phyre2 algorithm. The degree of amino acid residue conservation was determined by Consurf and was based on a multiple alignment of 14 UPF1 and 13 MOV10L1 homologues generated with Clustal Omega. Residues with the highest conservation score are depicted as sticks. Protein structures in **S4B**, **C** were generated by using PyMol (www.pymol.org).

C. Crystal structure of human UPF1 bound to RNA (PDB: 2XZO)(Chakrabarti et al., 2011). Helicase domains are shown in red and blue, as in (**S4B**).

D. Line representation model of superimposed amino acid chains of MOV10L1 (red) and UPF1 (blue). Important domains are indicated. RMSD: 1.183 Å. Image was generated using PyMol.

E. Space-filling model of MOV10L1 with electrostatic surface. The MOV10L1 structure is depicted at the same angle as in panel **S4B**. The solvent-accessible surface model of MOV10L1 was generated using the Adaptive Poisson–Boltzmann Solver (ABPS) on PDB2PQR Server (http://nbcr-222.ucsd.edu/pdb2pqr_1.8/)(Dolinsky et al., 2004) and visualized by Jmol (http://jmol.sourceforge.net).

Supplemental Fig. S5.



Supplemental Fig. S5. The MOV10L1 K778A mutation is inactive in vitro and shows a piRNA biogenesis defect in vivo.

A. *In vitro* RNA duplex unwinding assay using the immunopurified, K778A MOV10L1 point mutant.

B. Generation of the *Mov10l1* knockin (K778A) allele. The neo selection cassette is flanked with loxP sites and is removed by *Actb*-Cre.

C. Genome browser snapshots (mm9), showing prenatal (embryonic day 18 – E18) Mili bound piRNAs (Xiol et al., 2012), Repeat masker track, and four amplicons detected by qRT-PCR using E16.5 total RNA from heterozygote and homozygote MOV10L1 point mutant KI mice (related to **Fig. 5I**)

D. MOV10L1 co-immunoprecipitates with mZuc in a heterologous system. We expressed FLAG-tagged MOV10L1 and Myc-tagged mZuc in 293T in the combinations shown. FLAG IP was performed from 293T cell lysates, and protein eluates from IP beads were analyzed by SDS-PAGE and immunoblotted using anti-FLAG and anti-Myc antibodies. The asterisk denotes a non-specific band.

SUPPLEMENTAL TEXT

The ends of MOV10L1 CLIP tags are exposed to an endogenous nuclease that is active in the testis lysate during the IP step of CLIP.

MOV10L1 CLIP tags show A/(G) nucleotide bias at the 5′ end, and U/(C) bias one position upstream (which is not part of the CLIP tag); the 3′ end has a similar bias (**Supplemental Fig. S2E**). No process related to piRNA biogenesis is known to produce such nucleotide bias, and no exogenous nuclease was added during the preparation of these libraries. We compared these nucleotide biases with the respective biases of Miwi bound piRNA precursor CLIP tags (Vourekas et al., 2012). piRNA precursor intermediate fragments bound by Miwi (Vourekas et al., 2012) show a bias for U at the 5′ end, and in a library that was prepared without exogenous RNase treatment, the 3′ ends display the same U/(C)-A/(G) bias as the MOV10L1 CLIP tags (**Supplemental Fig. S2F**). Piwi proteins bind and stabilize the 5′ U of PPIFs and therefore protect their 5′ ends from nucleolytic attack *in vivo*, but also *in vitro* during the immunoprecipitation step of CLIP. On the other hand, the 3′ ends of Piwi bound PPIFs are available for exonucleolytic trimming. The presence of the U/(C)-A/(G) bias at the 3′ ends of Miwi CLIP tags suggests that after cell lysis, RNA crosslinked to Miwi is exposed to the same nucleolytic activity that created the 5′ and 3′ ends of the MOV10L1 CLIP tags. This explanation is further supported by the following observation: when crosslinked lysates are treated with exogenously provided RNase T1, Miwi loaded PPIFs retain a U bias at their 5′ end, but the 3′ end bias shifts towards G, reflecting the nucleotide preference of RNase T1 which cleaves RNAs after G (**Supplemental Fig. S2G**). Evidently, protected ends of bound RNAs retain the nucleotide biases generated by *in vivo* processes, whereas exposed ends can be further processed by exogenous and/or endogenous nucleases after cell lysis and during RNP purification, altering their characteristics. Since the crosslinked lysates were not treated with exogenous nucleases, the U/(C)-A/(G) bias is due to cleavage by an endogenous nuclease (unrelated to piRNA biogenesis), during the IP step of the CLIP experiment. We confirmed that crosslinked testis lysates exhibit enduring nuclease activity (our unpublished data). The sensitivity of the 5′ ends of MOV10L1 CLIP tags to nucleases indicates that they are not bound and protected by a Piwi protein. Further evidence that MOV10L1 binds the 5′ cleaved precursor is provided by the mean relative 5′-5′ distance between MOV10L1 CLIP tags and piRNAs which remains constant, indicating that the 5′ end of the piRNA precursor represented by the MOV10L1

