The conserved zinc-finger protein GTSF1 helps PIWI proteins achieve their full catalytic potential

Amena Arif[1,2,6], Shannon Bailey[2], Natsuko Izumi[3], Todd A. Anzelon[4], Deniz M. Ozata[2,7], Cecilia Andersson[2], Ildar Gainetdinov[2], Ian J. MacRae[4], Yukihide Tomari[3,5] & Phillip D. Zamore[2,*]

[1]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Chan Medical School, Worcester, MA 01605, USA

[2]Howard Hughes Medical Institute and RNA Therapeutics Institute, University of Massachusetts Chan Medical School, 368 Plantation Street, Worcester, MA 01605, USA

[3]Laboratory of RNA Function, Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan

[4]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

[5]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan

[6]Current address: Beam Therapeutics, Cambridge, MA 02139

[7]Current address: Department of Molecular Biosciences, Stockholm University, SE-106 91 Stockholm, Sweden

*Correspondence: phillip.zamore@umassmed.edu (P.D.Z.)

Abstract

Argonaute proteins use nucleic acid guides to find and bind specific DNA or RNA target sequences. Argonaute proteins can be found in all kingdoms of life, and play diverse biological functions including genome defense, gene regulation, and chromosome partitioning. Many Argonautes retain their ancestral endoribonuclease activity, cleaving the phosphodiester bond between target nucleotides t10 and t11. In animals, a specialized class of Argonautes, the PIWI proteins, use 21–35 nt PIWI-interacting RNAs (piRNAs) to direct transposon silencing, protect the germline genome, and regulate gene expression during gametogenesis[1]. The piRNA pathway is required for fertility in one or both sexes of nearly all animals. Both piRNA production and function require RNA cleavage catalyzed by PIWI proteins. Spermatogenesis in mice and other placental mammals requires three distinct, developmentally regulated PIWI proteins: MIWI (PIWIL1), MILI (PIWIL2), and MIWI2 (PIWIL4)[2–4]. The piRNA-guided endoribonuclease activities of MIWI and MILI are essential to produce functional sperm[5,6]. piRNA-directed silencing in mice and insects also requires Gametocyte-Specific Factor 1 (GTSF1), a PIWI-associated protein of unknown function[7–12]. Here, we report that GTSF1 potentiates the weak, intrinsic, piRNA-directed RNA cleavage activities of PIWI proteins, transforming them into efficient endoribonucleases. GTSF1 represents the first example of an auxiliary protein that potentiates the catalytic activity of an Argonaute protein.

Main

In animals, 21–35 nt PIWI-interacting RNAs (piRNAs) direct PIWI proteins to silence transposons and regulate gene expression[1,13–27]. Invertebrates produce piRNAs and PIWI proteins in both the soma and the germline[28–35] but mammalian piRNAs act only in the germline[36–43]. Mice lacking any of their three PIWI proteins—MIWI2 (PIWIL4), MILI (PIWIL2), and MIWI (PIWIL1)—or other proteins required for piRNA production are invariably male sterile[1–4]. As in other animals, mouse piRNA production requires catalytically active PIWI proteins, MILI and MIWI[5,6]. Transposon silencing is the ancestral function of piRNAs; uniquely, placental mammals also produce pachytene piRNAs, which first appear shortly after the onset of meiosis I (refs. 40–42,44–46) and reach a peak abundance in spermatocytes rivalling that of ribosomes[16]. Pachytene piRNAs tune the abundance of mRNAs required for spermiogenesis, the process by which round spermatids become sperm[47–50]. Pachytene piRNAs have been proposed to direct MIWI and MILI to cleave specific mRNAs, ensuring appropriate levels of their protein products[47,50–52]. Testing this idea has been thwarted by the absence of a cell-free system in which MIWI or MILI can be loaded with synthetic piRNAs of defined sequence.

Recombinant MIWI loaded with a piRNA of defined sequence

We used lentiviral transduction to engineer a stable HEK293T cell line over-expressing epitope-tagged MIWI. Tagged MIWI was captured from cell lysate using anti-FLAG antibody coupled to paramagnetic beads, incubated with a synthetic piRNA bearing a monophosphorylated 5′ terminus and 2′-*O*-methylated 3′ end, and eluted from the magnetic beads using 3XFLAG peptide. MIWI loaded with a synthetic piRNA (MIWI piRISC; Fig. 1a), but not unloaded apo-MIWI (Extended Data Fig. 1a), cleaved a 5′ ^{32}P-radiolabled target RNA fully complementary to the synthetic guide (Extended Data Fig. 1b, c and Supplementary Data Fig. 1a).

Recombinant MIWI bound stably to an RNA guide bearing a 5′ monophosphate but not to a guide with a 5′ hydroxyl group (Extended Data Fig. 1d). The MID domain of Argonaute proteins contains a 5′ monophosphate-binding pocket that anchors the RNA guide to the protein[53–59]. Mutations predicted to disrupt 5′ monophosphate-binding perturb PIWI function[55,60–64]. We immobilized MIWI on paramagnetic beads via its 3XFLAG tag, incubated it with guide RNA, washed the beads, eluted the MIWI piRISC with 3XFLAG tag peptide, and tested its ability to cleave a fully complementary target RNA. Incubation with a 5′ monophosphorylated but not an otherwise identical 5′ hydroxy guide yielded MIWI piRISC that cleaved target RNA (Extended Data Fig. 1d). In vivo, the methyltransferase HENMT1 adds a 2′-*O*-methyl group to the 3′ ends of piRNAs. *Henmt1*$^{-/-}$ male mice and *henn1*$^{-/-}$ worms are sterile[65–71]. Terminal 2′-*O*-methyl modification likely stabilizes small RNAs against degradation by cellular ribonucleases rather than secures the guide to the protein[72,73]. Consistent with this function for 2′-*O*-methylation, piRNAs bearing a 3′ terminal 2′ hydroxyl or 2′-*O*-methyl were equally functional in directing target cleavage, provided that the piRNA was 5′ monophosphorylated (Extended Data Fig. 1d).

Purified MIWI RISC is a sluggish enzyme

Although affinity-purified MIWI loaded with a piRNA bearing 5′ monophosphorylated and 3′, 2′-*O*-methylated termini specifically cleaved a fully complementary target RNA at the phosphodiester bond that links target nucleotides t10 to t11, the rate of cleavage (0.01 min^{-1}) was >300 times slower than that catalyzed by mouse AGO2 RISC (≥ 3 min^{-1})[74,75]. At physiological temperature (37°C), mouse AGO2 RISC catalyzes multiple rounds of target cleavage[76]. By contrast, 5 nM MIWI piRISC cleaved only ~15% of the target RNA (5 nM) after 1 h (Fig. 1b, Extended Data Fig. 1e, and Supplementary Data Fig. 1b). Inefficient target cleavage by MIWI piRISC was not caused by the presence of the amino-terminal 3XFLAG-SNAP tandem tag: removing the tandem tags

using tobacco etch virus protease generated an untagged protein (Extended Data Fig. 1f and Supplementary Data Fig. 1c) whose target cleavage kinetics were identical to that of piRISC assembled with the tagged MIWI (Extended Data Fig. 1g).

The ubiquitously expressed arginine methyltransferase PRMT5 modifies the amino terminus of PIWI proteins, allowing it to bind Tudor domain-containing proteins, many of which are required for piRNA biogenesis, gametogenesis, and fertility[77–79]. A potential explanation for the sluggish activity of recombinant MIWI piRISC is that it lacks arginine methylation. We used mass spectrometry to map the positions of methyl arginine in our affinity-purified MIWI. All of the arginine residues previously shown to be methylated in endogenous MIWI immunoprecipitated from mouse testis[77] were methylated in recombinant MIWI produced in HEK293T cells (Extended Data Fig. 1h).

Another possible explanation for the inefficiency of target cleavage by MIWI piRISC is that the recombinant protein, although properly arginine methylated, lacks other post-translational modifications. To test this idea, we immobilized apo-MIWI, incubated it with wild-type (C57BL/6) mouse testis lysate in the presence of an ATP regenerating system at 25°C for 15 min, removed the testis lysate by washing, and then loaded MIWI with a synthetic piRNA and eluted the resulting piRISC. Pre-incubation of MIWI with testis lysate either before or after loading with a piRNA failed to increase its target cleaving activity (Extended Data Fig. 2a, b). We conclude that neither a missing post-translational modification nor a tightly associated protein partner is likely to explain the low activity of MIWI piRISC compared to AGO2 RISC.

MIWI piRISC requires an auxiliary factor to efficiently cleave its target

Adding testis lysate increased the rate of target cleavage by affinity-purified MIWI piRISC ~20-fold (Fig. 1b; Extended Data Fig. 2b). This effect cannot be attributed to the lysate contributing additional piRISC, because testis lysate alone failed to generate detectable cleavage product (Fig. 1b).

MIWI is first produced as spermatogonia enter meiosis and differentiate into spermatocytes[3,80]. To determine whether the MIWI-potentiating factor is differentially expressed during spermatogenesis, we supplemented the target cleavage assay with lysate prepared from FACS-purified germ cells: piRISC-potentiating activity was greatest in lysate from secondary spermatocytes (Fig. 1c), a cell type in which pachytene-piRNA-directed target cleavage is readily detected in vivo[47,50]. Moreover, the potentiating activity was testis-specific: lysates from brain, liver, or kidney failed to enhance piRNA-directed target cleavage by MIWI (Extended Data Fig. 3a). Finally, the potentiating activity was specific for PIWI proteins and had no effect on the rate of target cleavage by mouse AGO2 (Extended Data Fig. 3b).

Three lines of evidence suggest that the MIWI-potentiating activity contains one or more structural Zn^{2+} ions. First, pre-treating testis lysate with the sulfhydryl alkylating agent *N*-ethylmaleimide inactivated the potentiating activity (Extended Data Fig. 3c), indicating reduced cysteine residues, which often bind divalent metal cations[81], are essential. Second, the MIWI-potentiating activity was unaltered by EGTA, which chelates Ca^{2+}, but was irreversibly inactivated by EDTA, which chelates many divalent metals, and by 1,10-Phenanthroline, which specifically chelates Zn^{2+} (Extended Data Fig. 3d). Adding additional metal ions failed to rescue the activity (Extended Data Fig. 3e–g), suggesting that loss of Zn^{2+} irreversibly denatures the MIWI-potentiating factor, a characteristic of zinc-finger proteins[82,83]. Third, the MIWI-potentiating activity bound more tightly to an immobilized metal-affinity resin charged with Zn^{2+} than to resin charged with Ni^{2+} (Extended Data Fig. 3h).

