Sexual Dimorphism and Evolutionary Innovation in piRNA-Guided Genome Defense

Thesis by

Peiwei Chen

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2024

(Defended May 23, 2024)

Peiwei Chen
ORCID: 0000-0001-7160-6673

ACKNOWLEDGEMENTS

I cannot pinpoint when or why I decided to pursue biology, but I also never felt that I had to pick. I remember I spent a lot of time on biology in high school, but I was not particularly good at it. Having grown up in Mainland China, I went to Hong Kong for college, where everything was taught in English, so my freshman biology class was basically an English class. I left several exam questions blank—because I did not understand the questions—and I got a B. Somehow that did not discourage me, and miraculously I got into a special program designed for people wanting to pursue a research career. I did research throughout undergrad, got into Caltech, and somehow, I am now finishing my PhD in Biology. Of course, such a miracle is only possible because of the people below.

I want to first thank my PhD advisor, Alexei Aravin. I was the only person in my cohort who did not have a lab to join after the first year, so I decided to do a fourth rotation. If that rotation did not work out, I was going to leave Caltech, but luckily it did. Alexei shook my hand after I agreed to join the lab, and he let me work on the fly testis, while the lab and the field as a whole had primarily studied the fly ovary. I thank Alexei for the tremendous amount of intellectual freedom he gave me, and the so many afternoon conversations where we talked for hours and hours. Alexei is always my toughest "reviewer", so I thought if I could convince him I could convince almost anyone. I simply would not be the scientist I am today without his mentorship, guidance, and support. He discovered piRNAs during his PhD, so he set the bar very, very high.

I want to thank my undergrad mentors who set me up for a research career. Tom Cheung gave me my first ever independent research project in my sophomore year, and he treated me like a colleague even though I was a rookie. I was over the moon when I identified the target of a microRNA that controls muscle stem cell activation in his lab. I spent a summer at UCLA with Alex Hoffmann, who let me do a time-course ChIP-seq to look at the kinetics of RNA polymerase II transcribing immune response genes. When that experiment worked, I felt like I could do almost anything. I worked on neuronal senescence with Karl Herrup, who told the lab to "keep pushing off the frontier of science" every day before he left the lab. That sentiment stayed with me even today. I was only able to do so much research in undergrad because King Chow believed in me and admitted me into the IRE program, even though my freshman-year GPA was 3.45.

I want to thank many faculty at Caltech who supported me throughout grad school. I learned what a phylogenetic tree is from my thesis committee chair, Joe Parker, who always asked me the most

thoughtful and most thought-provoking questions. I fell in love with flies in Angela Stathopoulos's lab, and her excitement for science is infectious—I always enjoy talking to her about science and felt really lucky that I could stop by her office any time. I thank Marianne Bronner for being the most supportive committee member anyone could ever ask for, and I felt lucky that I could ask her anything and that she is always just an email away. I thank Ellen Rothenberg for her unparalleled passion for science and her lucidity in explaining complex ideas and articulating deep insights. I truly learned a lot from her. I also thank Bruce Hay, Sarkis Mazmanian, Kata Fejes Tóth, Mitch Guttman, Rebecca Voorhees, Justin Bois, and others who have taught me different things and helped me along the way.

I want to thank my lab mates and friends. Yicheng Luo taught me almost everything about fly genetics. Maria Ninova taught me almost everything I know about bioinformatics. Baira Godneeva, Lena Fefelova, Andres Rodriguez, and Anton Kuzmenko were fantastic lab mates. Elena Udartseva was the best lab manager. I thank the undergrads who have worked with me—Katherine Pan, Eunice Park, and Suchitra Dara. I have wonderful friends—Juan Bravo, Reem Abdel-Haq, James McGehee, Frank Macabenta, Yuelin Shi, Xinhong Chen, Shumao Zhang, Cai Tong Ng, Ailin He, Aditi Narayanan, Tom Naragon, Lev Tsypin, and so many others—who make my time in California memorable. I thank Resty and Cesar for their friendship.

I would not have become the scientist I am today without the broader scientific community outside of Caltech. I thank the following biologists whom I deeply admire for their help and inspirations: Yukiko Yamashita, Grace Yuh Chwen Lee, Felipe Karam Teixeira, Jim Kennison, Phil Zamore, Amanda Larracuente, Harmit Malik, Mia Levine, Zhao Zhang, Mark Van Doren, Ching-Ho Chang and many others. I learned most of what I know about fly testis from reading Yukiko's papers, and she is both a fantastic scientist and an extremely supportive colleague.

I thank Liz Jackman and Kitty Cahalan for their help in making the LEAP project a reality. I had the fortune to turn an idea into an outreach program at Caltech, and they are the heroes behind it.

Lastly, I thank my parents for their sacrifice, love, and support. It was until I was much older did I realize my mother is my personal hero—she specialized in pediatric oncology in the 90s in a small Chinese city, and now she saves nine out of ten kids with leukemia. I like to think that she is the reason why I became a biologist. My father did everything he could to make sure that I grew up without worrying about anything. I am only able to be the first person in my family to ever set foot out of Mainland China to pursue my dreams because of his sacrifice. I hope I make them proud.

ABSTRACT

The genome is a battleground, where different genetic elements vie for inheritance. In particular, selfish genetic elements enhance their own transmission at the expense of host fitness, causing intragenomic conflicts that must be resolved to protect host reproduction. To keep selfish genes in check, animals employ several genome defense mechanisms, including the PIWI-interacting RNA (piRNA) pathway, where small non-coding piRNAs guide PIWI proteins to find complementary RNAs for silencing. While selfish genes are found across the tree of life, they are often sexually dimorphic and lineage-specific. Yet, it remains poorly understood how sex- and lineage-specific selfish genes are tamed by conserved genome defense mechanisms. To address this, I used the piRNA pathway in *Drosophila melanogaster* as the model system to study sexual dimorphism and evolutionary innovation in genome defense.

In this thesis, I first described my discovery of piRNA sexual dimorphism, which evolved in response to the sex-specific selfish gene landscape. Next, I dissected the genetic basis and molecular mechanisms that underpin piRNA sexual dimorphism, gaining mechanistic insights into how the biological sex modifies the piRNA pathway to tame distinct selfish genes in two sexes. Pivoting to evolutionary innovation, I discovered a novel piRNA locus on the Y chromosome, which I named *petrel*, that silences the expression of an X-linked host gene, which I named *pirate*, implicating piRNAs in resolving X-versus-Y sex chromosome conflicts. *petrel* piRNAs evolved very recently after the split of *D. melanogaster* from its sibling species, highlighting a recent piRNA innovation against a lineage-specific target. Finally, I described my discovery of a novel genome defense protein factor, which I named Trailblazer, that tames a radically expanded selfish gene family, *Stellate*. Mechanistically, Trailblazer is a germline transcription factor that, via recent innovation of its DNA-binding domain, up-regulates the expression of two piRNA pathway effectors to quantitatively match *Stellate* in abundance, indicating a new mode of defense innovation beyond target-specific repressors.

Collectively, my thesis shows that the genomic battleground against selfish genes differs substantially between sexes and across lineages, which selects for distinct innovations in the piRNA pathway to control different selfish genes, thereby safeguarding genome integrity, animal fertility, and species continuity.

PUBLISHED CONTENT AND CONTRIBUTIONS

Chapter II

P. Chen, A. A. Aravin (2024). Rapid evolution of the piRNA pathway. *Invited review for Mol Biol Evol, in preparation*

P.C. conceived the project, made all figures, and wrote the manuscript.

Chapter III

P. Chen, A. A. Kotov, B. K. Godneeva, S. S. Bazylev, L. V. Olenina, A. A. Aravin (2021). piRNA-mediated gene regulation and adaptation to sex-specific transposon expression in D. melanogaster male germline. *Genes Dev* **35**, 914–935. https://doi.org/10.1101/gad.345041.120

P.C. conceived the project, did most experiments and analysis, made all figures, and wrote the manuscript.

Chapter IV

P. Chen, Y. Luo, A. A. Aravin (2021). RDC complex executes a dynamic piRNA program during Drosophila spermatogenesis to safeguard male fertility. *PLoS Genet* **17**, e1009591. https://doi.org/10.1371/journal.pgen.1009591

P.C. conceived the project, did most experiments and analysis, made all figures, and wrote the manuscript.

Chapter V

P. Chen, A. A. Aravin (2023). Genetic control of a sex-specific piRNA program. *Curr Biol* **33**, 1825–1835.e3. https://doi.org/10.1016/j.cub.2023.03.057

P.C. conceived the project, did all experiments and analysis, made all figures, and wrote the manuscript.