CLIP tags had been defined before it was attacked by nucleases in the lysate, contrary to the 3′-5′ distance, which increases with the size of the CLIP tag (**Supplemental Fig. S2H**).

Structural Similarity of MOV10L1 with UPF1 RNA Helicase.

MOV10L1 contains two RecA-like helicase domains (Story and Steitz, 1992) that form a helicase core (**Fig. 5A**); the N-terminal domain harbors the conserved Walker A and B sequence motifs for ATP binding and hydrolysis (Walker et al., 1982). MOV10L1 is classified in the SF1 superfamily of DNA/RNA helicases (Fairman-Williams et al., 2010), and one of the closest homologues of MOV10L1 is UPF1, an RNA helicase involved in the nonsense mediated RNA decay (**Supplemental Fig. S4A**) (Kervestin and Jacobson, 2012). Intriguingly, nuclear UPF1 is also involved in telomere replication, and telomeres contain G-quadruplexes (Azzalin, 2012). We performed tertiary structure modeling using Phyre2 server (Kelley and Sternberg, 2009), and approximately two thirds of the MOV10L1 covering the C-terminal region were modeled with 100% confidence on the crystal structure of hUPF1, including domains similar to 1B and 1C that are specific for UPF1 (**Supplemental Fig. S4B, C, D**) (Chakrabarti et al., 2011; Saikrishnan et al., 2009). Analysis of the *Drosophila* MOV10L1 homolog Armitage produced similar results. All sequence motifs, and most residues previously identified as important for ATP binding, hydrolysis and coordination with the subsequent conformational changes that promote RNA binding and translocation (Chakrabarti et al., 2011; Cheng et al., 2007; Franks et al., 2010), are present and show remarkable degree of conservation between MOV10L1 and UPF1 (**Supplemental Fig. S4A, B**) (Fairman-Williams et al., 2010; Gorbalenya and Koonin, 1993; Iyer et al., 2004; Leipe et al., 2003; Weng et al., 1996). Residues found on the surface of the two RecA-like domains opposite to ATP binding, make contacts with the RNA phosphate backbone (Fairman-Williams et al., 2010; Saikrishnan et al., 2009). Mapping of the electrostatic potential on the surface of the modeled MOV10L1 reveals that the respective surface has a predicted net positive charge (**Supplemental Fig. S4E**), which is consistent with a conserved RNA binding role. To the degree that structural similarity supports similar molecular function, we postulated that MOV10L1 may utilize its conserved helicase core to translocate the piRNA precursor transcript, analogous to the translocation of UPF1 on mRNAs with a 5′ to 3′ directionality (Franks et al., 2010).

SUPPLEMENTAL TABLE LEGENDS

Supplementary Table 1

(A) Deep sequencing libraries – mapped reads; **(B)** *Intergenic piRNA cluster* coordinates; **(C)** CLIP tags mapping within intergenic piRNA clusters (abundance);

Supplementary Table 2

Numbers of putative piRNA sequences within deep sequencing libraries.