To identify the MIWI-potentiating activity, we developed a chromatographic purification scheme using cation-exchange, hydrophobic-interaction, and size-exclusion chromatography (Fig. 2a). Notably, the activity eluted from a Superdex 200 size-exclusion column as a single ~17 kDa peak (Fig. 2b). Taken together, our data suggest

that the MIWI-potentiating activity corresponds to a small, testis-specific, Zn^{2+}-binding protein abundantly expressed in meiotic and post-meiotic male germ cells (Fig. 2c).

GTSF1 is necessary and sufficient for efficient target cleavage by MIWI

Gametocyte Specific Factor 1 (GTSF1) is a conserved, 19,083 Da, tandem CHHC-type zinc-finger protein essential for piRNA function and fertility in flies[7,8,10], silk moths[11], worms[84], and mice[9,12]. Like *Miwi* and *Mili*, the mRNA abundance of *Gtsf1*, as well as its paralogs *Gtsf1L* and *Gtsf2*, rises sharply as male germ cells enter meiosis I, peaks in secondary spermatocytes, and then declines in round spermatids (Extended Data Fig. 4a). Immobilized mouse GTSF1 captures all three PIWI proteins from testis lysate, but GTSF1 and PIWI proteins co-expressed in HEK293T cells do not co-immunoprecipitate, leading to the suggestion that the GTSF1-PIWI interaction requires additional protein or RNA components[9]. MIWI2-bound piRNAs are lost in *Gtsf1*$^{-/-}$ male mice[9], which are sterile[12] likely because loss of GTSF1 causes loss of MIWI2-directed retrotransposon promoter methylation[4,85–87]. Although RNA cleavage by MIWI2 is not required for piRNA biogenesis or function, production of MIWI2 piRISC requires MILI-dependent piRNA amplification, a process that requires the endonuclease activity of MILI[6]. In flies, piRNA-directed transcriptional silencing of transposons by Piwi, but not Piwi piRISC-assembly, requires the GTSF1 ortholog Asterix[7,8].

To test whether the MIWI-potentiating activity in the testis lysate corresponded to GTSF1, we used CRISPR-Cas9 to engineer a mouse strain with a 3XFLAG-HA epitope tag inserted into the endogenous *Gtsf1* coding sequence (*Gtsf1*^{KI}; Fig. 2d). Because the MIWI potentiating activity was greatest in secondary spermatocytes, we prepared lysate from FACS-purified secondary spermatocytes from homozygous *Gtsf1*$^{KI/KI}$ male mice. Incubation with anti-FLAG antibody coupled to paramagnetic beads depleted the lysate of both epitope-tagged GTSF1 (Fig. 2e) and the MIWI-potentiating activity (Fig. 2f). The activity was recovered from the beads by elution with 3XFLAG peptide (Fig. 2f). In

contrast, lysate from C57BL/6 secondary spermatocytes retained the MIWI-potentiating factor after incubation with anti-FLAG antibody, and no activity was eluted from the beads after incubation with FLAG peptide. Thus, GTSF1 is necessary to potentiate the catalytic activity of MIWI.

GTSF1 is also sufficient to potentiate the catalytic activity of MIWI. Adding purified recombinant GTSF1 (Extended Data Fig. 4b) to MIWI-piRISC increased the rate of target cleavage by MIWI 19–100-fold for three different piRNA sequences (Fig. 3a-c). First, the addition of 500 nM GTSF1 to MIWI programmed with an artificial, 30-nt piRNA (5 nM) caused the pre-steady-state rate (k_{burst}) of target cleavage to increase from 0.010 min^{-1} to 1.5 min^{-1} (Fig. 3b), a rate similar to that of AGO2 (≥ 3 min^{-1})[74,75]. Second, programming MIWI with an endogenous mouse piRNA sequence from the 9-qC-10667.1 locus (*pi9*) that is antisense to the L1MC transposon increased the pre-steady-state rate of cleavage of a fully complementary target RNA ~19-fold: k_{burst} = 0.033 min^{-1} for piRISC alone and 0.62 min^{-1} with GTSF1 added (Fig. 3c and Extended Data Fig. 5). Finally, for a 6-qF3–28913(−),8009(+) locus (*pi6*) piRNA that directs cleavage at a partially complementary site in the *Scpep1* mRNA in diplotene spermatocytes [47], GTSF1 increased k_{burst} 80-fold, from 0.0040 min^{-1} to 0.32 min^{-1} (Fig. 3c and Extended Data Fig. 5).

Our data suggest that GTSF1 does not promote product release or turnover: the steady-state rate (k_{ss} <0.005 min^{-1}) of target cleavage under multiple-turnover conditions (i.e., GTSF1 >> target RNA >> MIWI) was essentially unchanged from the rate in the absence of GTSF1 (Fig. 3b). Perhaps MIWI remains bound to the cleaved products, preventing it from catalyzing multiple rounds of target cleavage. Supporting this idea, at incubation times >15 min, 3′-to-5′ exonucleases present in testis lysate degrade the uncut, full-length target RNA, but the 5′ cleavage product remains stable, consistent with its being protected by MIWI bound to its 3′ end (Fig. 1b). In vivo, product

release has been proposed to be facilitated in insects and mammals by the RNA-stimulated ATPase Vasa (DDX4)[88–90] and the Vasa-like protein DDX43 (ref. 91).

RNA-binding is essential for GTSF1 function

Only three known eukaryotic proteins contain CHHC zinc-fingers: the spliceosomal RNA-binding protein U11-48K, the TRM13 tRNA methyltransferase, and GTSF1 and its paralogs[92]. In vitro, the first zinc-finger of GTSF1 directly binds RNA; RNA binding requires four basic surface residues (R26, R29, K36 and K39)[93]). Potentiation of target cleavage required RNA-binding by GTSF1: purified mutant GTSF1^{R26A,R29A,K36A,K39A} (Extended Data Fig. 4b) did not detectably increase the rate of catalysis by either MIWI or MILI (Fig. 3d–f and Extended Data Table 2), suggesting that GTSF1 must interact with RNA to function.

GTSF1 function requires PIWI-binding

In flies and mice, GTSF1 interacts with PIWI proteins through conserved aromatic residues in its central region[7,9] (Fig. 2c). Mutating these residues in the recombinant protein—W98A, W107A, and W112A for mouse GTSF1—reduced the stimulatory effect of mouse GTSF1 on MIWI in a concentration-dependent manner, supporting the idea that GTSF1 binds MIWI directly when potentiating target cleavage (Fig. 3g and Extended Data Table 2) and suggesting that W98, W107, and W112 define a conserved surface of GTSF1-that interacts with PIWI proteins. We measured the pre-steady-state catalytic rate (k_{burst}) as a function of GTSF1 concentration. Fitting the data to a hyperbolic function revealed that GTSF1 binds MIWI piRISC >60-fold more tightly than GTSF1^{W98A,W107A,W112A} (GTSF1: K_D = 8 nM, 95% confidence interval [C.I.] = 6 – 9 nM; GTSF1^{W98A,W107A,W112A}: K_D = 500 nM, 95% C.I. = 100 – 900 nM) (Fig. 3g). Moreover, PIWI-binding appears to be the sole defect in GTSF1^{W98A,W107A,W112A}, since the first-order rates of target cleavage by MIWI at saturating concentrations of either wild-type or

mutant GTSF1 were essentially indistinguishable. (wild-type: 1.2 min^{-1}, 95% C.I. = 0.7 – 2 min^{-1}; mutant GTSF1: 0.8 min^{-1}, 95% C.I. = 0.6 – 0.9 min^{-1}).

The PIWI-stimulatory function of GTSF1 orthologs is evolutionarily conserved

GTSF1 orthologs can be found in many metazoan genomes[92], suggesting that GTSF1 may potentiate target cleavage by PIWI proteins in many animals. Supporting this idea, purified recombinant Gtsf1 from the arthropod *Bombyx mori* (Extended Data Fig. 4b) potentiated the catalytic activity of the *B. mori* PIWI protein, BmSiwi (Extended Data Fig. 6b).

The structure and kinetics of EfPiwi from the sponge *Ephydatia fluviatilis* were recently described[64]. Like MIWI, EfPiwi possesses inherently weak catalytic activity, a feature common to all PIWI proteins examined to date. Although the *E. fluviatilis* genome is yet to be sequenced, the closely related, fully sequenced genome of *Ephydatia muelleri* contains a readily identifiable GTSF1 ortholog. Purified recombinant EmGtsf1 (Extended Data Fig. 4c) stimulated the single-turnover catalytic rate of EfPiwi piRISC ~28 fold (Fig. 3h). Sponges (Porifera) are the sister group to all other animals, the Eumetazoa, having separated ~900 Mya. Thus, the last common ancestor of all animals likely required GTSF1 to potentiate target cleavage by PIWI proteins.

The GTSF1 tandem zinc-finger domains are conserved across phyla, whereas the GTSF1 central and carboxy-terminal sequence diverges substantially between mammals and arthropods (Extended Data Table 1). For example, the sequences of the mouse and rhesus macaque first and second zinc-fingers are 100% identical, whereas their C-terminal domains share 88.5% identity. The first zinc-finger of mouse GTSF1 is 37.5% and 45.8% identical to its fly and moth orthologs, but the mouse protein shares just 8% and 8.3% identity with the central and carboxy-terminal domains of the fly and moth proteins, respectively. Consistent with the evolutionary divergence of their central and carboxy terminal domains, testis lysate from rat or rhesus macaque enhanced

target cleavage by mouse MIWI piRISC, whereas lysate from *Drosophila melanogaster* or *Trichoplusia ni* ovaries, *T. ni* Hi5 cells, or purified recombinant *Bombyx mori* BmGtsf1 did not (Extended Data Fig. 4c, 5a, b and Supplementary Data Fig. 1d).

GTSF1 paralogs can distinguish among PIWI proteins within a species

Animal genomes often encode more than one GTSF protein[92] (Extended Data Fig. 5c). For example, *D. melanogaster* has four GTSF paralogs. The *D. melanogaster* OSC and OSS cell lines, which derive from somatic follicle cells that support oogenesis, express Piwi but lack the PIWI paralogs Aub or Ago3. Piwi-mediated, piRNA-guided transposon silencing in these cells requires Asterix, a GTSF1 ortholog[7,8,10]. In vivo, *asterix* mutants phenocopy *piwi* mutants and are female sterile, even though Piwi is successfully loaded with piRNAs and transits to the nucleus in the absence of Asterix[7,8]. Whether the other fly GTSF paralogs have a function in vivo, perhaps as auxiliary factors for Aub or Ago3, remains to be tested. Like fly Asterix, mouse GTSF1 is essential for piRNA function and fertility. In mice, two GTSF1 paralogs, GTSF1L and GTSF2 are also expressed during spermatogenesis and interact with PIWI proteins[94]. Unlike GTSF1, single and double *Gtsf1l* and *Gtsf2* knockout males are fertile[94]. Genes encoding the *Gtsf* paralogs are syntenic in mammals, while *Gtsf2* is lost in primates (Extended Data Fig. 6).