Chapter VI

P. Chen, K. C. Pan, E. H. Park, Y. Luo, Y. C. G. Lee, A. A. Aravin (2024). Adaptive piRNA pathway tuning tames recently expanded selfish genes. *submitted*

P.C. conceived the project, did most experiments and analysis, made all figures, and wrote the manuscript.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

Counter to our intuition and perhaps far from a harmonious place, the genome is a battleground, where every bit of DNA has to fight for inheritance and evolutionary survival. While most genes compete for inheritance by being "good citizens" of the genome—i.e., they help the host organism better survive and better reproduce, "hoping" that they would be inherited—selfish genes, on the other hand, cheat the entire process by enhancing their own propagation at the expense of the hosts. This results in the so-called intragenomic conflicts between selfish genes and the rest of the genome, which are most severe in the germline, so they must be resolved to protect host reproduction. To keep selfish genes in check, animals employ several genome defense mechanisms, including the PIWI-interacting RNA (piRNA) pathway, which is the focus of my thesis work.

At the core of the piRNA pathway, a small non-coding RNA, piRNA, guides the PIWI protein to find complementary targets for silencing. In this context, the piRNA provides the target specificity based on nucleotide sequence complementarity, whereas the PIWI protein silences the target identified by its piRNA guide, e.g., via cleaving the target RNA. Even though these core components of the piRNA pathway and their binary architecture are conceptually simple, over 30 proteins have been identified to be dedicated to this single pathway in the fruit fly, *Drosophila melanogaster*, alone, suggesting substantial elaborations of, and innovations in, the genome-defending piRNA pathway during metazoan evolution. However, the underlying driving force for such complexity remains largely unknown.

In addition, most prior studies of the piRNA pathway, in any animal, have been focused on one of the two sexes, leaving the piRNA biology virtually unexplored in the other sex. Given that genome defense against selfish genes is most critical in the heritable genome in the germ cell—arguably the most sexually dimorphic cell type—the piRNA pathway is very likely to differ between sexes. Thus, to what extent, how, and why the piRNA pathway is sexually dimorphic represent major gaps in our current knowledge of genome defense.

In this thesis, I sought to address two key questions of piRNA-guided genome defense: sexual dimorphism and evolutionary innovation.

Thesis overview

In Chapter II, I described the makeup of the piRNA pathway and highlighted its diversity and rapid evolution across metazoans. This is a review article invited by *Molecular Biology and Evolution*, currently in preparation.

In Chapter III, I described the first systematic definition of piRNAs in *D. melanogaster* testis and substantial sexual dimorphism of the piRNA program in the fly gonads. I showed that the sex-specific piRNA program parallels the sexually dimorphic transposon landscape, suggesting an adaptation of the piRNA program to distinct transposon threats in two sexes. In addition, I described the discovery of a new piRNA locus on the Y chromosome, *petrel*, which evolved very recently to control the expression of an X-linked gene, *pirate*, implicating the piRNA pathway in resolving X-versus-Y sex chromosome conflicts. This also assigned a novel function to the *D. melanogaster* Y chromosome,

which has been studied for over 100 years. Intriguingly, a different class of small RNAs, siRNAs, evolved in a sibling species to control the orthologous *pirate* gene, suggesting phenotypic convergence where two distinct small RNA pathways were recruited to tame a conserved gene in <3 million years of evolution. Chapter III is a research article published in 2021 in *Genes & Development.*

In Chapter IV, I studied a protein complex called the Rhino-Deadlock-Cutoff (RDC) complex, which licenses the non-canonical transcription of piRNA clusters in the *Drosophila* germline. It was thought that the RDC complex was female-specific, raising the question of how piRNA clusters are transcribed in the male counterpart. Challenging this dogma, I showed that the RDC complex operates in the male germline but exhibits several prominent sex differences. In particular, RDC differentially binds different piRNA clusters to mount a sex-specific piRNA program in each of the two sexes. Furthermore, RDC shows unique spatio-temporal dynamics during early spermatogenesis, activating different piRNA clusters at different stages of male germline development. Chapter IV is a research article published in 2021 in *PLoS Genetics*.

In Chapter V, I followed up with my discovery of piRNA sexual dimorphism in *Drosophila* gonads and showed that most of it originates from the germline, rather than gonadal somatic cells. Building on this, I dissected the underlying genetic control mechanisms of germline piRNA sexual dimorphism. Specifically, I employed unique genetic tools in *Drosophila* to independently manipulate sex chromosome compositions and germline sex determination. I found that the Y chromosome is both necessary and sufficient to recapitulate some aspects of the male piRNA program—in fact, the Y chromosome can produce otherwise male-specific piRNAs in female cellular environments.

On the other hand, by manipulating the sex determination pathway, I showed that sex determination directly controls piRNA sexual dimorphism via the key switch protein, Sxl. The effect of cellular sexual identity on piRNA sexual dimorphism is mediated, in part, through two chromatin proteins: a histone reader protein, Phf7, and a zinc finger protein, Kipferl. Together, my work showed that the sex chromosome and cellular sexual identity control different facets of piRNA sexual dimorphism, and it is their collective action that sculpts the piRNA sexual dimorphism. Chapter V is a research article published in 2023 in *Current Biology*.

In Chapter VI, I conducted a targeted *in vivo* RNAi screen and discovered a new piRNA pathway factor that I named Trailblazer. Trailblazer specifically protects the male germline against a male-specific selfish gene family, *Stellate*, representing the first piRNA pathway machinery known to be specifically required for genome defense in male, but not female, flies. Mechanistically, Trailblazer is a germline transcription factor that up-regulates the expression of two piRNA pathway effectors— Aub and AGO3—thereby silencing the highly abundant *Stellate* transcripts produced from the ampliconic *Stellate* genes. The ability of Trailblazer to tune the piRNA pathway results from its recent innovation in its DNA-binding domain, as a close ortholog from a sibling species that diverged <3 million years ago is unable to substitute its function in *D. melanogaster*. Based on these findings, I proposed a new mode of evolutionary innovation in genome defense: quantitatively tuning the general, target-nonspecific defense machinery to match selfish genes in abundance. Given that copy-number expansion is a recurrent feature of diverse selfish genes found across the tree of life, I envision that this is likely a widespread phenomenon and a central strategy in evolutionary battles against the expanding selfish genes. Chapter VI is a research article recently submitted for journal publication.

C h a p t e r 2

RAPID EVOLUTION OF THE PIRNA PATHWAY

P. Chen, A. A. Aravin (2024).

Invited review for Mol Biol Evol; in preparation

Main Text

Introduction

In 1961, a peculiar observation was made under the microscope: needle- or star-shaped crystals accumulate in the spermatocyte nuclei of *Drosophila melanogaster* XO males that lack a Y chromosome . These crystals were only found in XO males, but never in XY. The locus responsible for the production of these crystals was mapped to the X chromosome and named "*Stellate*", which describes things arranged in radiating patterns, after the shape of crystals observed (*1*). Meanwhile, the Y-linked locus that suppresses the appearance of these Stellate crystals was named *crystal* (*cry*) and *Suppressor of Stellate* (*Su(Ste)*) by two different groups (*2*, *3*) (referred to as *Su(Ste)* hereafter). Both *Stellate* and *Su(Ste)* are ampliconic, tandem repeats. For over a decade, it was unclear how *Su(Ste)*—a Y-linked locus that shares sequence homology with the X-linked *Stellate* locus (*4*)—acts to suppress *Stellate*. Inspired by work on homology-dependent silencing via RNA interference (*5*–*8*), it was reported in 2001 that *Su(Ste)* produces short RNAs that are just slightly longer than 21-23nt small interfering RNA (siRNA), which seem to confer silencing (*9*). Five years later, a distinct class of small RNAs were named PIWI-interacting RNAs (piRNAs), which specifically bind PIWI-clade Argonaute proteins and thus differ from siRNA and microRNA that bind AGO-clade Argonaute proteins (*10*–*15*). *Su(Ste)* short RNAs, initially thought to be slightly longer, repeat-associated siRNAs (rasiRNAs), were then recognized as piRNAs (*15*, *16*).