Supplementary Table 3

(A) Coordinates of predicted G-quadruplexes within mouse intergenic piRNA clusters (IPC), loose prediction algorithm; **(B)** Coordinates of G-quadruplexes within mouse intergenic piRNA clusters (IPC), strict prediction algorithm; **(C)** Analysis of piRNA density relative to G-quadruplexes (loose algorithm), within IPCs. **(D)** G-Quadruplex prediction in shuffled mouse piRNA clusters; **(E)** G-quadruplex prediction in shuffled human piRNA clusters; **(F)** G-quadruplex prediction in shuffled macaque piRNA clusters; **(G)** G-quadruplex prediction in shuffled rat piRNA clusters;

SUPPLEMENTAL MATERIALS AND METHODS

qRT-PCR.

Quantitative reverse transcription PCR (qRT-PCR) was performed using StepOnePlus™ (Life Technologies) and Prism (Applied Biosystems) Real-Time PCR Systems, Power SYBR Green reaction mix (Applied Biosystems 4367659) and primer sets (for pachytene piRNA clusters and other transcripts described in **Supplemental Fig. S3E**), as described before using the ddCt method (Vourekas et al., 2012; Zheng and Wang, 2012; Zheng et al., 2010). For qRT-PCR experiments shown in **Supplemental Fig. S1D** and **Figure 4** (detection of pachytene piRNA precursors in MOV10L1 and BG4 IPs), the following primers were used: Renilla luciferase, F: 5′-CGCTGAAAGTGTAGTAGATGTG-3′ and R: 5′-TCCACGAAGAAGTTATTCTCCA-3′; IPC 204a F: 5′-AGTCTGTGTAGTAGTTTCCTGAG-3′ and R: 5′-TGTCCACTTCCATGTTACCT-3′; IPC 266, F: 5′-GTTGCCCCAAGAGAATGTGT-3′ and R: 5′-TTCCACAGGTCCAGCCTTAG-3′; IPC 494a F: 5′-TGTTCACTTACATATCAGGGTC-3′ and R: 5′-GTAAAGCCCAAGAGCAAGAG-3′; IPC 507b, F: 5′-CTCCCAATGGCAACATCTTT-3′ and R: 5′-TGACCCTTCAGGGAATTCAG-3′. For qRT-PCR experiments shown in **Figure 5I** (detection of prenatal piRNA precursor transcripts in total RNA samples from heterozygote and homozygote point mutant MOV10L1 KI mice), we examined mapping locations of the embryonic day 18 Mili IP piRNAs identified by the Pillai group (Xiol et al., 2012) to identify areas of high prenatal piRNA density, that constitute the prenatal piRNA clusters. We designed three pairs of primers that amplify the following amplicons: Amplicon 1 chr7:6590715-6590867 (minus strand) forward primer 5′-ATGTCCACCACTTCTCCCAG-3′, reverse primer 5′-TATGTGGGACCCTGACCTCT-3′; Amplicon 2 chr7:6589486-6589684 (minus strand) forward primer 5′-ACATTGGGCAGCAGATGTTG-3′, reverse primer 5′-GGGACCAATGTGACGAGAGA-3′; Amplicon 3 chr7:20343672-20343865 (plus strand) forward primer 5′-AGCCAGAAGAAGCCAGGTAG-3′, reverse primer 5′-CGTCGGTCTTTTGCTACTCG-3′. One more pair of primers was designed to detect Hist1h1a mRNA, which is also processed into piRNAs (**Supplemental Fig. S5C**): forward primer 5′-GTCGCTCAGGCTGCTTCTAC-3′, reverse primer 5′-GGATGCTTTCGCCTTCACTGAC-3′. Care was taken so that the above primers hybridize on unique sequences and not on repeat sequences (**Supplemental Fig. S5C**). GAPDH was used as internal reference.

Histological, immunofluorescence and western blotting analyses.