The central and carboxy-terminal domains of GTSF orthologs are more similar among closely related species than among GTSF paralogs within the same species (Extended Data Fig. 4b and 5c and Extended Data Table 1), further supporting the view that this domain has evolved to bind specific PIWI proteins. Consistent with this idea, mouse GTSF1, GTSF1L, and GTSF2 differ in their ability to potentiate target cleavage by MIWI and MILI. While GTSF1 accelerated target cleavage by both MIWI and MILI, purified, recombinant GTSF1L and GTSF2 (Extended Data Fig. 4b) efficiently potentiated target cleavage by MIWI but not MILI (Fig. 3d–f and Extended Data Table 2). GTSF2 was unable to detectably increase the rate of target cleavage by MILI.

Although GTSF1L had a modest effect on the rate of target cleavage by MILI piRISC, this enhancement was less than one-sixteenth that provided by GTSF1 and half that of the PIWI-binding mutant GTSF1^{W98A,W107A,W112A} (Fig. 3d–f and Extended Data Table 2). We conclude that GTSF1L and GTSF2 are specialized to potentiate target cleavage by MIWI piRISC.

Silk moth BmGtsf1-like is more similar to mouse GTSF1 than to BmGtsf1 (Extended Data Fig. 6c and Extended Data Table 1); BmGtsf1-like increased amount of target cleaved by MIWI piRISC by 2.6-fold (S.D. = 0.1; p = 0.0027 by ANOVA with Dunnett's post-hoc test) but had no detectable effect on MILI (Extended Data Fig. 6b), further supporting the idea that the GTSF central domain determines the affinity of the protein for specific PIWI proteins. In vivo, BmGtsf1 interacts with Siwi and is required for transposon silencing and sex determination[11]. BmGtsf1 increased the rate of target cleavage by affinity-purified BmSiwi but not that of the other silk moth PIWI protein, BmAgo3 (Extended Data Fig. 6b).

Target cleavage by MIWI or MILI requires extensive pairing between piRNA and target

piRNA:target RNA complementarity from g2–g22 (i.e., 21 base-pairs) is required for efficient target cleavage directed by endogenous piRNAs loaded into MIWI piRISC immunoprecipitated from adult mouse testis[5]. However, GTSF1 does not detectably co-immunoprecipitate with MIWI piRISC from mouse testis[5,77]. A requirement for 21 bp complementarity between target and guide is unprecedented among Argonaute proteins: fly Ago2 slices a target with as few as 11 contiguous base pairs[95]; mammalian AGO2 requires only 11 contiguous base pairs for detectable cleavage[96]; and the eubacterial DNA-guided DNA endonuclease TtAgo requires as few as 14 (ref. 97).

Affinity-purified MIWI, programmed with either of two different synthetic, 30 nt piRNAs, readily cleaved target RNA complementary to guide nucleotides g2–21 but not

a target complementary to g2–g16 (Extended Data Fig. 7a, b). Under multiple-turnover conditions with saturating amounts of purified, recombinant GTSF1 ([GTSF1] >> [target] > [piRISC]), MIWI readily cleaved a target RNA with 19 nucleotides (g2–g20) complementary to its synthetic piRNA guide (Fig. 4a, top). The lower background of single-turnover experiments ([GTSF1] >> [piRISC] > [target]) allowed longer incubation times. Using these conditions, we could detect GTSF1-stimulated cleavage of a target RNA with as few as 15 complementary nucleotides (g2–g16; Fig. 4a, bottom). We note that the pachytene stage of meiosis in mouse spermatogenesis lasts about 175 h, and the pachytene piRNA pathway components are expressed until at least the round spermatid stage, a time interval spanning >400 h.

Guide length limits the rate of target cleavage by MIWI

In vivo, piRNAs are trimmed to a length characteristic of the PIWI protein in which they reside: ~30 nt for MIWI and ~26–27 nt for MILI[41,42,87]. An attractive hypothesis is that these piRNA lengths are optimal for target cleavage catalyzed by the specific PIWI protein. Our data suggest a more complex relationship between piRNA length, PIWI protein identity, and target complementarity. In the presence of GTSF1, MIWI loaded with a 30 nt piRNA and MILI loaded with a 26 nt piRNA readily cleaved a fully complementary target RNA in a 60 min reaction (Fig. 4b). By contrast, neither piRISC cleaved a target complementary to piRNA positions g2–g16 (Fig. 4a, b and Extended Data Fig. 7a, b). Similarly, MIWI loaded with a 26 nt or MILI loaded with a 26 or 21 nt guide produced little cleavage for a target complementary to positions g2–g16 although both guide lengths supported cleavage of a fully complementary target (Fig. 4b). In fact, for MIWI, a 21mer was more active than a 30 nt guide (Fig. 4b). Without GTSF1, MIWI or MILI loaded with any of these guide lengths produced little cleaved target in 60 min (Fig. 4b). Remarkably, MIWI or MILI loaded with a 16 nt guide RNA, a piRNA length not present in vivo, readily cleaved the RNA target (Fig. 4b).

GTSF1 accelerates the rate of pre-steady state target cleavage (k_{burst}) by MIWI and MILI but has little effect on the nearly slow rate of steady-state cleavage (k_{ss}), suggesting that piRISC remains bound to its cleavage by-products (Fig. 3e). We incubated GTSF1 and MIWI—loaded with a 30, 26, 21, or 16 nt guide—with a target RNA fully complementary to each guide and measured k_{burst} and k_{ss} (Fig. 4c). As the guide length decreased, k_{burst} decreased and k_{ss} increased. Compared to its native 30 nt guide length, the 16 nt guide increased k_{ss} ~9-fold and decreased k_{burst} >50-fold. These data suggest that as the guide was shortened, the cleaved products were released more rapidly.

We estimated the binding affinity (ΔG) of the piRNA for its target using the nearest-neighbor rules for base pairing at 37°C. As the strength of the piRNA base pairing to the target increased, the rate of pre-steady-state cleavage increased (Fig. 5a), consistent with base pairing serving to extract the 3′ end of the piRNA from the PAZ domain, facilitating the transition to a more catalytically competent piRISC conformation[98]. Conversely, decreased base-pairing strength increased the steady-state rate of target cleavage, supporting the view that for biologically relevant piRNA lengths, product release is the rate-determining step for MIWI-catalyzed target cleavage (Fig. 5a).

Discussion

In nearly all animals, both piRNA biogenesis and piRNA function require the PIWI endoribonuclease activity. Yet purified mouse MIWI and MILI are intrinsically slow to cleave complementary target RNAs. Our data show that unlike AGO proteins, PIWI proteins require an auxiliary factor, GTSF1, to efficiently cleave their RNA targets.

The ability of GTSF1 to potentiate PIWI-catalyzed target cleavage provides a biochemical explanation for the genetic requirement for this small zinc-finger protein in the piRNA pathway. GTSF1 function requires that it bind both RNA and the PIWI

protein, and differences among carboxy-terminal domains restrict individual GTSF1 paralogs to specific PIWI proteins.

We propose a testable kinetic scheme for target cleavage by MIWI (Fig. 5b, c and Extended Data Fig. 8) that incorporates the requirement for GTSF1 and the observation that a 16 nt guide changes the rate-determining step of target cleavage catalyzed by MIWI. As originally proposed for fly Ago2 (ref. 98), piRISC bound to a target is presumed to exist in two-states: one in which the piRNA 3′ end is secured in the PAZ domain and a competing, pre-catalytic conformation in which the piRNA is fully base paired to its target. Structures of Piwi-A from the freshwater sponge *Ephydatia fluviatilis* show that extensive pairing between a piRNA and its target induces a catalytically competent conformation in which the PAZ domain is rotated away from the piRNA 3′ end[64]. We propose that the PAZ-bound state is more favorable for guides bound to PIWI proteins than to AGOs, requiring a high degree of complementarity between guide and target to extract the piRNA 3′ terminus from the PAZ domain. Our data also suggest that a slow rate of product release after cleavage results directly from the extensive piRNA-target RNA complementarity required for this conformational change. A 16 nt piRNA allows MIWI more easily to assume a pre-catalytic state, perhaps because the 3′ end of the short guide cannot reach the PAZ domain. Notably, single-stranded siRNA guides as short as 14 nt allow mammalian AGO3, initially believed to have lost its endonuclease activity, to efficiently cleave RNA targets[99]. In golden hamsters, piRNAs bound to PIWIL1 are initially ~29 nt long, but a shorter population of ~23 nt piRNAs appears at metaphase II and predominates in two-cell embryos[100]. piRNAs bound to PIWIL3, a female-specific PIWI protein absent from mice, are ~19 nt long in hamster and ~20 nt in human oocytes[100,101]. We speculate that these short piRNAs allow PIWIL1 and PIWIL3 to function as multiple-turnover endonucleases.

Our model proposes that GTSF1 recognizes the pre-catalytic piRISC state, facilitating a second conformational change in PIWI proteins that may occur

spontaneously in AGO clade proteins. GTSF1 binding likely stabilizes the catalytically active conformation, facilitating target cleavage, but has no detectable effect on subsequent release of the cleaved products. Slow product release may be a general property of PIWI proteins: purified *Ephydatia fluviatilis* (freshwater sponge) PIWI similarly catalyzes only a single round of target cleavage[64].

References

1. Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D. & Zamore, P. D. PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **20**, 89-108 (2018).

2. Kuramochi-Miyagawa, S. et al. *Mili*, a mammalian member of *piwi* family gene, is essential for spermatogenesis. *Development* **131**, 839-849 (2004).

3. Deng, W. & Lin, H. *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* **2**, 819-830 (2002).

4. Carmell, M. A. et al. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* **12**, 503-514 (2007).

5. Reuter, M. et al. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* **480**, 264-267 (2011).

6. De Fazio, S. et al. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480**, 259-263 (2011).

7. Dönertas, D., Sienski, G. & Brennecke, J. *Drosophila* Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. *Genes Dev.* **27**, 1693-1705 (2013).

8. Ohtani, H. et al. DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the *Drosophila* ovary. *Genes Dev.* **27**, 1656-1661 (2013).