The discovery of the piRNA pathway, where small non-coding piRNAs guide PIWI proteins to silence complementary RNAs, solved several decades-old riddles of transposon control in *Drosophila melanogaster*. The *flamenco* locus that silences the *gypsy* transposon turned out to encode piRNAs rather than proteins (*17*, *18*), the X telomere-associated

sequence (X-TAS) that controls the *P-elemen*t transposon was found to produce piRNAs (*18, 19*), and the maternally supplied materials that confer silencing of *P-element* and *I-element* transposons in the offspring were also piRNAs (*20–22*). In the 18 years following the discovery of piRNAs in 2006, piRNAs and their PIWI protein partners have been found in the vast majority of animals, but not in plants, fungi, or a few metazoan lineages (*23–26*). Because transposon-targeting piRNAs can be found in almost all animals, transposon control seems to be the deeply conserved function of piRNAs. However, the most striking feature of the piRNA pathway is its rapid evolution: novel piRNAs recurrently evolve to regulate new targets that differ among closely related species, the piRNA pathway frequently employs young and fast-evolving proteins for its essential function, and the overall piRNA pathway architecture markedly diverges across taxa. This trait of rapid evolution distinguishes the piRNA pathway from other small RNA pathways in animals, but how and why the piRNA pathway evolves so fast remain important puzzles in the field.

In this Review, we will first describe the makeup of the piRNA pathway in two focal species (flies and mice) and then take a comparative approach to examine its rapid evolution across metazoans. Next, we will discuss the mechanisms, driving forces, and consequences of rapid piRNA pathway evolution and highlight open questions in the field. While the conserved function of the piRNA pathway in transposon control already implies its ability to quickly change and adapt to the ever-changing transposon landscape, mechanistically how the piRNA pathway does so remain elusive. We further argue that the ubiquitous piRNAs against non-transposon targets, such as *Stellate* in flies (*9, 15, 16*) and those described in the initial discovery of the piRNA pathway in mammals (*10–13*), hint at additional forces driving piRNA pathway evolution. We propose that genetic conflicts,

including but far beyond the host-transposon conflict, underpin rapid piRNA pathway evolution.

Makeup of the piRNA pathway

Cutting: post-transcriptional silencing

At the core of the piRNA pathway, a small non-coding piRNA loads onto a PIWI protein and guides it to find complementary RNAs for silencing. Conceptually similar to AGO proteins loaded with siRNA guides, cytoplasmic PIWI proteins exert post-transcriptional silencing by cleaving the target RNAs that are complementary to piRNA guides (*18*, *27*–*29*) (Fig. 1a). While the exact targeting rules of piRNAs is still a subject of active research, PIWI-mediated cleavage can happen when there is high, but not necessarily perfect, sequence complementarity between piRNAs and targets (*30*), and such cleavage occurs at exactly 10 nt away from the piRNA 5' end, which is characteristic of Argonaute proteins (*31*). Cytoplasmic PIWI proteins are often found at the peri-nuclear cytoplasmic compartment called "nuage" (*18*, *32*), which contains the DEAD-box RNA helicase—Vasa in flies and MVH/DDX4 in mice—that defines a class of germline-specific granules across metazoans.

Rather than simply being degraded, the target RNA cleaved by PIWI can be turned into yet another piRNA. This new piRNA with the opposite polarity to the original piRNA can repeat the process and ultimately produce a piRNA similar to the original piRNA, resulting in the so-called "ping-pong cycle" that selectively amplifies piRNA species with abundant complementary targets, analogous to adaptive immunity (*18*, *27*). This was proposed based on the observation that complementary piRNAs are often found to have 10 nt 5'-to-5' overlap, so PIWI-mediated cleavage often also defines the 5' end of a new

piRNA. Interestingly, flies and mice both have multiple cytoplasmic PIWI proteins, among which piRNA ping-pong pairs are partitioned unevenly. In flies, two cytoplasmic PIWI proteins, Aub and Ago3, disproportionally load transposon-antisense and transposon-sense piRNAs, respectively, suggesting that the two PIWI proteins participate in sequential steps of the ping-pong cycle (*18*, *27*). Similarly, in mice, MIWI2/PIWIL4 and MILI/PIWIL2 ping-pong (*33*).

Cycles of ping-pong processing and, more generally, localizations and functions of PIWI proteins require dynamic regulation. Just three years following the discovery of the piRNA pathway, PIWI proteins were found to contain symmetrically dimethylated arginine (sDMA) (*34*–*38*), a post-translational modification bound by Tudor domain-containing proteins (TDRDs) (*39*). Since then, a set of TDRDs have been biochemically identified in PIWI interactome and genetically implicated in PIWI functions, including Vret/TDRD1, Papi/TDRD2 (a.k.a. TDRKH), Qin/TDRD4 (a.k.a. RNF17), Tej/TDRD5, Tud/TDRD6, Tapas/TDRD7, Spn-E/TDRD9, Krimp (a fly-specific TDRD), and TDRD12 which has three paralogs in flies: Yb, Bo(Yb), and So(Yb) (*40*–*47*). This set of TDRDs is thought to regulate the ping-pong cycle and other PIWI functions dynamically.

Without cutting: co-transcriptional silencing

While the piRNA pathway can silence targets post-transcriptionally in the cytoplasm—a function shared with metazoan siRNA and microRNA pathways—the piRNA pathway also uniquely has a second, nuclear arm that silences targets co-transcriptionally on chromatin (Fig. 1b). Instead of mediating an enzymatic cleavage of the nascent target RNA, the nuclear PIWI (Piwi in flies and MIWI2/PIWIL4 in mice) initiates heterochromatin formation at the target genomic locus. This is done by H3K9me3

deposition in both flies (*48–51*) and mice (*52*), as well as DNA methylation in mice (*33*, *53–55*) (note that flies are a unique animal group that lack DNA methylation altogether). PIWI, through a set of auxiliary proteins—Arx (*56–58*), Panx (*59*, *60*), and Nxf2 (*61–64*) in flies and SPOCD1 (*65*) and TEX15 (*66*, *67*) in mice, recruits histone lysine methyltransferase Egg/SETDB1 to deposit H3K9me3 (*59*, *60*) and DNA methyltransferases DNMT3C and their co-factor DNMT3L to methylate DNA (*33*, *55*, *68*). Beyond a dedicated set of proteins, protein SUMOylation, a post-translational modification, also emerged as a key regulator of piRNA-guided heterochromatin formation (*69–71*).

Of note, Arx that partners with the nuclear Piwi to exert co-transcriptional silencing is one of the four fly homologs of the conserved gametocyte-specific factor 1 (GTSF1) protein family (*56–58*), whose murine ortholog was recently shown to accelerate the target cleavage of the cytoplasmic PIWI proteins—MIWI/PIWIL1 and MILI/PIWIL2—in mice (*72*). Additionally, the two GTSF1 paralogs in silkworm were shown to orthogonally activate two cytoplasmic PIWI proteins for target cleavage (*73*). These recent findings suggest that different GTSF1 paralogs likely partner with different PIWI proteins for their functions, by either post-transcriptionally enhancing target cleavage (as shown in mice) or co-transcriptionally facilitating heterochromatin formation (as shown in flies).

Before chopping: transcription and export of piRNA precursors

piRNAs guide the PIWI proteins to silence targets, but how are piRNAs made? Unlike microRNAs and proteins, which are encoded in individual genes, piRNAs are encoded in extended genomic loci termed “piRNA clusters”, each of which encodes many piRNA species (*10*, *11*, *18*). piRNA precursors are transcribed from the piRNA clusters by RNA

polymerase II as single-stranded long RNAs, which, following nuclear export, are "chopped" into short piRNAs in the cytoplasm.

piRNA clusters differ between species and even across cell types and developmental stages within a species. In flies, the piRNA pathway is active in germ cells and gonadal somatic cells that surround and support germ cells (*18*). Both cell types have the so-called "uni-strand piRNA clusters", whose transcription happens from one DNA strand as canonical transcription, but germline also has "dual-strand piRNA clusters", from which piRNA precursors are transcribed from both DNA strands (Fig. 1c). Dual-strand piRNA clusters bend most conventional rules of gene expression: they lack defined promoters, are transcribed without splicing and polyadenylation, and are embedded in H3K9me3-decorated heterochromatin. Such non-canonical transcription of dual-strand piRNA clusters is enabled by a fly-specific tripartite protein complex, Rhi-Del-Cuff (RDC) complex (*74–78*). RDC biology is mediated, in part, through a set of specialized protein machinery: the zinc-finger protein Kipf recruits RDC to certain piRNA clusters (*79*); for many piRNA clusters, RDC initiates transcription via a specialized TFIIA-L paralog Moon (*80*) and enables the nuclear export by recruiting Boot and Nxf3 (*81, 82*). On the other hand, uni-strand piRNA clusters behave mostly like normal protein-coding genes, though two nucleoporins, Nup54 and Nup58, appear particularly important for their nuclear export (*83*).