For histology, testes were fixed in Bouin's solution, dehydrated in ethanol, embedded in paraffin, sectioned, and stained with hematoxylin and eosin. For immunofluorescence, E16.5 testes were fixed in 4% PFA, dehydrated in sucrose, embedded in tissue freezing medium, frozen in dry ice/ethanol, and sectioned with a cryostat. The following primary antibodies were used for immunofluorescence and Western blotting: anti-MOV10L1 (Zheng et al., 2010), 17.8 mouse monoclonal anti-MILI (Kirino et al., 2009), 13E-3 anti-MILI, anti-MIWI2 (Pandey et al., 2013), anti-LINE1 Orf1p (Martin and Branciforte, 1993), anti-Myc (Vazyme), M2 anti-FLAG (Sigma), and anti-ACTB (Sigma)

Immunoprecipitation and detection of piRNAs.

Embryonic day 16.5 (E16.5) testes were used. Testicular extract preparation, immunoprecipitation, and 5′ end labeling of piRNAs were carried out as previously described (Kirino et al., 2011; Reuter et al., 2009). Equal numbers of E16.5 *Mov10l1*$^{KI/+}$ and *Mov10l1*$^{KI/KI}$ testes were used in experiments on small RNAs associated with MILI and MIWI2.

MOV10L1 and mZuc expression in 293T cells, and FLAG immunoprecipitation from cell lysates

We co-expressed FLAG-tagged MOV10L1 (N-terminal tag, pRK5 vector, see above) and Myc-tagged mZuc (ENSMUST00000125307, 221 aa; C-terminal Myc-tag, pcDNA3.1 vector) in 293T cells. The cells were lysed in 50 mM Tris (pH 7.5), 100 mM NaCl, 1 mM EDTA, 0.5% Triton X-100, 10% Glycerol, 50mM NaF, 5 mM β-glycerophosphate, and protease inhibitors cocktail, and centrifuged for 30 min at 4°C. Supernatants were pre-cleared with sepharose protein A beads and incubated overnight with anti-FLAG antibody coated beads at 4°C. Beads were washed three times with lysis buffer, and incubated at 95°C for 5 min with SDS loading buffer. IP eluates were analyzed on 10% NuPAGE gels and electrotransferred to PVDF membranes. Western blot analysis was performed using anti-FLAG and anti-Myc antibodies.

MOV10L1 sequence and structure analysis.

We analyzed the primary sequence of mouse MOV10L1 using PFAM (http://pfam.sanger.ac.uk)(Punta et al., 2012) at the default settings. We used the Phyre 2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index)(Kelley and Sternberg, 2009) to generate a model of MOV10L1 and Armitage tertiary structure. Normal and intensive settings were used to produce the same overall structures. To further substantiate the similarity between MOV10L1 and UPF1, we acquired 14 UPF1 and 13 MOV10L1 sequences from Uniprot database, aligned them with Clustal Omega(Sievers et al., 2011), and used the Consurf server(Ashkenazy et al., 2010; Celniker et al., 2013) to visualize conserved residues within the predicted structure of mouse MOV10L1.

The following 14 UPF1 and 13 MOV10L1 sequences were retrieved from Uniprot (http://www.uniprot.org/):

UPF1 homologs: Q9FJR0|UPF1_ARATH; Q09820|UPF1_SCHPO; Q9EPU0|UPF1_MOUSE; A8I9D7|UPF1_CHLRE; Q92900|UPF1_HUMAN; Q9VYS3|UPF1_DROME; F1LY19|UPF1_RAT; Q7ZVZ4|UPF1_DANRE; P30771|UPF1_YEAST; K7B5Y2|UPF1_PANTR; F6VP39|UPF1_XENTR; E1BEK9|UPF1_BOVIN; E1C0J4|UPF1_CHICK; Q54I89|UPF1_DICDI.

MOV10L1 homologs: E1BY80|M10L1_CHICK; Q9BXT6|M10L1_HUMAN; Q99MV5|M10L1_MOUSE; D3ZH42|M10L1_RAT; Q0IHQ7|M10L1_XENTR; G3X7G5|M10L1_BOVIN; G3VZQ4|M10L1_SARHA; H3DEX1|M10L1_TETNG; F6YK83|M10L1_MONDO; F1RXR7|M10L1_PIG; Q6J5K9|ARMI_DROME; Q08BA5|M10L1_DANRE; G3UKW3|M10L1_LOXAF.