9. Yoshimura, T. et al. Mouse GTSF1 is an essential factor for secondary piRNA biogenesis. *EMBO Rep.* **19**, e42054 (2018).

10. Muerdter, F. et al. A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in *Drosophila*. *Mol. Cell.* **50**, 736-748 (2013).

11. Chen, K. et al. Gtsf1 is essential for proper female sex determination and transposon silencing in the silkworm, *Bombyx mori*. *PLoS Genet.* **16**, e1009194 (2020).

12. Yoshimura, T. et al. *Gtsf1/Cue110*, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testes. *Dev. Biol.* **335**, 216-227 (2009).

13. Mohn, F., Handler, D. & Brennecke, J. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* **348**, 812-817 (2015).

14. Han, B. W., Wang, W., Li, C., Weng, Z. & Zamore, P. D. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science* **348**, 817-821 (2015).

15. Homolka, D. et al. PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. *Cell Rep* **12**, 418-428 (2015).

16. Gainetdinov, I., Colpan, C., Arif, A., Cecchini, K. & Zamore, P. D. A Single Mechanism of Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA Production in Most Animals. *Mol. Cell.* **71**, 775-790.e5 (2018).

17. Ipsaro, J. J., Haase, A. D., Knott, S. R., Joshua-Tor, L. & Hannon, G. J. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* **491**, 279-283 (2012).

18. Ding, D. et al. PNLDC1 is essential for piRNA 3′ end trimming and transposon silencing during spermatogenesis in mice. *Nat Commun* **8**, 819 (2017).

19. Zhang, Y. et al. An essential role for PNLDC1 in piRNA 3′ end trimming and male fertility in mice. *Cell Res* **27**, 1392-1396 (2017).

20. Nishimura, T. et al. PNLDC1, mouse pre-piRNA Trimmer, is required for meiotic and post-meiotic male germ cell development. *EMBO Rep.* **19**, (2018).

21. Wang, H. et al. Antagonistic roles of Nibbler and Hen1 in modulating piRNA 3' ends in *Drosophila*. *Development* **143**, 530-539 (2016).

22. Feltzin, V. L. et al. The exonuclease Nibbler regulates age-associated traits and modulates piRNA length in *Drosophila*. *Aging Cell* **14**, 443-452 (2015).

23. Izumi, N., Shoji, K., Suzuki, Y., Katsuma, S. & Tomari, Y. Zucchini consensus motifs determine the mechanism of pre-piRNA production. *Nature* **578**, 311-316 (2020).

24. Izumi, N. et al. Identification and Functional Analysis of the Pre-piRNA 3' Trimmer in Silkworms. *Cell* **164**, 962-973 (2016).

25. Haase, A. D. et al. Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*. *Genes Dev.* **24**, 2499-2504 (2010).

26. Gunawardane, L. S. et al. A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5′ End Formation *in Drosophila*. *Science* **315**, 1587-1590 (2007).

27. Brennecke, J. et al. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell* **128**, 1089-1103 (2007).

28. Lewis, S. H. et al. Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat Ecol Evol* **2**, 174-181 (2018).

29. Fu, Y. et al. The genome of the Hi5 germ cell line from *Trichoplusia ni*, an agricultural pest and novel model for small RNA biology. *eLife* **7**, e31628 (2018).

30. Jehn, J. et al. PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. *Commun Biol* **1**, 137 (2018).

31. Lim, R. S., Anand, A., Nishimiya-Fujisawa, C., Kobayashi, S. & Kai, T. Analysis of Hydra PIWI proteins and piRNAs uncover early evolutionary origins of the piRNA pathway. *Dev. Biol.* **386**, 237-251 (2014).

32. Juliano, C. E. et al. PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells. *Proc. Natl. Acad. Sci. U S A* **111**, 337-342 (2014).

33. Praher, D. et al. Characterization of the piRNA pathway during development of the sea anemone *Nematostella vectensis*. *RNA Biol* **14**, 1727-1741 (2017).

34. Grimson, A. et al. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**, 1193-1197 (2008).

35. Kerner, P., Degnan, S. M., Marchand, L., Degnan, B. M. & Vervoort, M. Evolution of RNA-binding proteins in animals: insights from genome-wide analysis in the sponge *Amphimedon queenslandica*. *Mol. Biol. Evol.* **28**, 2289-2303 (2011).

36. Murchison, E. P. et al. Conservation of small RNA pathways in platypus. *Genome Res.* **18**, 995-1004 (2008).

37. Houwing, S. et al. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**, 69-82 (2007).

38. Toombs, J. A. et al. Xenopus Piwi proteins interact with a broad proportion of the oocyte transcriptome. *RNA* **23**, 504-520 (2017).

39. Sun, Y. H. et al. Domestic chickens activate a piRNA defense against avian leukosis virus. *eLife* **6**, e24695 (2017).

40. Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* **20**, 1709-1714 (2006).

41. Aravin, A. et al. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203-207 (2006).

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

43. Lim, S. L. et al. Conservation and expression of PIWI-interacting RNA pathway genes in male and female adult gonad of amniotes. *Biol. Reprod.* **89**, 136 (2013).

44. Lau, N. C. et al. Characterization of the piRNA complex from rat testes. *Science* **313**, 363-367 (2006).

45. Özata, D. M. et al. Evolutionarily conserved pachytene piRNA loci are highly divergent among modern humans. *Nat Ecol Evol* **4**, 156-168 (2020).

46. Li, X. Z. et al. An Ancient Transcription Factor Initiates the Burst of piRNA Production during Early Meiosis in Mouse Testes. *Mol. Cell.* **50**, 67-81 (2013).

47. Wu, P. H. et al. The evolutionarily conserved piRNA-producing locus *pi6* is required for male mouse fertility. *Nat. Genet.* **52**, 728-739 (2020).

48. Choi, H., Wang, Z. & Dean, J. Sperm acrosome overgrowth and infertility in mice lacking chromosome 18 pachytene piRNA. (2020).

49. Gou, L. T. et al. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res* 1-21 (2014).

50. Goh, W. S. et al. piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes Dev.* **29**, 1032-1044 (2015).

51. Vourekas, A. et al. Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nat. Struct. Mol. Biol.* **19**, 773-781 (2012).

52. Zhang, P. et al. MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. *Cell Res* **25**, 193-207 (2015).

53. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209-213 (2008).

54. Schirle, N. T. & MacRae, I. J. The crystal structure of human Argonaute2. *Science* **336**, 1037-1040 (2012).

55. Matsumoto, N. et al. Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. *Cell* **167**, 484-497.e9 (2016).

56. Parker, J. S., Roe, S. M. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* **23**, 4727-4737 (2004).

57. Ma, J. B. et al. Structural basis for 5′-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434**, 663-666 (2005).

58. Yuan, Y. R. et al. Crystal structure of *A. aeolicus* Argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell.* **19**, 405-419 (2005).

59. Ma, J. B. et al. Structural basis for 5′-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434**, 666-670 (2005).

60. Le Thomas, A., Marinov, G. K. & Aravin, A. A. A transgenerational process defines piRNA biogenesis in Drosophila virilis. *Cell Rep* **8**, 1617-1623 (2014).

61. Yamaguchi, S. et al. Crystal structure of Drosophila Piwi. *Nat Commun* **11**, 858 (2020).

62. Kawaoka, S., Izumi, N., Katsuma, S. & Tomari, Y. 3′ end formation of PIWI-interacting RNAs in vitro. *Mol. Cell.* **43**, 1015-1022 (2011).

63. Cora, E. et al. The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. *RNA* **20**, 773-781 (2014).

64. Anzelon, T. A. et al. Structural basis for piRNA targeting. *Nature* **597**, 285-289 (2021).

65. Kirino, Y. & Mourelatos, Z. The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA* **13**, 1397-1401 (2007).

66. Kirino, Y. & Mourelatos, Z. 2′-*O*-methyl modification in mouse piRNAs and its methylase. *Nucleic Acids Symp Ser (Oxf)* 417-418 (2007).

67. Lim, S. L. et al. HENMT1 and piRNA Stability Are Required for Adult Male Germ Cell Transposon Repression and to Define the Spermatogenic Program in the Mouse. *PLoS Genet.* **11**, e1005620 (2015).

68. Kamminga, L. M. et al. Hen1 is required for oocyte development and piRNA stability in zebrafish. *EMBO J.* **29**, 3688-3700 (2010).

69. Montgomery, T. A. et al. PIWI associated siRNAs and piRNAs specifically require the Caenorhabditis elegans HEN1 ortholog henn-1. *PLoS Genet.* **8**, e1002616 (2012).

70. Billi, A. C. et al. The *Caenorhabditis elegans* HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs. *PLoS Genet.* **8**, e1002617 (2012).

71. Kamminga, L. M. et al. Differential impact of the HEN1 homolog HENN-1 on 21U and 26G RNAs in the germline of Caenorhabditis elegans. *PLoS Genet.* **8**, e1002702 (2012).

72. Pelisson, A., Sarot, E., Payen-Groschene, G. & Bucheton, A. A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates

complementary sense *gypsy* transcripts in somatic cells of the *Drosophila* ovary. *J. Virol.* **81**, 1951-1960 (2007).

73. Ameres, S. L. et al. Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328**, 1534-1539 (2010).

74. Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D. & Serebrov, V. Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell* **162**, 84-95 (2015).

75. Becker, W. R. et al. High-Throughput Analysis Reveals Rules for Target RNA Binding and Cleavage by AGO2. *Mol. Cell.* **75**, 741-755.e11 (2019).

76. Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056-2060 (2002).

77. Vagin, V. V. et al. Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev.* **23**, 1749-1762 (2009).

78. Siomi, M. C., Mannen, T. & Siomi, H. How does the royal family of Tudor rule the PIWI-interacting RNA pathway? *Genes Dev.* **24**, 636-646 (2010).

79. Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. *Mol. Cell.* **33**, 1-13 (2009).

80. Kuramochi-Miyagawa, S. et al. Two mouse *piwi*-related genes: *miwi* and *mili*. *Mech Dev* **108**, 121-133 (2001).

81. Michael, S. F., Kilfoil, V. J., Schmidt, M. H., Amann, B. T. & Berg, J. M. Metal binding and folding properties of a minimalist Cys2His2 zinc finger peptide. *Proc. Natl. Acad. Sci. U S A* **89**, 4796-4800 (1992).

82. Matt, T., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. The CBP/p300 TAZ1 domain in its native state is not a binding partner of MDM2. *Biochem. J.* **381**, 685-691 (2004).

83. Nyborg, J. K. & Peersen, O. B. That zincing feeling: the effects of EDTA on the behaviour of zinc-binding transcriptional regulators. *Biochem. J.* **381**, e3-4 (2004).