In mice, piRNAs and their genomic origins differ substantially between developmental stages. In prenatal/fetal (E16.5) and neonatal (on/before P10) mouse testes, where germ cells have not yet reached the pachytene stage and their piRNAs are called "pre-pachytene", uni- and dual-strand piRNA clusters similar to those in flies can be found;

however, individual transposons and mRNAs (in particular, their 3' UTR) seem to be major sources of fetal and neonatal piRNAs, respectively (*33*, *53*). Meanwhile, in adult testis (on/after P14), most piRNAs—referred to as the "pachytene piRNAs", as germ cells containing them are at the pachytene stage—are encoded in divergently transcribed loci (*10*, *11*), where the transcription factor A-MYB transcribes many piRNA precursors (*84*) from one DNA strand on one side of the promoter and the other DNA strand on the other side of the promoter. Overall, mechanisms by which piRNA precursors are made appear to the most diverged aspects of the piRNA pathway between flies and mice.

Chopping: phased fragmentation of piRNA precursors into piRNAs

Once single-stranded piRNA precursors have been exported to the cytoplasm, they need to be cut into short piRNAs that can load onto PIWI proteins and guide them to complementary RNAs for targeting. A key enzyme that generates the 5' ends of piRNAs is called Zuc in flies and PLD6 (or MitoPLD) in mice, which is anchored at the mitochondria outer membrane (*85–89*). Though homologous to phospholipases, Zuc/PLD6 shows nuclease, rather than phospholipase, activities (*88*, *89*). Once Zuc/PLD6 has made a cut on piRNA precursors, it often makes another cut(s) downstream of the previous cut (Fig. 1d), leading to phased production of serial piRNAs from one transcript (*90*, *91*). As opposed to the ping-pong cycle that increases the abundance of piRNAs, Zuc/PLD6-mediated phasing is thought to diversify piRNA species, converting sequences downstream of a piRNA target into potentially novel piRNAs. For unknown reasons, the outer mitochondrial membrane (OMM) seems to be a key processing site for piRNAs, housing Zuc/PLD6 and other proteins critical for piRNA production: Mino/GPAT2 (*92*), Gasz/GASZ (*93–96*), and a fly-specific protein, Daed (*96*).

In flies, the Zuc-mediated cut generates not just the 5’ end of a piRNA but oftentimes also the 3’ end of a neighboring piRNA (*90, 91*). In mice, however, the 3’ end generated by the PLD6 cut has to be trimmed by PNLDC1 (*97–99*) to make mature piRNAs, and this process involves TDRD2/TDRKH (*90, 91*). Once the piRNAs of correct lengths are made, they are 2’-*O*-methylated by Hen1/HENMT1 (*100–103*) and loaded onto PIWI proteins by chaperone Hsp90/HSP90 and co-chaperone Shu/FKBP6 (*104*, *105*).

How does Zuc/PLD6 select the correct RNAs for piRNA biogenesis? In fly germ cells, PIWI cuts (in particular, Ago3 cuts) in nuage are often upstream of Zuc cuts at the mitochondrial surface (*90, 91*). In other words, RNAs cleaved by PIWI proteins in nuage are identified as piRNA precursors and brought to mitochondria for phased piRNA biogenesis, which seems to be facilitated by the helicase Armi/MOV10L1. Blocking the PIWI or Zuc cuts enriches Armi in nuage and mitochondria, respectively, suggesting that Armi shuttles between two subcellular compartments (*106*, *107*). In the fly gonadal somatic cells and mouse germ cells, however, the source of specificity for Zuc/PLD6-mediated piRNA biogenesis are still unknown, although Armi and Yb (a fly-specific TDRD12 protein) are known to be involved in defining the input RNAs for Zuc in the gonadal soma of flies (*108, 109*). Interestingly, artificially tethering an unrelated RNA to Armi or Yb is sufficient to trigger piRNA production from that RNA (*110, 111*), supporting a role of Armi and Yb in selecting piRNA precursors among cellular RNAs for processing.

The complexity

Even though the core architecture of the piRNA pathway—short piRNAs guide PIWI proteins to find targets for silencing—is conceptually simple, over 30 proteins have been

described to be dedicated to this pathway, many of which are lineage-specific proteins. While the functions of some proteins remain obscure, e.g., the conserved protein Mael/MAEL (*32*, *48*, *112–114*) and the fly-specific Squ (*85*, *86*), four modules of the piRNA pathway have merged (Fig. 1). Nonetheless, distinctions between the biogenesis and functions of piRNAs have become increasingly blurry. For example, post-transcriptional silencing also produces piRNAs via the ping-pong cycle and feeds into phased piRNA biogenesis at the mitochondria surface; co-transcriptional silencing also sets up the chromatin environment for certain piRNA precursor transcription; and phased piRNA biogenesis is arguably also a type of post-transcriptional silencing. Because of such interconnectivities, as well as difficulties to reconstitute the complex piRNA pathway *in vitro*, the molecular events initiating different parts of the piRNA pathway are still unclear, and spatiotemporally ordering different processes within the pathway remains nontrivial.

Diversity and rapid evolution of the piRNA pathway

The complexity of the piRNA pathway is further exacerbated by its overwhelming diversity and rapid evolution across metazoans. We describe below the diversity and rapid evolution of i) non-coding piRNAs, ii) protein machinery of the piRNA pathway, and iii) their cell type ranges of operation in different species.

Non-coding piRNAs

piRNAs against transposons

As the sources of specificity, piRNA species dictate the targets and functions of the piRNA pathway. In almost all the animal species examined to date, piRNAs can be found in the germline (or cells that harbor the “heritable genome” in asexual lineages), and many

of them target transposons (Fig. 2). This observation suggests that the ancient function of the piRNA pathway is transposon control in heritable genome, like germline genome in sexually reproducing animals and some somatic stem cell genome in asexual lineages. As transposon sequence landscape differs among closely related species, sequences of transposon-targeting piRNAs are poorly conserved. Even when syntenic, piRNA-producing loci can be found between related species, the underlying sequences diverge (*10*, *11*, *115–119*), let alone the complete lack of conservation in more distantly related species. Thus, unlike microRNAs that are conserved across long evolutionary spans, piRNAs often evolve fast and diverge across species.

piRNAs against host genes

Outside transposon control, piRNAs can participate in gene regulation. In the initial description of piRNAs in adult mammalian testes, pachytene piRNAs were found to be depleted of transposon sequences and they were diverse genome-unique sequences that map to almost only themselves (*10*, *11*). Because of the poorly defined targeting rules of piRNAs, it was not until recently that two pachytene piRNA loci were experimentally demonstrated to each regulate a few host genes during mouse spermiogenesis (*120*, *121*). Nonetheless, pachytene piRNAs evolve fast and their sequences are poorly conserved (*117*). Although flies only express piRNAs before meiosis and therefore lack pachytene piRNAs found in mammals, the first described piRNA species—*Su(Ste)* piRNAs—regulate the expression of *Stellate* genes in the male germline of flies (*9*, *15*). Subsequently, another piRNA locus called *AT-chX* was found to produce piRNAs that target the *vasa* gene in the fly male germline (*16*), and, more recently, yet another piRNA locus called *petrel* was discovered to regulate the *pirate* gene (*122*). Notably, all three gene-regulating piRNAs were only found in *Drosophila melanogaster*, but not in its

sibling species, highlighting recent piRNA innovations. Meanwhile, in the silkworm *Bombyx mori*, the W-linked *Feminizer* piRNA silences the Z-linked *Masculinizer* gene, thereby determining the femaleness of ZW animals (*123*). While the *Masculinizer* gene can be found in most Lepidopteran species, the *Feminizer* piRNA seems to be a *Bombyx* innovation to silence *Masculinizer* (*124–127*). Yet, in a distantly related species, the diamondback moth *Plutella xylostella*, distinct piRNA species arose on the W chromosome to silence the *Masculinizer* gene, suggesting convergent evolution (*128*). Interestingly, in *C. elegans* and *C. briggsae*, different piRNA species regulate the expression of the orthologous *xol-1* gene—a key regulator of sex determination and dosage compensation (*129*). Finally, in other invertebrates such as mosquitos, locusts, and planarians, gene-regulating piRNAs have been identified in the soma (*130–132*). Together, contrary to the deeply conserved role of piRNAs in transposon control, gene regulations seems to be lineage-specific innovations of piRNAs at the tips of the animal tree (Fig. 2).

piRNAs against (or from) viruses and satDNAs

In addition to transposon control and gene regulation, piRNAs have taken on the antiviral function in a few phylogenetically isolated species—mosquitos (*133*, *134*), chicken (*135*), and koala (*136*)—but not in a broader range of animals. Furthermore, piRNAs are produced from two phylogenetically unrelated satDNAs in mosquitos (*131*) and flies (*122*, *137*, *138*), respectively, further adding to the diversity of piRNAs. These findings emphasize the lineage-specific involvement of piRNAs in virus and satDNA biology (Fig. 2).