The above sequences were aligned using the Clustal Omega server (http://www.ebi.ac.uk/Tools/msa/clustalo/).

The multiple sequence alignment (MSA) was used together with the pdb file of the MOV10L1 model acquired from Phyre2 server for further analysis in the Consurf server (http://consurf.tau.ac.il) using the default (JTT) evolutionary substitution model and Bayesian calculation.

The surface electrostatic charge model of MOV10L1 was generated and visualized by Jmol on PDB2PQR Server (http://nbcr-222.ucsd.edu/pdb2pqr_1.8/)(Dolinsky et al., 2004).

Bioinformatic analyses

Sequenced reads pre-processing.

All sequenced reads were trimmed from the 3′ end to remove bases identified with lower quality. This pre-processing step corrects for Illumina Analyzer bias that calls bases with lower accuracy towards the 3′ end of the reads. The trimming was done as described in the BWA alignment software (Li and Durbin 2009). The 3′ end ligated adaptor (GTGTCAGTCACTTCCAGCGGTCGTATGCCGTCTTCTGCTTG) was removed from the sequences using the cutadapt software and a 0.25 acceptable error rate for the alignment of the adaptor on the read. To eliminate reads in which the adaptor was ligated more than one time, adaptor removal was performed 3 times. Reads shorter than 15 bases were excluded from further analyses.

Alignment and annotation.

Sequencing reads were aligned to the mouse (mm9) genome using the BWA alignment program and setting the default program parameters allowing for 0.04 fraction of missing alignments given a 2% uniform base error rate. All aligned reads shorter than 20nt that mapped to more than a single position on the genome were excluded from further analysis. The mouse (mm9) genome sequences, the genic and repeat (RepeatMasker) annotation were obtained from the University of California, Santa Cruz (UCSC) Genome Browser.

*Fine-tuning of intergenic piRNA clusters (IPC) definition (**Supplemental Table S1**).*

In our previous publication (Vourekas et al. 2012), the intergenic piRNA hotspots (IPH, elsewhere also called intergenic piRNA clusters, i.e. in (Li et al., 2013)) were defined in a strand independent way. Since almost all clusters are strand specific, we re-assembled these hotspots by splitting them into smaller pieces and defining a strand for each piece depending on the ratio of reads originating from each strand. The strand with the highest number of reads was assigned to the corresponding part. To be consistent with the nomenclature, the identified 119 stranded clusters are referred herein as Intergenic piRNA Clusters (IPC) (**Supplementary Table S1**). They take up the exact same length in absolute genomic coordinates as IPHs previously, and were used for all analyses involving IPH. IPH-IPC sequence space contains the overwhelming majority of intergenic piRNAs (see the Supplementary Figure 5 in (Vourekas et al., 2012)), and are in agreement with the respective clusters identified by (Li et al., 2013).

*MOV10L1 CLIP tag density on mRNAs (**Fig. 1E**).*

Transcript annotation was acquired from UCSC. mRNA regions (5′ UTR, CDS, and 3′ UTR) were divided in 20 bins. The average RPKM (reads that map within each bin per million mapped reads, per kilobase) of 20 bins for 5′ UTR, CDS and 3′ UTR is plotted for all mRNAs.

*Percent of piRNAs contained in CLIP libraries (**Supplemental Table S2**).*

We calculated the number of uniquely mapped reads in the CLIP libraries that are in the piRNA size (23-31nts) and that have the same 5′ end with pachytene and/or pre-pachytene Mili piRNAs from standard IP libraries (Aravin et al., 2007; Kirino et al., 2009).

MOV10L1 CLIP libraries and Mili piRNA overlap.

We measured the degree of tag overlap between our MOV10L1 CLIP libraries and piRNAs (23-31 nt) extracted from two combined Mili CLIP libraries enriched for piRNAs (MILI_1_small, MILI_3_small from (Vourekas et al., 2012)). Putative piRNA sequences were removed from the MOV10L1 libraries by selecting tags that start with a U and that are 27-30 nt long.