84. Almeida, M. V. et al. GTSF-1 is required for formation of a functional RNA-dependent RNA Polymerase complex in *Caenorhabditis elegans*. *EMBO J.* **37**, e99325 (2018).

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

86. Kojima-Kita, K. et al. MIWI2 as an Effector of DNA Methylation and Gene Silencing in Embryonic Male Germ Cells. *Cell Rep* **16**, 2819-2828 (2016).

87. Aravin, A. A. et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell.* **31**, 785-799 (2008).

88. Wenda, J. M. et al. Distinct Roles of RNA Helicases MVH and TDRD9 in PIWI Slicing-Triggered Mammalian piRNA Biogenesis and Function. *Dev. Cell* **41**, 623-637.e9 (2017).

89. Xiol, J. et al. RNA Clamping by Vasa Assembles a piRNA Amplifier Complex on Transposon Transcripts. *Cell* (2014).

90. Nishida, K. M. et al. Respective functions of two distinct Siwi complexes assembled during PIWI-interacting RNA biogenesis in *Bombyx* germ cells. *Cell Rep* **10**, 193-203 (2015).

91. Murakami, R., Sumiyoshi, T., Negishi, L. & Siomi, M. C. DEAD-box polypeptide 43 facilitates piRNA amplification by actively liberating RNA from Ago3-piRISC. *EMBO Rep.* **22**, e51313 (2021).

92. Andreeva, A. & Tidow, H. A novel CHHC Zn-finger domain found in spliceosomal proteins and tRNA modifying enzymes. *Bioinformatics* **24**, 2277-2280 (2008).

93. Ipsaro, J. J., O'Brien, P. A., Bhattacharya, S., Palmer III, A. G. & Joshua-Tor, L. Asterix/Gtsf1 links tRNAs and piRNA silencing of retrotransposons. *Cell Reports* **34**, 108914 (2021).

94. Takemoto, N., Yoshimura, T., Miyazaki, S., Tashiro, F. & Miyazaki, J. *Gtsf1l and* Gtsf2 Are Specifically Expressed in Gonocytes and Spermatids but Are Not Essential for Spermatogenesis. *PLoS One* **11**, e0150390 (2016).

95. Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nat. Struct. Mol. Biol.* **11**, 599-606 (2004).

96. Jackson, A. L. et al. Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21**, 635-637 (2003).

97. Wang, Y. et al. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**, 754-761 (2009).

98. Tomari, Y. & Zamore, P. D. Perspective: machines for RNAi. *Genes Dev.* **19**, 517-529 (2005).

99. Park, M. S., Sim, G., Kehling, A. C. & Nakanishi, K. Human Argonaute2 and Argonaute3 are catalytically activated by different lengths of guide RNA. *Proceedings of the National Academy of Sciences* **117**, 28576-28578 (2020).

100. Ishino, K. et al. Hamster PIWI proteins bind to piRNAs with stage-specific size variations during oocyte maturation. *bioRxiv* 2020.12.01.407411; doi: https://doi.org/10.1101/2020.12.01.407411 (2020).

101. Yang, Q. et al. Single-cell CAS-seq reveals a class of short PIWI-interacting RNAs in human oocytes. *Nature Comm.* **10**, 1-15 (2019).

Methods

Plasmids and cell lines

Supplementary Data Table 1 provides the sequences of all oligonucleotides used. To create pScalps_Puro_EGFP, an IRES-driven EGFP was inserted downstream of the multiple cloning site and upstream of the puromycin coding sequence of the lentivirus transfer vector pScalps_Puro. MIWI cDNA was obtained from Mammalian Gene Collection (https://genecollections.nci.nih.gov/MGC/). MILI cDNA was amplified by RT-PCR from mouse testis total RNA. Gibson assembly and restriction cloning were used to clone the MILI and MIWI coding sequences into pScalps_Puro_EGFP, fusing them in-frame with N-terminal 3XFLAG and SNAP tags. Lentivirus transfer vectors were packaged by co-transfection with psPAX2 and pMD2.G (4:3:1) using TransIT-2020 (Mirus Bio) in HEK293T cells. Supernatant containing lentivirus was used to transduce HEK293T cells in the presence of 16 ug/ml polybrene (Sigma) to obtain stable PIWI-expressing cell lines. Three sequential transductions were performed to maximize recombinant protein production. The transduced cells were selected in the presence of 2 ug/ml Puromycin for two weeks, then the cells expressing the 5–10% highest EGFP fluorescence were selected by FACS (UMASS Medical School Flow Cytometry Core). Selected cells stably expressing the recombinant PIWI proteins were expanded, harvested, and cell pellets flash-frozen and stored at −80°C.

Mouse GTSF1 cDNA was synthesized at Twist Biosciences and cloned into pCold-GST (Takara Bio) bacterial expression vector by restriction cloning. GTSF1 mutants, GTSF1L, and GTSF2- expressing pCold-GST vectors were synthesized at Twist Biosciences. pIZ-FLAG6His-Siwi was described previously[102]. FLAG-tagged MILI and MIWI-expressing vectors were the kind gift of Shinpei Kawaoka (Kyoto University, Kyoto, Japan). MmGtsf1 cDNA was amplified by RT-PCR from mouse spermatogonial stem cell total RNA[103]. BmGtsf1 and BmGtsf1-like cDNAs were amplified by RT-PCR from BmN4 cell total RNA. The amplified cDNA fragment and a DNA fragment coding

V5SBP were cloned into pcDNA5/FRT/TO vector (Thermo Fisher Scientific) by In-fusion cloning (Takara). The *Ephydatia muelleri* Gtsf1 coding sequence was obtained from [104] and cloned into bacterial expression vector pSV272 (IJM).

Mice

Generation of 3XFLAG-HA-tagged GTSF1 mice (*Gtsf1*$^{KI/KI}$) was performed at Cyagen. The coding sequence for the tags was inserted into the endogenous locus by CRISPR-Cas9. Briefly, fertilized mouse embryos were injected with the sgRNA targeting the sequence GTCTTCCATGCTGATGGCAAAGG (PAM), a 3XFLAG-HA-tag cassette HDR donor, and Cas9 mRNA. Supplementary Data Table 1 provides the sequences of the HDR donor and oligonucleotide primers used for genotyping. Founder F0 animals were genotyped and bred to generate F1 animals carrying the germline-transmitted knock-in allele. Mice were maintained and used according to the guidelines of the Institutional Animal Care and Use Committee of the University of Massachusetts Chan Medical School (A201900331). C57BL/6J mice (RRID: IMSR_JAX:000664) were used as wild-type control, where indicated.

Recombinant protein purification

PIWI proteins. PIWI-expressing stable cells were harvested by centrifugation and stored at −80°C until lysed in 10 ml lysis buffer (30 mM HEPES-KOH, pH 7.5, 100 mM potassium acetate, 3.5 mM magnesium acetate, 1 mM DTT, 0.1% v/v Triton X-100, 20% v/v glycerol, and 1× protease inhibitor cocktail [1 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride [Sigma; A8456], 0.3 μM Aprotinin, 40 μM betanin hydrochloride, 10 μM. E-64 [Sigma; E3132], 10 μM leupeptin hemisulfate]) per g frozen cells. Cell lysis was monitored by staining with trypan blue. Crude cytoplasmic lysate was clarified at 20,000 (S20) or 100,000 × *g* (S100) for 30 min, aliquoted, flash-frozen, and stored at −80°C. To capture PIWI proteins, clarified

lysate was incubated with 20 µl Anti-FLAG M2 paramagnetic beads (Sigma) per ml of lysate for 2 h to overnight rotating at 4°C. Beads were washed five times with high salt wash buffer (30 mM HEPES-KOH, pH 7.5, 2 M potassium acetate, 3.5 mM magnesium acetate, 1 mM DTT), twice with low salt buffer (high salt wash buffer except containing 100 mM potassium acetate and 0.01% v/v Triton X-100). To assemble RISC, beads were resuspended in low salt buffer containing 100 nM synthetic guide piRNA and incubated with rotation at 37°C or room temperature for 30 min. MIWI RISC or unloaded apo-MIWI was eluted from the beads with 200 ng/ul 3XFLAG peptide in lysis buffer without 0.1% v/v Triton X-100 or protease inhibitors for 2 h at 4°C. Eluate containing PIWI piRISC was aliquoted and stored at −80°C. EfPiwi piRISC purification has been described[64]. AGO2 siRISC was purified using oligo-affinity purification as described[105].

GTSF1, GTSF1 mutants, and GTSF1 homologs. pCold-GST GTSF-expression vectors were transformed into Rosetta-Gami 2 competent cells (Novagen). Cells were grown in the presence of 1 µM $ZnSO_4$ at 37°C until OD_{600} ~0.6–0.8, then chilled on ice for 30 min to initiate cold shock. Protein expression was induced with 0.5 mM IPTG for 18 h at 15°C. Cells were harvested by centrifugation, washed twice with PBS, and cell pellets were flash frozen and stored at −80°C. Cell pellets were resuspended in lysis/GST column buffer containing 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM DTT, 5% v/v glycerol, and complete EDTA-free protease inhibitor cocktail (Roche). Cells were lysed by a single pass at 18,000 psi through a high-pressure microfluidizer (Microfluidics M110P), and the resulting lysate clarified at 30,000 × *g* for 1 h at 4°C. Clarified lysate was filtered through a 0.2 µm low-protein binding syringe filter (Millex Durapore; EMD Millipore) and applied to glutathione Sepharose 4b resin (Cytiva) equilibrated with GST column buffer. After draining the flowthrough, the resin was washed with 50 column-volumes GST column buffer. To elute the bound protein and cleave the GST tag in a single step, 50 U HRV3C Protease (Novagen) was added to the column, the column

was sealed and incubated for 3 h at 4°C, following which, the column was drained to collect the cleaved protein. The protein was diluted to 50 mM NaCl and further purified using a Hi-Trap Q (Cytiva) anion exchange column equilibrated with 20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM DTT, and 5% v/v glycerol. The bound protein was eluted using a 100–500 mM NaCl gradient in the same buffer. Peak fractions were analyzed for purity by SDS-PAGE and the purest were pooled and dialyzed into storage buffer containing 30 mM HEPES-KOH, pH 7.5, 100 mM potassium acetate, 3.5 mM magnesium acetate, 1 mM DTT, 20% v/v glycerol. To avoid precipitation of GTSF2, which has a pI = 7.3, 20 mM Tris-HCl, pH 8.8, was substituted for HEPES-KOH during purification and dialysis.