Key features of piRNAs: diverse sequences, varying abundance, and autoimmune constraints

A key feature of piRNAs seen across the animal tree is their sequence diversity. piRNAs take up a vast sequence space and, with just a few mismatches, they can map to a large number of cellular transcripts. Coupled with the lack of evolutionary conservation in piRNA sequences, a central challenge in piRNA biology is to identify *bona fide* targets among diverse sequences that piRNAs can map to through imperfect complementarity. While piRNA sequences are so diverse, their abundance varies substantially, with an overall positive correlation between the abundance of piRNAs and their regulatory impacts (*139*). This is in part because the ping-pong cycle is thought to selectively amplify piRNAs that recognize actively transcribed target RNAs, rendering their abundance higher than other piRNAs that do not actively engage with targets (*18*, *27*). Thus, abundant piRNAs are more likely to be active regulators of target expression, whereas those of low abundance probably do not have a regulatory function. This also reflects the "autoimmune" constraints faced by the piRNA pathway: vastly diverse piRNAs have the potential to, but must not, silence essential host genes (*140*). Consequently, as diverse piRNAs evolve to tame changing transposon landscape and acquire lineage-specific new targets, the abundance of many piRNAs is gated to minimize autoimmunity.

Protein machinery of the piRNA pathway

PIWI proteins: diverse subcellular localization, sex specificity, and cell type specificity

As implied by their names—PIWI-interacting RNAs—piRNAs are defined by their interaction with PIWI proteins, and they exert regulatory functions through their PIWI protein partners. The number of PIWI proteins differ drastically across metazoans (Fig. 2). Hydra, silkworm, zebrafish, and chicken only have two cytoplasmic PIWI proteins (*135*, *141–144*), likely reflecting an ancestral state. Meanwhile, planarian, fruit fly, and mammals have a third, nuclear PIWI protein—SMEDWI-2, Piwi, and PIWIL4,

respectively—but they are not direct orthologs with each other (*145–147*), suggesting at least three independent evolution of nuclear PIWI proteins. In many mammals including cattle, golden hamsters, and humans—with the notable exception of rodents—there is a fourth PIWI protein, PIWIL3 (*148–153*), that is cytoplasmic. Intriguingly, PIWIL3 is ovary-specific whereas PIWIL4 is testis-specific, implicating PIWI protein specialization in sex-specific piRNA biology. Remarkably, mosquitos have an expanded PIWI protein family of seven members—one (Piwi7) is embryo-specific, one (Piwi3) is ovary-specific, one (Piwi2) is germline-specific, and the remaining four are ubiquitously expressed (*154*). This level of protein family expansion and cell type specificity imply substantial specialization of different PIWI proteins in mosquitos. Lineage-specific PIWI duplication is not unique to mosquitos. For example, besides XILI/PIWIL2 and XIWI2/PIWIL4, frogs have two PIWIL1 paralogs: XIWI1a/PIWIL1a and XIWI1b/PIWIL1b (*155*). Lastly, nematode is a unique animal clade where many lineages have lost PIWI genes and the piRNA pathway altogether (*24*), and *C. elegans* has only one PIWI protein (*156–158*), whose functions and associated piRNAs differ considerably from most other animal clades (*159*). Collectively, PIWI proteins exhibit striking diversity across animals.

Subcellular topologies, functions, and actions

Subcellular localizations of PIWI proteins profoundly impact the piRNA pathway by determining how it silences targets in different species. In flies and mice, piRNAs can load onto cytoplasmic and nuclear PIWI proteins to silence targets post-transcriptionally in the cytoplasm and co-transcriptionally in the nucleus, respectively (*33*, *48*). However, many other taxa lack nuclear PIWI proteins and therefore the nuclear arm of the piRNA pathway altogether, exclusively silencing targets post-transcriptionally in the cytoplasm. On the other hand, the gonadal somatic cells in flies have co-opted the nuclear, but not

cytoplasmic, arm of the piRNA pathway to only silence targets co-transcriptionally in the nucleus (*115*, *160*). Thus, across different cell types in different species, the piRNA pathway can adopt one of the three subcellular topologies: nuclear only, cytoplasmic only, or both nuclear and cytoplasmic. These distinct subcellular topologies translate to different functions and modes of actions of the piRNA pathway.

The diversity and rapid evolution of the piRNA pathway machinery can also be seen in TDRDs and GTSFs that interact with PIWI proteins, sets of protein machinery that drive piRNA cluster expression, as well as those at the mitochondria surface facilitating piRNA biogenesis. Furthermore, at an even higher level, the cell type range of piRNA pathway operation also exhibits striking diversity. These are not described in detail here but will be discussed in the final manuscript submitted for journal publication.

References

1. R. W. Hardy, Crystal aggregates in the primary spermatocytes of XO males in Drosophila melanogaster. *Drosophila Inform. Serv.* **55**, 54–55 (1980).

2. S. Pimpinelli, S. Bonaccorsi, M. Gatti, L. Sandler, The peculiar genetic organization of Drosophila heterochromatin. *Trends in Genetics* **2**, 17–20 (1986).

3. K. J. Livak, Detailed structure of the Drosophila melanogaster stellate genes and their transcripts. *Genetics* **124**, 303–316 (1990).

4. K. J. Livak, Organization and mapping of a sequence on the Drosophila melanogaster X and Y chromosomes that is transcribed during spermatogenesis. *Genetics* **107**, 611–634 (1984).

5. A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, C. C. Mello, Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–811 (1998).

6. A. J. Hamilton, D. C. Baulcombe, A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950–952 (1999).

7. S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon, An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**, 293–296 (2000).

8. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25–33 (2000).

9. A. A. Aravin, N. M. Naumova, A. V. Tulin, V. V. Vagin, Y. M. Rozovsky, V. A. Gvozdev, Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Current Biology* **11**, 1017–1027 (2001).

10. A. Aravin, D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino, P. Morris, M. J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J. J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan, T. Tuschl, A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207 (2006).

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

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

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

14. K. Saito, K. M. Nishida, T. Mori, Y. Kawamura, K. Miyoshi, T. Nagami, H. Siomi, M. C. Siomi, Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev.* **20**, 2214–2222 (2006).

15. V. V. Vagin, A. Sigova, C. Li, H. Seitz, V. Gvozdev, P. D. Zamore, A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).

16. K. M. Nishida, K. Saito, T. Mori, Y. Kawamura, T. Nagami-Okada, S. Inagaki, H. Siomi, M. C. Siomi, Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad. *RNA* **13**, 1911–1922 (2007).

17. A. Pélisson, S. U. Song, N. Prud'homme, P. A. Smith, A. Bucheton, V. G. Corces, Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. *EMBO J* **13**, 4401–4411 (1994).

18. J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G. J. Hannon, Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128**, 1089–1103 (2007).

19. S. Ronsseray, M. Lehmann, D. Anxolabéhère, The maternally inherited regulation of P elements in Drosophila melanogaster can be elicited by two P copies at cytological site 1A on the X chromosome. *Genetics* **129**, 501–512 (1991).

20. G. Picard, Non-mendelian female sterility in Drosophila melanogaster: hereditary transmission of I factor. *Genetics* **83**, 107–123 (1976).

21. M. G. Kidwell, J. F. Kidwell, J. A. Sved, Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. *Genetics* **86**, 813–833 (1977).

22. J. Brennecke, C. D. Malone, A. A. Aravin, R. Sachidanandam, A. Stark, G. J. Hannon, An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**, 1387–1392 (2008).

23. A. Grimson, M. Srivastava, B. Fahey, B. J. Woodcroft, H. R. Chiang, N. King, B. M. Degnan, D. S. Rokhsar, D. P. Bartel, Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**, 1193–1197 (2008).