*Relative position of CLIP library reads with respect to piRNAs (**Supplemental Fig. S2C**).*

To examine the relative positioning of the reads of our libraries with the reads of a reference piRNA library, we calculated the distribution of the distances of the 5′ positions between all pairwise combinations of overlapping reads between the two libraries.

*Nucleotide enrichment of MOV10L1 CLIP tags mapping within intergenic piRNA clusters (**Fig. 3A, inset**).*

Statistical test: The nucleotide enrichment of the reads that map within IPCs was evaluated against a random background of reads. In detail, we created 1000 random sets of reads (the reads initially do not have sequence information) with same number and size distribution of reads as the original library. Each of the random reads was assigned in random positions within the IPCs. The sequence of the reads was extracted from the assigned positions. We measured the read nucleotide composition for the random sets and compared them to those of the original library. To calculate the p-value we measure how many times the random sets have higher or equal nucleotide composition than the original library and divide by the number of permutations. The p-value for G was smaller than 0.05 for all MOV10L1 CLIP libraries.

*Nucleotide enrichment of intergenic piRNA clusters **(Fig. 3B)**.*

We defined two extra sets of regions: a) intergenic but not piRNA clusters, b) intergenic non-piRNA clusters and non-repeat (Repeat Masker). We measured nucleotide content within these regions, and performed a pairwise Welch Two Sample t-test between any two nucleotide measurements.

*G-quadruplex prediction (**Fig. 3C-F, Supplemental Table S3**)*

The genome wide G-quadruplex prediction was performed using a sensitive general pattern: G_{L1}-N_{L2}-G_{L3}-N_{L4}-G_{L5}-N_{L6}-G_{L7}, where G corresponds to guanine and N to any nucleotide including G (adapted from (Ryvkin et al., 2010; Todd et al., 2005)). L1 to L7 correspond to the length of the stretch of G/N and are allowed to vary independently between a minimum and maximum value. For the loose prediction L1, L3, L5 and L7 vary from 2 to 4, while L2, L4 and L6 vary from 1 to 7. For the strict prediction, L1, L3, L5 and L7 can be 3 or 4, while L2, L4 and L6 vary from 1 to 7. The loose prediction identified 7,005,529 G-quadruplexes along the mouse genome, from which 2887 G4s were found on the transcribed strand and 2623 on the complementary strand of Intergenic piRNA Clusters. G4s cover 4.68% of the transcribed strand length (nt). The respective numbers for the strict prediction are 566,094 predicted G-quadruplexes along the mouse genome, of which 260 G4s found on the transcribed strand and 181 on the complementary strand of IPC loci, covering 0.48% of the IPC length (only for the transcribed strand). Locations of predicted G quadruplexes are provided in **Supplemental Table S3**.

*piRNA density within intergenic piRNA clusters and upstream of G-quadruplexes (**Fig. 3F, Supplemental Table S3**)*

Mili standard IP piRNAs were used for this analysis (it was established that CLIP piRNA and standard IP piRNA populations are essentially the same, (Vourekas et al., 2012)). The number of piRNAs within each intergenic piRNA cluster (IPC) over the sequence length of the IPC produced the average piRNA density, for all IPCs (**Supplemental Table S3**). If piRNAs are distributed homogeneously within IPC, no difference should be expected between average density and density in any particular sequence window within the IPC length. We calculated piRNA density within a 60 nt length window upstream of G4 5′ ends and compared with IPC average piRNA density, to examine whether piRNA generating loci are linked with G4 occurrence (this window length was chosen because it is long enough for just one intermediate piRNA processing fragment). Only clusters for which G4 quadruplexes were predicted were used in the following statistical test. For loose prediction 107 out of 119 clusters contained 2887 predicted G4s. Paired t-test was used for

statistical comparison of the piRNA density distributions. The piRNA density within the 60 nt window is significantly higher ($p<0.001$, paired t-test). For the strict prediction only 58 clusters out of 119 contained only 260 predicted G quadruplexes in total. The density in the 60 nt upstream windows was not found significantly higher, most probably because very few G quadruplexes were identified to reach statistical significance.