For Extended Data Fig.6b, HEK293T cells were transfected with mouse and insect GTSF1 coding sequences cloned into pcDNA5 expression vectors using Lipofectamine 3000 (Thermo Fisher Scientific) according to the manufacturer's instructions. Cells were harvested 36 h later and homogenized in lysis buffer [20 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1.5 mM $MgCl_2$, 0.15% v/v Triton X-100, 100 µg/ml RNase A (Qiagen), 0.5 mM DTT, 1× Complete EDTA-free protease inhibitor (Roche)]. After centrifugation at 17,000 × *g* for 20 min at 4°C, additional NaCl (0.65 M f.c.) and 1% v/v Triton X-100 (f.c.) were added to the supernatant and the lysate was incubated with Streptavidin-Sepharose High Performance (Cytiva) beads at 4°C for 1 h. The beads were washed with wash buffer [20 mM Tris-HCl (pH 7.4), 1 M NaCl, 1.5 mM $MgCl_2$, 1% v/v Triton X-100, 0.5 mM DTT] five times and rinsed with lysis buffer without RNase A. SBP-tagged recombinant proteins were eluted with elution buffer (30 mM HEPES-KOH, pH 7.4, 100 mM potassium acetate, 2 mM magnesium acetate, 2.5 mM biotin, 0.5 mM DTT).

For Figure 3h, EmGtsf1 expression vector was transformed into BL21(DE3) cells (New England Biolabs). Transformed cells were grown in LB supplemented with 1 µM ZnSO4 at 37°C until OD600 ~0.6–0.8. The incubation temperature was lowered to 16°C

and protein expression was induced by addition of 1 mM IPTG for 4 h. Cells were harvested by centrifugation and cell pellets were flash frozen in liquid nitrogen and stored at −80°C for future use. Thawed cell pellets were resuspended in lysis buffer (50 mM Tris, pH 8, 300 mM NaCl, 0.5 mM TCEP) and passed through a high-pressure (18,000 psi) microfluidizer (Microfluidics M110P) to induce cell lysis. The resulting lysate was clarified by centrifugation at 30,000 × *g* for 20 min at 4°C. Clarified lysate was applied to Ni-NTA resin (Qiagen) and incubated for 1 h. The resin was extensively washed with Nickel Wash Buffer (300 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, 50 mM Tris, 8). Protein was eluted in four column volumes of Nickel Elution Buffer (300 mM NaCl, 300 mM imidazole, 0.5 mM TCEP, 50 mM Tris, pH 8). TEV protease was added to the eluted protein to induce cleavage and removal of the N-terminal His6 and MBP tags. The resulting mixture was dialyzed against Hi-Trap Dialysis Buffer (300 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, 50 mM Tris, pH 8) at 4°C overnight. The dialyzed protein was then passed through a 5-ml Hi-Trap Chelating column (Cytiva) and the unbound material collected. Unbound material was concentrated and further purified by size exclusion chromatography using a Superdex 200 Increase 10/300 column (Cytiva) equilibrated in 50 mM Tris, pH 8, 300 mM NaCl, and 0.5 mM TCEP. Peak fractions were analyzed for purity by SDS-PAGE, and the purest were pooled, concentrated to 150 µM, aliquoted, and stored at −80°C.

Immunodepletion and Western Blotting

Approximately 300,000 FACS-purified Secondary Spermatocytes (Sp2) from wild-type (C57BL/6) or 3XFLAG-HA-GTSF1-expressing (*Gtsf1*$^{KI/KI}$) mice were lysed in lysis buffer (30 mM HEPES-KOH, pH 7.5, 100 mM potassium acetate, 3.5 mM magnesium acetate, 1 mM DTT, 0.1% v/v Triton X-100, 20% v/v glycerol, and 1× protease inhibitor cocktail [1 mM 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride [Sigma; A8456], 0.3 µM Aprotinin, 40 µM betanin hydrochloride, 10 µM. E-64 [Sigma; E3132], 10 µM leupeptin

hemisulfate]). Crude cytoplasmic lysate was clarified at 20,000 × *g* (S20) for 30 min at 4°C. The S20 protein concentration was measured using Bradford Assay. The S20 was incubated with Anti-FLAG paramagnetic beads overnight at 4°C. The supernatant was removed, and the beads washed thrice with lysis buffer but containing 1M potassium acetate (high salt wash) and then twice with lysis buffer. Bound proteins were eluted by incubation with 200 ng/ul 3XFLAG peptide for 2 h at 4°C. The extent of depletion was assayed by western blotting. Briefly, the samples were resolved on SDS-PAGE and transferred to nitrocellulose membrane (Amersham Protran 0.45 NC; Cytiva). The membrane was probed using mouse anti-FLAG antibody (Sigma, F3165; 1:1,000 in blocking buffer, Rockland Immunochemicals, MB-070) and rabbit anti-Tubulin (Cell Signaling, 2144; 1:1000) for 2 hours at room temperature. After washing 5× with TBS (+0.1% v/v Tween-20), the membrane was incubated with secondary antibodies - donkey anti-rabbit IRDye 680RD (LI-COR, 926-68073; 1:15,000) and goat anti-mouse IRDye 800CW (LI-COR, 926-32210; 1:15,000) for 30 minutes at room temperature. The membrane was washed 5× with TBS (+0.1% v/v Tween-20) and imaged using Odyssey Infrared Imaging System (LI-COR).

Northern Blotting

Northern blotting was performed as described[106]. Briefly, piRNA guide standards and PIWI RISCs were first resolved on a denaturing 15% polyacrylamide gel, followed by transfer to Hybond-NX (Cytiva) neutral nylon membrane by semi-dry transfer at 20 V for 1 h. Next, crosslinking was performed in the presence of 0.16 M EDC in 0.13 M 1-methylimidazole, pH 8.0, at 60°C for 1 h. The crosslinked membrane was pre-hybridized in Church's buffer (1% w/v BSA, 1 mM EDTA, 0.5 M phosphate buffer, and 7% w/v SDS) at 45°C for 1 h. Radiolabeled, 5′ ^{32}P-DNA probe (25 pmol) in Church's buffer was added to the membrane and allowed to hybridize overnight at 45°C, followed by five

washes with 1× SSC containing 0.1% w/v SDS. The membrane was air dried and exposed to a storage phosphor screen.

Chromatographic fractionation of the MIWI-potentiating activity

Dissected animal tissues were homogenized in lysis buffer in a Dounce homogenizer using 10 strokes of the loose-fitting pestle A, followed by 20 strokes of tight-fitting pestle B. Lysate was clarified at 20,000 × *g*, followed by 0.2 μm filtration to yield an S20 for further chromatographic purification. Lysates used without further purification were directly prepared in 30 mM HEPES-KOH, pH 7.5, 100 mM potassium acetate, 3.5 mM magnesium acetate, 1 mM DTT, and 20% v/v glycerol ("dialysis buffer") with 1× protease inhibitor homemade cocktail; column fractions were dialyzed into this buffer before assaying. Protein concentration was measured using the BCA assay. Chromatography buffers were filtered prior to use.

For chromatography, lysate was prepared as described except using 30 mM HEPES-KOH, pH 7.5, 50 mM NaCl, 1 mM DTT, 5% v/v glycerol, and protease inhibitors. The lysate was applied to HiTrap SP column (Cytiva) equilibrated with the lysis buffer. The column was washed, and the bound proteins eluted stepwise using increasing NaCl concentrations. The NaCl content of the SP column fractions containing the peak MIWI-potentiating activity was adjusted to 2 M and applied to HiTrap Phenyl (Cytiva) equilibrated with column buffer containing 2 M NaCl. Bound proteins were eluted stepwise using decreasing NaCl concentrations. The peak MIWI-potentiating fractions elute from the HiTrap Phenyl column were pooled, concentrated (10 kDa MWCO Amicon Ultra filter), and applied to Superdex 200 Increase 10-300 GL size-exclusion chromatography column (bed volume ~24 ml) equilibrated with the dialysis buffer but containing 5% v/v glycerol. The void volume (V_0) of the gel filtration column was determined with blue dextran, and all fractions (0.5 ml each), starting from just before V_0 to the end of the column (V_t) were assayed for MIWI potentiating activity. The

molecular weight of the potentiating activity was determined relative to molecular weight markers (beta amylase, 200 kDa; alcohol dehydrogenase, 150 kDa; albumin, 66 kDa; carbonic anhydrase, 29 kDa; and cytochrome C, 12.4 kDa).

Zn^{2+} and Ni^{2+} immobilized-metal affinity chromatography

HiTrap Chelating HP (Cytiva) columns were charged with 0.1 M $NiSO_4$ or $ZnSO_4$, washed with water, and then equilibrated in column buffer (20 mM potassium phosphate buffer, pH 7.5, 500 mM NaCl, 0.5 mM DTT, 5% v/v glycerol). S20 testis lysate was applied to the column, the flow-through collected, and the column washed with the column buffer until absorbance at 280 nm stabilized. Bound proteins were eluted in two steps: first, with 20 mM potassium phosphate, pH 7.5, 2 M ammonium chloride, 0.5 mM DTT, 5% v/v glycerol (elution buffer 1), and, second, with 20 mM potassium phosphate, pH 7.5, 500 mM NaCl, 200 mM imidazole, pH 8.0, 0.5 mM DTT, 5% v/v glycerol (elution buffer 2). The peak of each step was dialyzed into the dialysis buffer and assayed for the ability to potentiate MIWI catalysis.

Target cleavage assays

Target RNA substrates for in vitro cleavage assays were prepared as described[74,107,108]. Briefly, piRNA target site-containing templates were amplified by PCR, in vitro transcribed with T7 RNA polymerase, purified by urea-PAGE, and radiolabeled using α-^{32}P GTP (3000 Ci/mmol; Perkin Elmer), *S*-adenosyl methionine, and vaccinia virus RNA guanylyl transferase as described[108]. Unincorporated α-^{32}P GTP was removed using a G-25 spin column (Cytiva), and target RNA gel purified. In Extended Data Fig. 8a, target cleavage was monitored using synthetic RNA oligonucleotides radiolabeled by ligating [5′-^{32}P] cytidine 3′,5′-bisphosphate to the 3′ end of the target with T4 RNA ligase I (Ambion). The [5′-^{32}P] cytidine 3′,5′-bisphosphate was prepared by incubating 1 mM cytidine 3′-monophosphate (Sigma) with 312.5 pmole [γ-^{32}P] ATP (6000 Ci/mmol;

Perkin Elmer) with 25 U T4 polynucleotide kinase (NEB) at 37°C for 1 h, followed by 70°C for 30 min to inactivate the kinase.