24. P. Sarkies, M. E. Selkirk, J. T. Jones, V. Blok, T. Boothby, B. Goldstein, B. Hanelt, A. Ardila-Garcia, N. M. Fast, P. M. Schiffer, C. Kraus, M. J. Taylor, G. Koutsovoulos, M. L. Blaxter, E. A. Miska, Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages. *PLoS Biol* **13**, e1002061 (2015).

25. M. Mondal, P. Klimov, A. S. Flynt, Rewired RNAi-mediated genome surveillance in house dust mites. *PLoS Genet* **14**, e1007183 (2018).

26. S. Fontenla, G. Rinaldi, J. F. Tort, Lost and Found: Piwi and Argonaute Pathways in Flatworms. *Front Cell Infect Microbiol* **11**, 653695 (2021).

27. L. S. Gunawardane, K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, M. C. Siomi, A slicer-mediated mechanism for repeat-associated siRNA 5’ end formation in Drosophila. *Science* **315**, 1587–1590 (2007).

28. S. De Fazio, N. Bartonicek, M. Di Giacomo, C. Abreu-Goodger, A. Sankar, C. Funaya, C. Antony, P. N. Moreira, A. J. Enright, D. O’Carroll, The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480**, 259–263 (2011).

29. M. Reuter, P. Berninger, S. Chuma, H. Shah, M. Hosokawa, C. Funaya, C. Antony, R. Sachidanandam, R. S. Pillai, Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* **480**, 264–267 (2011).

30. I. Gainetdinov, J. Vega-Badillo, K. Cecchini, A. Bagci, C. Colpan, D. De, S. Bailey, A. Arif, P.-H. Wu, I. J. MacRae, P. D. Zamore, Relaxed targeting rules help PIWI proteins silence transposons. *Nature* **619**, 394–402 (2023).

31. S. M. Elbashir, W. Lendeckel, T. Tuschl, RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* **15**, 188–200 (2001).

32. A. A. Aravin, G. W. van der Heijden, J. Castañeda, V. V. Vagin, G. J. Hannon, A. Bortvin, Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genet* **5**, e1000764 (2009).

33. A. A. Aravin, R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor, G. J. Hannon, A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**, 785–799 (2008).

34. C. Chen, J. Jin, D. A. James, M. A. Adams-Cioaba, J. G. Park, Y. Guo, E. Tenaglia, C. Xu, G. Gish, J. Min, T. Pawson, Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi. *Proc Natl Acad Sci U S A* **106**, 20336–20341 (2009).

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

36. K. M. Nishida, T. N. Okada, T. Kawamura, T. Mituyama, Y. Kawamura, S. Inagaki, H. Huang, D. Chen, T. Kodama, H. Siomi, M. C. Siomi, Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. *EMBO J* **28**, 3820–3831 (2009).

37. M. Reuter, S. Chuma, T. Tanaka, T. Franz, A. Stark, R. S. Pillai, 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 (2009).

38. V. V. Vagin, J. Wohlschlegel, J. Qu, Z. Jonsson, X. Huang, S. Chuma, A. Girard, R. Sachidanandam, G. J. Hannon, A. A. Aravin, 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).

39. J. Côté, S. Richard, Tudor domains bind symmetrical dimethylated arginines. *J Biol Chem* **280**, 28476–28483 (2005).

40. V. S. Patil, T. Kai, Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein Tejas. *Curr Biol* **20**, 724–730 (2010).

41. D. Handler, D. Olivieri, M. Novatchkova, F. S. Gruber, K. Meixner, K. Mechtler, A. Stark, R. Sachidanandam, J. Brennecke, A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J* **30**, 3977–3993 (2011).

42. H.-Y. Huang, S. Houwing, L. J. T. Kaaij, A. Meppelink, S. Redl, S. Gauci, H. Vos, B. W. Draper, C. B. Moens, B. M. Burgering, P. Ladurner, J. Krijgsveld, E. Berezikov, R. F. Ketting, Tdrd1 acts as a molecular scaffold for Piwi proteins and piRNA targets in zebrafish. *EMBO J* **30**, 3298–3308 (2011).

43. L. Liu, H. Qi, J. Wang, H. Lin, PAPI, a novel TUDOR-domain protein, complexes with AGO3, ME31B and TRAL in the nuage to silence transposition. *Development* **138**, 1863–1873 (2011).

44. H. Qi, T. Watanabe, H.-Y. Ku, N. Liu, M. Zhong, H. Lin, The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. *J Biol Chem* **286**, 3789–3797 (2011).

45. A. L. Zamparini, M. Y. Davis, C. D. Malone, E. Vieira, J. Zavadil, R. Sachidanandam, G. J. Hannon, R. Lehmann, Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. *Development* **138**, 4039–4050 (2011).

46. Z. Zhang, J. Xu, B. S. Koppetsch, J. Wang, C. Tipping, S. Ma, Z. Weng, W. E. Theurkauf, P. D. Zamore, Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. *Mol. Cell* **44**, 572–584 (2011).

47. A. Anand, T. Kai, The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in Drosophila. *EMBO J* **31**, 870–882 (2012).

48. G. Sienski, D. Dönertas, J. Brennecke, Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**, 964–980 (2012).

49. X. A. Huang, H. Yin, S. Sweeney, D. Raha, M. Snyder, H. Lin, A major epigenetic programming mechanism guided by piRNAs. *Dev Cell* **24**, 502–516 (2013).

50. A. Le Thomas, A. K. Rogers, A. Webster, G. K. Marinov, S. E. Liao, E. M. Perkins, J. K. Hur, A. A. Aravin, K. F. Tóth, Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* **27**, 390–399 (2013).

51. N. V. Rozhkov, M. Hammell, G. J. Hannon, Multiple roles for Piwi in silencing Drosophila transposons. *Genes Dev.* **27**, 400–412 (2013).

52. D. Pezic, S. A. Manakov, R. Sachidanandam, A. A. Aravin, piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev* **28**, 1410–1428 (2014).

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

54. M. A. Carmell, A. Girard, H. J. G. van de Kant, D. Bourc'his, T. H. Bestor, D. G. de Rooij, G. J. Hannon, MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* **12**, 503–514 (2007).

55. S. Kuramochi-Miyagawa, T. Watanabe, K. Gotoh, Y. Totoki, A. Toyoda, M. Ikawa, N. Asada, K. Kojima, Y. Yamaguchi, T. W. Ijiri, K. Hata, E. Li, Y. Matsuda, T. Kimura, M. Okabe, Y. Sakaki, H. Sasaki, T. Nakano, DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**, 908–917 (2008).

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

57. F. Muerdter, P. M. Guzzardo, J. Gillis, Y. Luo, Y. Yu, C. Chen, R. Fekete, G. J. Hannon, A Genome-wide RNAi Screen Draws a Genetic Framework for Transposon Control and Primary piRNA Biogenesis in Drosophila. *Molecular Cell* **50**, 736–748 (2013).

58. H. Ohtani, Y. W. Iwasaki, A. Shibuya, H. Siomi, M. C. Siomi, K. Saito, DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the Drosophila ovary. *Genes Dev.* **27**, 1656–1661 (2013).

59. G. Sienski, J. Batki, K.-A. Senti, D. Dönertas, L. Tirian, K. Meixner, J. Brennecke, Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. *Genes Dev.* **29**, 2258–2271 (2015).

60. Y. Yu, J. Gu, Y. Jin, Y. Luo, J. B. Preall, J. Ma, B. Czech, G. J. Hannon, Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science* **350**, 339–342 (2015).

61. J. Batki, J. Schnabl, J. Wang, D. Handler, V. I. Andreev, C. E. Stieger, M. Novatchkova, L. Lampersberger, K. Kauneckaite, W. Xie, K. Mechtler, D. J. Patel, J. Brennecke, The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. *Nat. Struct. Mol. Biol.* **26**, 720–731 (2019).

62. M. H. Fabry, F. Ciabrelli, M. Munafò, E. L. Eastwood, E. Kneuss, I. Falciatori, F. A. Falconio, G. J. Hannon, B. Czech, piRNA-guided co-transcriptional silencing coopts nuclear export factors. *Elife* **8** (2019).

63. K. Murano, Y. W. Iwasaki, H. Ishizu, A. Mashiko, A. Shibuya, S. Kondo, S. Adachi, S. Suzuki, K. Saito, T. Natsume, M. C. Siomi, H. Siomi, Nuclear RNA export factor variant initiates piRNA-guided co-transcriptional silencing. *EMBO J.* **38**, e102870 (2019).