*Intergenic piRNA cluster sequence shuffling and G-quadruplex prediction (**Supplemental Table S3**).*

The sequences of all IPC loci were shuffled ten thousand (10000) times, and to each shuffled sequence we performed G-quadruplex prediction with the strict and the loose algorithm. The predicted G-quadruplexes were always fewer in the shuffled sequences (p-value $<10^{-4}$).

*Human, rat and macaque piRNAs (**Fig. 3E, G**).*

The human piRNA library was downloaded from NCBI using Entrez codes DQ569913 to DQ601958 (Girard et al., 2006) and was aligned to the human genome (hg19). Rat piRNAs (Lau et al., 2006) were acquired from GEO (Series GSE5026) and were aligned to rat genome (rn3). Macaque (Macaca mulatta) piRNAs (Rosenkranz and Zischler, 2012) were acquired from SRA database (accession SRX200124) and were aligned to the respective genome (NCBI build 1.2). Alignment was performed with the same parameters as the rest of the libraries.

*Genomic nucleotide preference around 5′ ends of piRNAs (**Fig. 3G**).*

The coordinates of 5′ ends of piRNAs were marked on the genome, and the nucleotide preference at upstream and downstream positions (+/- 50 bases) was identified. This analysis was performed using all piRNAs (23-31 nt) from Mili CLIP library 1 (mouse) (Vourekas et al., 2012), and piRNA libraries of human (Girard et al., 2006), rat (Lau et al., 2006), and Rhesus macaque (Rosenkranz and Zischler, 2012). For mouse piRNAs, we also performed the analysis using uniquely mapped piRNAs or after multiple piRNA copies in the libraries were collapsed, to avoid biases introduced by piRNA copy number, and we observed no qualitative difference in the nucleotide preference distribution (data not shown).

*Intergenic piRNA cluster secondary structure prediction (**Fig. 3H, I**).*

IPC sequences were folded using RNAfold (Hofacker et al., 1994). For sequences larger than 3kb, the sequence was cut in 3kb pieces with 1kb overlaps and the results were "stitched" together

keeping the bound/unbound identification of the fragment with the nearest center. RNAfold was run with -g flag allowing for G-quadruplex.

For 100 nt upstream and downstream of the mid point of each tag that maps within intergenic piRNA clusters and per nucleotide, we count the number of bound and unbound nucleotides. The percentage bound is bound/(bound+unbound).

Transcript expression calculation.

We calculated the expression for protein-coding transcripts by counting the number of RNA-Seq reads that map within the exons of each transcript. The counts were normalized using upper quartile normalization, effectively dividing each count by the upper quartile of all counts (Bullard et al., 2010).

SUPPLEMENTAL REFERENCES

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science (80-.). *316*, 744–747.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. *38*, W529–W533.

Azzalin, C.M. (2012). UPF1: a leader at the end of chromosomes. Nucleus *3*, 16–21.

Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. *5*, 182–186.

Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics *11*, 94.

Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., and Ben-Tal, N. (2013). ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. Isr. J. Chem. *53*, 199–206.

Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., and Conti, E. (2011). Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. Mol. Cell *41*, 693–703.

Cheng, Z., Muhlrad, D., Lim, M.K., Parker, R., and Song, H. (2007). Structural and functional insights into the human Upf1 helicase core. EMBO J. *26*, 253–264.

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., and Baker, N. a (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. *32*, W665–W667.

Fairman-Williams, M.E., Guenther, U.-P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. Curr. Opin. Struct. Biol. *20*, 313–324.

Franks, T.M., Singh, G., and Lykke-Andersen, J. (2010). Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense- mediated mRNA decay. Cell *143*, 938–950.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199–202.

Gorbalenya, A.E., and Koonin, E. V (1993). Helicases: amino acid sequence comparisons and structure-function relationships. Curr. Opin. Struct. Biol. *3*, 419–429.