Radiolabeled target (3–100 nM f.c.) was added to purified PIWI piRISC (2–8 nM), plus ~1 µg tissue or sorted germ cell lysate per 10 µl reaction volume or 0.5 µM (f.c.) of purified GTSF protein. At the indicated times, an aliquot of a master reaction was quenched in 4 volumes 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 25 mM EDTA, 1% w/v SDS, then proteinase K (1 mg/ml f.c.) was added and incubated at 45°C for 15 min. An equal volume of urea loading buffer (8 M urea, 25 mM EDTA) was added to the reaction time points, heated at 95°C for 2 min, and resolved by 7–10% denaturing PAGE. Gels were dried, exposed to a storage phosphor screen, and imaged on a Typhoon FLA 7000 (GE).

The raw image file was used to quantify the substrate and product bands, corrected for background. Data were fit to the reaction scheme

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \overset{k_2}{\rightarrow} EP \overset{k_3}{\rightarrow} E + P$$

using the burst-and-steady-state equation, Equation 1.

$$[P] = f(t) = [E]([k_2/(k_2 + k_3)]^2 \times (1 - e^{-[k2+k]t}) + [k_2k_3/(k_2 + k_3)]t) \qquad \text{(Equation 1)}$$

The time-dependence of product formation corresponded to a pre-steady-state exponential burst ($k_{burst} = k_2 + k_3$) followed by a linear steady-state phase, described by k_{cat}, *where* $k_{cat} = k_{ss} = k_2k_3/(k_2 + k_3)$.

The affinity (K_D) of wild-type or GTSF1^{W98A,W107A,W112A} for MIWI piRISC-target ternary complex and the maximum observable rate (k_{pot}) were estimated by measuring the pre-steady-state rate of target cleavage (k_{burst}) of MIWI at increasing concentrations of GTSF1 (1–5,000 nM) and fitting the data to Equation 2 (ref. 109:

$$k_{burst} = k_{pot}\,[\text{GTSF1}]/(K_D + [\text{GTSF1}]) \qquad \text{(Equation 2).}$$

FLAG-Siwi, MIWI, and MILI target-cleavage assays

FLAG-tagged BmSiwi, MIWI, or MILI-expressing vectors were transfected into BmN4 cells with X-tremeGENE HP DNA Transfection Reagent (Sigma). Preparation of loading lysates and single-stranded RNA loading were as described[110]. After loading with synthetic guide RNAs in the cell lysate, piRISC was immunoprecipitated with anti-FLAG antibody (Sigma) conjugated to Dynabeads protein G superparamagnetic beads (Thermo Fisher Scientific). The beads were washed five times with lysis buffer containing 30 mM HEPES-KOH, pH 7.4, 100 mM potassium acetate, 2 mM magnesium acetate, 0.5 mM DTT, and 0.1% v/v Empigen. Target cleavage assays were performed at 25°C for 3.5 h in 8 µl reaction containing 2.4 µl "40×" reaction mix[108], 100 nM recombinant GTSF1 protein and 0.5 nM ^{32}P cap-radiolabeled 28 nt target RNA as described[24].

BmAgo3 target-cleavage assays

Naive BmN4 cells were resuspended in lysis buffer (30 mM HEPES-KOH, pH 7.4, 100 mM potassium acetate, 2 mM magnesium acetate, 0.05% v/v Triton X-100, 0.5 mM DTT, 1× Complete EDTA-free protease inhibitor [Roche]) and lysed using a Dounce homogenizer. After centrifugation at 17,000 × *g* for at 4°C 20 min, the supernatant was incubated at 4°C for 1 h with anti-BmAgo3 antibody[23] conjugated to Dynabeads protein G (Thermo Fisher Scientific). The superparamagnetic beads were washed five times with wash buffer (30 mM HEPES-KOH, pH 7.4, 100 mM potassium acetate, 2 mM magnesium acetate, 0.5% v/v Empigen, 0.5% v/v Triton X-100, 0.5 mM DTT) and rinsed with lysis buffer. Target cleavage assays were performed at 25°C for 2.5 h in 8 µl reactions containing 2.4 µl "40×" reaction mix, 55 nM recombinant protein, and 0.5 nM ^{32}P cap-radiolabeled target 1 (complementary to endogenous piRNA #1) or target 2 (complementary to endogenous piRNA #2), prepared by in vitro transcription as described[111].

EfPiwi target cleavage assays

EfPiwi piRISC was purified as described[64]. Purified EfPiwi piRISC (100 nM f.c.) was incubated at 37°C with a 5′-^{32}P-radiolabeled target RNA complementary to g2-g21 (10 nM f.c.) in reaction buffer composed of 20 mM Tris, pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 2 mM $MnCl_2$, and 0.5 mM TCEP. Where indicated, purified *Ephydatia muelleri* GTSF1 (500 nM f.c.) or 0.1 mg/mL baker's yeast tRNA were also included. Target cleavage was stopped at indicated times by mixing aliquots of each reaction with an equal volume of denaturing gel loading buffer (98% w/v formamide, 0.025% w/v xylene cyanol, 0.025% w/v bromophenol blue, 10 mM EDTA, pH 8.0). Intact and cleaved target RNAs were resolved by denaturing PAGE (15%) and visualized by phosphorimaging. Quantification of signal was performed using ImageQuant TL (GE Healthcare).

Mouse germ cell purification

Germ cells from mouse testes were sorted and purified as described[16,47]. Briefly, freshly dissected mouse testes were decapsulated with 0.4 mg/ml collagenase type IV (Worthington) in 1× Gey's balanced salt solution (GBSS) at 33°C for 15 min. The separated seminiferous tubules were treated with 0.5 mg/ml trypsin and 1 μg/ml DNase I in 1× GBSS at 33°C for 15 min. Trypsin was then inactivated by adding 7.5% v/v fetal bovine serum (FBS). The cell suspension was filtered through a 70-μm cell strainer, and cells pelleted at 300 × *g* at 4°C for 10 min. Cell staining was performed at 33°C for 15 min with 5 μg/ml Hoechst 33342 prepared in 1× GBSS, 5% v/v FBS, and 1 μg/ml DNase I, then the cells were treated with 0.2 μg/ml propidium iodide, followed by final pass through a 40-μm cell strainer before sorting. Sorted cells were pelleted at 100 × *g* for 5 min, the buffer removed, and the cell pellets flash frozen and stored at −80°C. Cell lysates were prepared as described for HEK293T cells. Protein concentration was estimated using the BCA assay, and an equal amount of total protein from each cell type was used to assay for the ability to potentiate MIWI catalysis.

Analysis of RNA-seq data

Publicly available datasets[45,112–114] were analyzed. rRNA reads were removed using Bowtie 2.2.5 with default parameters[115]. After rRNA removal, the remaining reads were mapped to corresponding genomes (mouse, mm10; rat, rn6; macaque, rheMac8; human, hg19) using STAR. 2.3 with default parameters that allowed ≤ 2 mismatches and 100 mapping locations (ref. 116). Mapped results were generated in SAM format, duplicates removed and translated to BAM format using SAMtools 1.8 (ref. 117). HTSeq 0.9.1 with default parameters was used to count uniquely mapping reads[118] (steady-state transcript abundance was reported in reads per kilobase per million uniquely mapped reads [RPKM]).

Methylarginine analysis

Recombinant MIWI was immunopurified and resolved by electrophoresis on a 4–20% gradient SDS-polyacrylamide gel. The gel was fixed, stained with Coomassie G-250 (Simply Blue, Invitrogen), and the recombinant MIWI band excised and analyzed at the UMASS Mass Spectrometry Core. Gel slices were chopped into ~1 mm^2 pieces, 1 ml water added, followed by 20 µl 45 mM DTT in 250 mM ammonium bicarbonate. Samples were incubated at 50°C for 30 min, cooled to room temperature, then 20 µl 100 mM iodoacetamide (IAA) was added and incubated for 30 min. The solution was removed, and the gel pieces were three times washed with 1 ml water, 1 ml 50 mM ammonium bicarbonate:acetonitrile (1:1), quenched with 200 µl acetonitrile, and dried in a SpeedVac. Gel pieces were then rehydrated in 50 ul 50 mM ammonium bicarbonate containing 4 ng/µl trypsin (Promega, Madison, WI) and 0.01% proteaseMAX (Promega) and incubated at 37°C for 18 h. Supernatants were collected and extracted with 200 µl an 80:20 solution of acetonitrile: 1% (v/v) formic acid in water. Supernatants were then combined, and the peptides lyophilized in a SpeedVac and re-suspended in 25 µl 5% acetonitrile, 0.1% (v/v) formic acid and subject to mass spectrometry analysis. Data

were acquired using a NanoAcquity UPLC (Waters Corporation, Milford, MA) coupled to an Orbitrap Fusion Lumos Tribrid (Thermo Fisher Scientific, Waltham, MA) mass spectrometer. Peptides were trapped and separated using an in-house 100 µm I.D. fused-silica pre-column (Kasil frit) packed with 2 cm ProntoSil (Bischoff Chromatography, DE) C18 AQ (200 Å, 5 µm) media and configured to an in-house packed 75 µm I.D. fused-silica analytical column (gravity-pulled tip) packed with 25 cm ProntoSil (Bischoff; 100 Å, 3 µm) media. Mobile phase A was 0.1 % (v/v) formic acid in water; mobile phase B was 0.1 % (v/v) formic acid in acetonitrile. Following a 3.8 µl sample injection, peptides were trapped at flow rate of 4 µl/min with 5% B for 4 min, followed by gradient elution at a flow rate of 300 nl/min from 5–35% B in 90 min (total run time, 120 min). Electrospray voltage was delivered by liquid junction electrode (1.5 kV) located between the columns and the transfer capillary to the mass spectrometer was maintained at 275°C. Mass spectra were acquired over *m/z* 300–1750 Da with a resolution of 120,000 (*m/z* 200), maximum injection time of 50 ms, and an AGC target of 400,000. Tandem mass spectra were acquired using data-dependent acquisition (3 s cycle) with an isolation width of 1.6 Da, HCD collision energy of 30%, resolution of 15,000 (*m/z* 200), maximum injection time of 22 ms, and an AGC target of 50,000.