64. K. Zhao, S. Cheng, N. Miao, P. Xu, X. Lu, Y. Zhang, M. Wang, X. Ouyang, X. Yuan, W. Liu, X. Lu, P. Zhou, J. Gu, Y. Zhang, D. Qiu, Z. Jin, C. Su, C. Peng, J.-H. Wang, M.-Q. Dong, Y. Wan, J. Ma, H. Cheng, Y. Huang, Y. Yu, A Pandas complex adapted for piRNA-guided transcriptional silencing and heterochromatin formation. *Nat. Cell Biol.* **21**, 1261–1272 (2019).

65. A. Zoch, T. Auchynnikava, R. V. Berrens, Y. Kabayama, T. Schöpp, M. Heep, L. Vasiliauskaitė, Y. A. Pérez-Rico, A. G. Cook, A. Shkumatava, J. Rappsilber, R. C.

Allshire, D. O'Carroll, SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation. *Nature* **584**, 635–639 (2020).

66. T. Schöpp, A. Zoch, R. V. Berrens, T. Auchynnikava, Y. Kabayama, L. Vasiliauskaitė, J. Rappsilber, R. C. Allshire, D. O'Carroll, TEX15 is an essential executor of MIWI2-directed transposon DNA methylation and silencing. *Nat Commun* **11**, 3739 (2020).

67. F. Yang, Y. Lan, R. R. Pandey, D. Homolka, S. L. Berger, R. S. Pillai, M. S. Bartolomei, P. J. Wang, TEX15 associates with MILI and silences transposable elements in male germ cells. *Genes Dev* **34**, 745–750 (2020).

68. J. Barau, A. Teissandier, N. Zamudio, S. Roy, V. Nalesso, Y. Hérault, F. Guillou, D. Bourc'his, The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354**, 909–912 (2016).

69. M. Ninova, Y.-C. A. Chen, B. Godneeva, A. K. Rogers, Y. Luo, K. Fejes Tóth, A. A. Aravin, Su(var)2-10 and the SUMO Pathway Link piRNA-Guided Target Recognition to Chromatin Silencing. *Mol Cell* **77**, 556-570.e6 (2020).

70. V. I. Andreev, C. Yu, J. Wang, J. Schnabl, L. Tirian, M. Gehre, D. Handler, P. Duchek, M. Novatchkova, L. Baumgartner, K. Meixner, G. Sienski, D. J. Patel, J. Brennecke, Panoramix SUMOylation on chromatin connects the piRNA pathway to the cellular heterochromatin machinery. *Nat Struct Mol Biol* **29**, 130–142 (2022).

71. M. Ninova, H. Holmes, B. Lomenick, K. Fejes Tóth, A. A. Aravin, Pervasive SUMOylation of heterochromatin and piRNA pathway proteins. *Cell Genom* **3**, 100329 (2023).

72. A. Arif, S. Bailey, N. Izumi, T. A. Anzelon, D. M. Ozata, C. Andersson, I. Gainetdinov, I. J. MacRae, Y. Tomari, P. D. Zamore, GTSF1 accelerates target RNA cleavage by PIWI-clade Argonaute proteins. *Nature* **608**, 618–625 (2022).

73. N. Izumi, K. Shoji, T. Kiuchi, S. Katsuma, Y. Tomari, The two Gtsf paralogs in silkworms orthogonally activate their partner PIWI proteins for target cleavage. *RNA* **29**, 18–29 (2022).

74. C. Klattenhoff, H. Xi, C. Li, S. Lee, J. Xu, J. S. Khurana, F. Zhang, N. Schultz, B. S. Koppetsch, A. Nowosielska, H. Seitz, P. D. Zamore, Z. Weng, W. E. Theurkauf, The Drosophila HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters. *Cell* **138**, 1137–1149 (2009).

75. A. Pane, P. Jiang, D. Y. Zhao, M. Singh, T. Schüpbach, The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. *EMBO J* **30**, 4601–4615 (2011).

76. F. Mohn, G. Sienski, D. Handler, J. Brennecke, The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in Drosophila. *Cell* **157**, 1364–1379 (2014).

77. Z. Zhang, J. Wang, N. Schultz, F. Zhang, S. S. Parhad, S. Tu, T. Vreven, P. D. Zamore, Z. Weng, W. E. Theurkauf, The HP1 Homolog Rhino Anchors a Nuclear Complex that Suppresses piRNA Precursor Splicing. *Cell* **157**, 1353–1363 (2014).

78. Y.-C. A. Chen, E. Stuwe, Y. Luo, M. Ninova, A. Le Thomas, E. Rozhavskaya, S. Li, S. Vempati, J. D. Laver, D. J. Patel, C. A. Smibert, H. D. Lipshitz, K. Fejes Toth, A. A. Aravin, Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. *Molecular Cell* **63**, 97–109 (2016).

79. L. Baumgartner, D. Handler, S. W. Platzer, C. Yu, P. Duchek, J. Brennecke, The Drosophila ZAD zinc finger protein Kipferl guides Rhino to piRNA clusters. *Elife* **11**, e80067 (2022).

80. P. R. Andersen, L. Tirian, M. Vunjak, J. Brennecke, A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* **549**, 54–59 (2017).

81. M. F. ElMaghraby, P. R. Andersen, F. Pühringer, U. Hohmann, K. Meixner, T. Lendl, L. Tirian, J. Brennecke, A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production. *Cell* **178**, 964-979.e20 (2019).

82. E. Kneuss, M. Munafò, E. L. Eastwood, U.-S. Deumer, J. B. Preall, G. J. Hannon, B. Czech, Specialization of the Drosophila nuclear export family protein Nxf3 for piRNA precursor export. *Genes Dev.* **33**, 1208–1220 (2019).

83. M. Munafò, V. R. Lawless, A. Passera, S. MacMillan, S. Bornelöv, I. U. Haussmann, M. Soller, G. J. Hannon, B. Czech, Channel nuclear pore complex subunits are required for transposon silencing in Drosophila. *Elife* **10**, e66321 (2021).

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

85. A. Pane, K. Wehr, T. Schüpbach, zucchini and squash Encode Two Putative Nucleases Required for rasiRNA Production in the Drosophila Germline. *Developmental Cell* **12**, 851–862 (2007).

86. A. D. Haase, S. Fenoglio, F. Muerdter, P. M. Guzzardo, B. Czech, D. J. Pappin, C. Chen, A. Gordon, G. J. Hannon, Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. *Genes Dev* **24**, 2499–2504 (2010).

87. D. Olivieri, M. M. Sykora, R. Sachidanandam, K. Mechtler, J. Brennecke, An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *EMBO J* **29**, 3301–3317 (2010).

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

89. H. Nishimasu, H. Ishizu, K. Saito, S. Fukuhara, M. K. Kamatani, L. Bonnefond, N. Matsumoto, T. Nishizawa, K. Nakanaga, J. Aoki, R. Ishitani, H. Siomi, M. C. Siomi, O. Nureki, Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* **491**, 284–287 (2012).

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

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

92. V. V. Vagin, Y. Yu, A. Jankowska, Y. Luo, K. A. Wasik, C. D. Malone, E. Harrison, A. Rosebrock, B. T. Wakimoto, D. Fagegaltier, F. Muerdter, G. J. Hannon, Minotaur is critical for primary piRNA biogenesis. *RNA* **19**, 1064–1077 (2013).

93. L. Ma, G. M. Buchold, M. P. Greenbaum, A. Roy, K. H. Burns, H. Zhu, D. Y. Han, R. A. Harris, C. Coarfa, P. H. Gunaratne, W. Yan, M. M. Matzuk, GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. *PLoS Genet* **5**, e1000635 (2009).

94. B. Czech, J. B. Preall, J. McGinn, G. J. Hannon, A Transcriptome-wide RNAi Screen in the Drosophila Ovary Reveals Factors of the Germline piRNA Pathway. *Molecular Cell* **50**, 749–761 (2013).

95. D. Handler, K. Meixner, M. Pizka, K. Lauss, C. Schmied, F. S. Gruber, J. Brennecke, The Genetic Makeup of the Drosophila piRNA Pathway. *Molecular Cell* **50**, 762–777 (2013).

96. M. Munafò, V. Manelli, F. A. Falconio, A. Sawle, E. Kneuss, E. L. Eastwood, J. W. E. Seah, B. Czech, G. J. Hannon, Daedalus and Gasz recruit Armitage to mitochondria, bringing piRNA precursors to the biogenesis machinery. *Genes Dev* **33**, 844–856 (2019).