Gregersen, L.H., Schueler, M., Munschauer, M., Mastrobuoni, G., Chen, W., Kempa, S., Dieterich, C., and Landthaler, M. (2014). MOV10 Is a 5’ to 3' RNA Helicase Contributing to UPF1 mRNA Target Degradation by Translocation along 3' UTRs. Mol. Cell 1–13.

Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., and Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. Monatshefte Für Chemie Chem. Mon. *125*, 167–188.

Iyer, L.M., Leipe, D.D., Koonin, E. V, and Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. J. Struct. Biol. *146*, 11–31.

Jankowsky, E., and Putnam, A. (2010). Duplex unwinding with DEAD-box proteins. Methods Mol. Biol. *587*, 245–264.

Kelley, L. a, and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. *4*, 363–371.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nat. Rev. Mol. Cell Biol. *13*, 700–712.

Kirino, Y., Kim, N., de Planell-Saguer, M., Khandros, E., Chiorean, S., Klein, P.S., Rigoutsos, I., Jongens, T.A., and Mourelatos, Z. (2009). Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. Nat Cell Biol *11*, 652–658.

Kirino, Y., Vourekas, A., Khandros, E., and Mourelatos, Z. (2011). Immunoprecipitation of piRNPs and Directional, Next Generation Sequencing of piRNAs. Methods Mol. Biol. *725*, 281–293.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA complex from rat testes. Science (80-.). *313*, 363–367.

Leipe, D.D., Koonin, E. V., and Aravind, L. (2003). Evolution and Classification of P-loop Kinases and Related Proteins. J. Mol. Biol. *333*, 781–815.

Lewandoski, M., Meyers, E.N., and Martin, G.R. (1997). Analysis of Fgf8 gene function in vertebrate development. Cold Spring Harb. Symp. Quant. Biol. *62*, 159–168.

Li, X.Z., Roy, C.K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B.W., Xu, J., Moore, M.J., Schimenti, J.C., Weng, Z., et al. (2013). An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol. Cell *50*, 67–81.

Martin, S.L., and Branciforte, D. (1993). Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. Mol. Cell. Biol. *13*, 5383–5392.

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. Proc. Natl. Acad. Sci. U. S. A. *12*.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res. *40*, D290–D301.

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. Nat. Struct. Mol. Biol. *16*, 639–646.

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

Ryvkin, P., Hershman, S.G., Wang, L., and Johnson, F.B. (2010). Computational approaches to the detection and analysis of sequences with intramolecular G-quadruplex forming potential. Methods Mol. Biol. *608*, 39–50.

Saikrishnan, K., Powell, B., Cook, N.J., Webb, M.R., and Wigley, D.B. (2009). Mechanistic basis of 5'-3' translocation in SF1B helicases. Cell *137*, 849–859.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Story, R.M., and Steitz, T.A. (1992). Structure of the recA protein-ADP complex. Nature *355*, 374–376.

Todd, A.K., Johnston, M., and Neidle, S. (2005). Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res. *33*, 2901–2907.

Vourekas, A., and Mourelatos, Z. (2014). HITS-CLIP (CLIP-Seq) for Mouse Piwi Proteins. Methods Mol. Biol. *1093*, 73–95.

Vourekas, A., Zheng, Q., Alexiou, P., Maragkakis, M., Kirino, Y., Gregory, B.D., and Mourelatos, Z. (2012). Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. Nat. Struct. Mol. Biol. *19*, 773–781.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. *1*, 945–951.

Weng, Y., Czaplinski, K., and Peltz, S.W. (1996). Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. Mol. Cell. Biol. *16*, 5477–5490.

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. Mol. Cell *47*, 970–979.

Zheng, K., and Wang, P.J. (2012). Blockade of pachytene piRNA biogenesis reveals a novel requirement for maintaining post-meiotic germline genome integrity. PLoS Genet. *8*, e1003038.

Zheng, K., Xiol, J., Reuter, M., Eckardt, S., Leu, N.A., McLaughlin, K.J., Stark, A., Sachidanandam, R., Pillai, R.S., and Wang, P.J. (2010). Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. Proc. Natl. Acad. Sci. U. S. A. *107*, 11841–11846.