Raw data were processed using Proteome Discoverer 2.1.1.21 (Thermo Fisher Scientific), and the database search performed by Mascot 2.6.2 (Matrix Science, London, UK) using the Swiss-Prot human database (download 04/09/2019). Search parameters were: semi-tryptic digestion with up to two missed cleavages; precursor mass tolerance 10 ppm; fragment mass tolerance 0.05 Da; peptide N-terminal acetylation, cysteine carbamidomethylation, methionine oxidation, N-terminal glutamine to pyroglutamate conversion, arginine methylation and arginine demethylation were specified as variable modifications. Peptide and protein validation and annotation was done in Scaffold 4.8.9 (Proteome Software, Portland, OR) employing Peptide Prophet[119]

and Protein Prophet[120]. Peptides were filtered at 1% FDR, while protein identification threshold was set to greater than 99% probability and with a minimum of two identified peptides per protein. Only arginine modification sites detected in all three replicates from separate immunoprecipitation experiments are reported in the figure.

Statistics and reproducibility

Figs. 1b, 3, 4 and Extended Data Figs. 4a, 5c show mean ± SD for three independent trials. Figs. 2a, b, Extended Data Figs. 1a–c and 1h (three independent trials) and Figs. 2e, 2f, Extended Data Figs. 1f–g, 2b, 3a, 3c (two independent trials) and 4b, 4c show representative data. The experiments in Figs. 1c, 4, 5a, and Extended Data Figs. 1d–e, 2a, 3b, 3d–h, 5a, 7a–b were performed once. Protein sequences were aligned using Clustal Omega; unrooted tree was constructed using Randomized Axelerated Maximum Likelihood with default parameters[121] and visualized in Interactive Tree of Life[122].

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data are available from the authors upon request.

102. Kawaoka, S. et al. The *Bombyx* ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *RNA* **15**, 1258-1264 (2009).

103. Aoshima, K., Baba, A., Makino, Y. & Okada, Y. Establishment of alternative culture method for spermatogonial stem cells using knockout serum replacement. *PLoS One* **8**, e77715 (2013).

104. Kenny, N. J. et al. Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge Ephydatia muelleri. *Nat Commun* **11**, 3676 (2020).

105. Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Rapid and specific purification of Argonaute-small RNA complexes from crude cell lysates. *RNA* **19**, 271-279 (2013).

106. Pall, G. S. & Hamilton, A. J. Improved northern blot method for enhanced detection of small RNA. *Nat Protoc* **3**, 1077-1084 (2008).

107. Wee, L., Flores-Jasso, C., Salomon, W. & Zamore, P. Argonaute Divides Its RNA Guide Into Domains with Distinct Functions and RNA-Binding Properties. *Cell* **151**, 1055-1067 (2012).

108. Haley, B., Tang, G. & Zamore, P. D. In vitro analysis of RNA interference in *Drosophila melanogaster*. *Methods* **30**, 330-336 (2003).

109. Carlson, K. D., Johnson, R. E., Prakash, L., Prakash, S. & Washington, M. T. Human DNA polymerase kappa forms nonproductive complexes with matched primer termini but not with mismatched primer termini. *Proc. Natl. Acad. Sci. U S A* **103**, 15776-15781 (2006).

110. Kawaoka, S. et al. Zygotic amplification of secondary piRNAs during silkworm embryogenesis. *RNA* **17**, 1401-1407 (2011).

111. Yoda, M. et al. ATP-dependent human RISC assembly pathways. *Nat. Struct. Mol. Biol.* **17**, 17-23 (2010).

112. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338**, 1593-1599 (2012).

113. Yu, T. et al. Long first exons and epigenetic marks distinguish conserved pachytene piRNA clusters from other mammalian genes. *Nat Commun* **12**, 73 (2021).

114. Gainetdinov, I. et al. Terminal modification, sequence, length, and PIWI-protein identity determine piRNA stability. *Mol. Cell.* **81**, 4826-4842. e8 (2021).

115. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).

116. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

117. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

118. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).

119. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383-5392 (2002).

120. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646-4658 (2003).

121. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).

122. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242-W245 (2016).

Acknowledgements We are grateful to Ken-Edwin Aryee (Greiner/Bremh Lab) for providing pScalps_Puro, psPAX2, pMD2.G plasmids and for guidance on lentiviral transduction; Pei-Hsuan Wu for help with mouse testis germ cell FACS and for providing sorted cells; Katharine Cecchini for mouse tissues; Elaine Norowski (Greiner/Bremh labs) for rat testes; Kyle Orwig (University of Pittsburgh) for rhesus macaque testes; Paul Albosta and Cindy Tipping for guidance with *T. ni* and *D. melanogaster* dissection; Leemor Joshua-Tor (Cold Spring Harbor Laboratory) for sharing data prior to publication; Tiffanie Gardner for help formatting the manuscript; the UMASS Flow Cytometry and Mass Spectrometry cores; and members of the Zamore lab for discussions and comments on the manuscript. This work was supported in part by NIGMS grants R37 GM062862 and R35 GM136275 (P.D.Z.) and R35 GM127090 (I.J.M) and JSPS KAKENHI grants 18H05271 (Y.T.) and 19K06484 (N.I.).

Author contributions A.A., P.D.Z., I.J.M., and Y.T. conceived and designed the experiments. A.A., S.M.B., N.I., T.A.A., D.M.O., C.A., performed the experiments. D.M.O. and I.G. provided and analyzed the sequencing data. P.D.Z. supervised the research. A.A. and P.D.Z. wrote the manuscript. All the authors discussed the results and approved the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at .

Correspondence and requests for materials should be addressed to P.D.Z.

Figure legends

Figure 1 | A component of mouse testis lysate potentiates piRNA-directed target RNA cleavage by MIWI.

a, Strategy for programming recombinant MIWI with synthetic piRNA. **b,** Top**:** representative denaturing polyacrylamide gel image for target RNA cleavage by MIWI piRISC with and without added testis lysate. Bottom: product formed as a function of time by MIWI piRISC (mean ± SD, *n* =3). Rate constants were determined by fitting the data to the burst-and-steady-state equation (Equation 1 in Methods). $[E]_{active}$: the apparent concentration of active MIWI piRISC estimated from data fitting; k_{burst}: the pre-steady-state rate; k_{ss}: the steady-state rate. **c,** Target RNA cleavage by MIWI piRISC in the presence of lysate from either whole testis or FACS-purified germ cells. Spg, spermatogonia; PS, pachytene spermatocytes; DS, diplotene spermatocytes; SpII, secondary spermatocytes; Sptd, spermatids.

Figure 2 | Testis protein GTSF1 potentiates target RNA cleavage by MIWI piRISC.

a, Scheme to purify the MIWI-potentiating factor from mouse testis lysate. **b,** Target cleavage assay to estimate the apparent molecular weight of the MIWI-potentiating activity by size-exclusion chromatography (SEC). Arrowheads indicate the peak concentration of the molecular weight standards, and their peak elution volumes. V_0: void volume. **c,** The properties of the MIWI-potentiating activity compared to those of GTSF1. **d,** Strategy for creating knock-in mouse expressing 3XFLAG-HA-GTSF1 (*Gtsf1*$^{KI/KI}$). Arrows represent sequencing primers used to validate the insertion. The exon encodes an incomplete Isoleucine codon, which is completed upon splicing. **e,** Western Blotting to show immuno-depletion of 3XFLAG-HA-GTSF1 from secondary spermatocyte lysate using anti-FLAG paramagnetic beads. I, Input; S, Supernatant; E, 3XFLAG peptide eluate. **f,** Product formed by MIWI piRISC in the presence of the

indicated components. Abbreviations as in **e**. Numbers below the gel indicate relative product formed in each condition.

Figure 3 | GTSF1 paralogs can distinguish between MIWI and MILI.

a, Representative denaturing polyacrylamide gel image showing that purified GTSF1 recapitulates the effect of testis lysate on MIWI catalysis. **b,** Product generated as a function of time (mean ± SD, *n* = 3). Data were fit to the burst-and-steady-state equation. Data for MIWI RISC alone are from Fig. 1. **c,** Product formed as a function of time by MIWI programmed with *pi6* and *pi9* piRNAs for two independent trials. The mean values of the two trials were fit to the burst-and-steady-state equation. **d**, **e**, Representative denaturing polyacrylamide gel images of the assay to test GTSF1 mutants and paralogs in target cleavage by MIWI (**d**) or MILI (**e**) piRISC. **f,** Product formed as a function of time (mean ± SD, *n* = 3) fit to the burst-and-steady-state equation. **g,** Observed pre-steady-state first order rate constants of target cleavage by MIWI piRISC in the presence of either wild-type or PIWI-interacting mutant GTSF1^{W98A,W107A,W112A} were plotted as a function of GTSF1 concentration. **h,** Single-turnover rate of target cleavage by EfPiwi piRISC in the presence or absence of EmGtsf1 or yeast tRNA (mean ± SD, *n* = 3).

Figure 4 | Cleavage by MIWI piRISC is sensitive to the extent of complementarity between the piRNA and target.

a, MIWI piRISC target cleavage in the presence of GTSF1 for targets with increasing complementarity to synthetic piRNA guide. Top: multiple-turnover conditions; bottom: single-turnover conditions. All reactions contained saturating amounts of GTSF1. **b,** Target cleavage assay using targets complementary to piRNA guide nucleotides g2–g16 or g2–g30 and MIWI or MILI loaded with piRNAs of the indicated lengths, with or without GTSF1 (mean ± SD, *n* = 3). **c,** Absolute (mean ± SD, *n* = 3) and relative pre-

steady-state and steady-state rates of cleavage of the g2–g30 target by MIWI loaded with piRNAs of the indicated lengths in the presence of GTSF1.

Figure 5 | A model for the function of guide length and GTSF1 in target cleavage by PIWI proteins.

a, The energy of base pairing, estimated by standard nearest-neighbor methods vs. pre- and steady-state rates of GTSF1-potentiated target cleavage (mean ± SD, *n* = 3), directed by piRNAs of different lengths loaded into MIWI. The same target, which was fully complementary to each piRNA, was used in all experiments. **b,** piRNA-directed, MIWI-catalyzed target cleavage is envisioned to require two sequential conformational changes: (1) a target-dependent conformational change in piRISC (E_{PAZ}) in which the piRNA 3′ end leaves the PAZ domain, allowing extensive base pairing of the piRNA with the target substrate (S); and (2) GTSF1-dependent conversion of this piRISC pre-catalytic state (E_C) to the fully competent catalytic state ($E_{C'}$). P: cleaved target products. **c,** Proposed effects of different piRNA lengths, extents of guide:target complementarity, and GTSF1 on the forward and reverse rates of the two conformational rearrangements. The wide, central cleft like that observed in the cryo-EM structure of *Ephydatia fluviatilis* Piwi-A[64] is envisioned to allow the central region of a 30 nt piRNA to be mobile and exposed to solvent when its 3′ end is secured to the PAZ domain (upper right).

Arif et al., Figure 1