97. D. Ding, J. Liu, K. Dong, U. Midic, R. A. Hess, H. Xie, E. Y. Demireva, C. Chen, PNLDC1 is essential for piRNA 3’ end trimming and transposon silencing during spermatogenesis in mice. *Nat Commun* **8**, 819 (2017).

98. Y. Zhang, R. Guo, Y. Cui, Z. Zhu, Y. Zhang, H. Wu, B. Zheng, Q. Yue, S. Bai, W. Zeng, X. Guo, Z. Zhou, B. Shen, K. Zheng, M. Liu, L. Ye, J. Sha, An essential role for PNLDC1 in piRNA 3’ end trimming and male fertility in mice. *Cell Res* **27**, 1392–1396 (2017).

99. T. Nishimura, I. Nagamori, T. Nakatani, N. Izumi, Y. Tomari, S. Kuramochi-Miyagawa, T. Nakano, PNLDC1, mouse pre-piRNA Trimmer, is required for meiotic and post-meiotic male germ cell development. *EMBO Rep* **19**, e44957 (2018).

100. M. D. Horwich, C. Li, C. Matranga, V. Vagin, G. Farley, P. Wang, P. D. Zamore, The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* **17**, 1265–1272 (2007).

101. Y. Kirino, Z. Mourelatos, Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nat Struct Mol Biol* **14**, 347–348 (2007).

102. T. Ohara, Y. Sakaguchi, T. Suzuki, H. Ueda, K. Miyauchi, T. Suzuki, The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nat Struct Mol Biol* **14**, 349–350 (2007).

103. K. Saito, Y. Sakaguchi, T. Suzuki, T. Suzuki, H. Siomi, M. C. Siomi, Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev* **21**, 1603–1608 (2007).

104. D. Olivieri, K.-A. Senti, S. Subramanian, R. Sachidanandam, J. Brennecke, The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in Drosophila. *Mol Cell* **47**, 954–969 (2012).

105. J. Xiol, E. Cora, R. Koglgruber, S. Chuma, S. Subramanian, M. Hosokawa, M. Reuter, Z. Yang, P. Berninger, A. Palencia, V. Benes, J. Penninger, R. Sachidanandam, R. S. Pillai, A role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing. *Mol Cell* **47**, 970–979 (2012).

106. H. Huang, Y. Li, K. E. Szulwach, G. Zhang, P. Jin, D. Chen, AGO3 Slicer activity regulates mitochondria-nuage localization of Armitage and piRNA amplification. *J Cell Biol* **206**, 217–230 (2014).

107. D. T. Ge, W. Wang, C. Tipping, I. Gainetdinov, Z. Weng, P. D. Zamore, The RNA-Binding ATPase, Armitage, Couples piRNA Amplification in Nuage to Phased piRNA Production on Mitochondria. *Mol Cell* **74**, 982-995.e6 (2019).

108. H. Ishizu, T. Kinoshita, S. Hirakata, C. Komatsuzaki, M. C. Siomi, Distinct and Collaborative Functions of Yb and Armitage in Transposon-Targeting piRNA Biogenesis. *Cell Rep* **27**, 1822-1835.e8 (2019).

109. H. Yamashiro, M. Negishi, T. Kinoshita, H. Ishizu, H. Ohtani, M. C. Siomi, Armitage determines Piwi-piRISC processing from precursor formation and quality control to inter-organelle translocation. *EMBO Rep* **21**, e48769 (2020).

110. R. R. Pandey, D. Homolka, K.-M. Chen, R. Sachidanandam, M.-O. Fauvarque, R. S. Pillai, Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in Drosophila ovaries. *PLoS Genet* **13**, e1006956 (2017).

111. A. K. Rogers, K. Situ, E. M. Perkins, K. F. Toth, Zucchini-dependent piRNA processing is triggered by recruitment to the cytoplasmic processing machinery. *Genes Dev* **31**, 1858–1869 (2017).

112. S. F. C. Soper, G. W. van der Heijden, T. C. Hardiman, M. Goodheart, S. L. Martin, P. de Boer, A. Bortvin, Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev Cell* **15**, 285–297 (2008).

113. J. Castañeda, P. Genzor, G. W. van der Heijden, A. Sarkeshik, J. R. Yates, N. T. Ingolia, A. Bortvin, Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice. *EMBO J* **33**, 1999–2019 (2014).

114. T. H. Chang, E. Mattei, I. Gainetdinov, C. Colpan, Z. Weng, P. D. Zamore, Maelstrom Represses Canonical Polymerase II Transcription within Bi-directional piRNA Clusters in Drosophila melanogaster. *Molecular Cell* **73**, 291-303.e6 (2019).

115. C. D. Malone, J. Brennecke, M. Dus, A. Stark, W. R. McCombie, R. Sachidanandam, G. J. Hannon, Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **137**, 522–535 (2009).

116. G.-W. Chirn, R. Rahman, Y. A. Sytnikova, J. A. Matts, M. Zeng, D. Gerlach, M. Yu, B. Berger, M. Naramura, B. T. Kile, N. C. Lau, Conserved piRNA Expression from a Distinct Set of piRNA Cluster Loci in Eutherian Mammals. *PLoS Genet* **11**, e1005652 (2015).

117. D. M. Özata, T. Yu, H. Mou, I. Gainetdinov, C. Colpan, K. Cecchini, Y. Kaymaz, P.-H. Wu, K. Fan, A. Kucukural, Z. Weng, P. D. Zamore, Evolutionarily conserved pachytene piRNA loci are highly divergent among modern humans. *Nat Ecol Evol* **4**, 156–168 (2020).

118. D. Gebert, L. K. Neubert, C. Lloyd, J. Gui, R. Lehmann, F. K. Teixeira, Large Drosophila germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. *Mol Cell* **81**, 3965-3978.e5 (2021).

119. S. Signor, J. Vedanayagam, B. Y. Kim, F. Wierzbicki, R. Kofler, E. C. Lai, Rapid evolutionary diversification of the flamenco locus across simulans clade Drosophila species. *PLoS Genet* **19**, e1010914 (2023).

120. P.-H. Wu, Y. Fu, K. Cecchini, D. M. Özata, A. Arif, T. Yu, C. Colpan, I. Gainetdinov, Z. Weng, P. D. Zamore, The evolutionarily conserved piRNA-producing locus pi6 is required for male mouse fertility. *Nat Genet* **52**, 728–739 (2020).

121. H. Choi, Z. Wang, J. Dean, Sperm acrosome overgrowth and infertility in mice lacking chromosome 18 pachytene piRNA. *PLoS Genet* **17**, e1009485 (2021).

122. P. Chen, A. A. Kotov, B. K. Godneeva, S. S. Bazylev, L. V. Olenina, A. A. Aravin, piRNA-mediated gene regulation and adaptation to sex-specific transposon expression in D. melanogaster male germline. *Genes Dev* **35**, 914–935 (2021).

123. T. Kiuchi, H. Koga, M. Kawamoto, K. Shoji, H. Sakai, Y. Arai, G. Ishihara, S. Kawaoka, S. Sugano, T. Shimada, Y. Suzuki, M. G. Suzuki, S. Katsuma, A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature* **509**, 633–636 (2014).

124. T. Fukui, M. Kawamoto, K. Shoji, T. Kiuchi, S. Sugano, T. Shimada, Y. Suzuki, S. Katsuma, The Endosymbiotic Bacterium Wolbachia Selectively Kills Male Hosts by Targeting the Masculinizing Gene. *PLoS Pathog* **11**, e1005048 (2015).

125. J. Lee, T. Kiuchi, M. Kawamoto, T. Shimada, S. Katsuma, Identification and functional analysis of a Masculinizer orthologue in Trilocha varians (Lepidoptera: Bombycidae). *Insect Mol Biol* **24**, 561–569 (2015).

126. Y.-H. Wang, X.-E. Chen, Y. Yang, J. Xu, G.-Q. Fang, C.-Y. Niu, Y.-P. Huang, S. Zhan, The Masc gene product controls masculinization in the black cutworm, Agrotis ipsilon. *Insect Sci* **26**, 1037–1044 (2019).

127. T. Fukui, K. Shoji, T. Kiuchi, Y. Suzuki, S. Katsuma, Masculinizer is not post-transcriptionally regulated by female-specific piRNAs during sex determination in the Asian corn borer, Ostrinia furnacalis. *Insect Biochem Mol Biol* **156**, 103946 (2023).

128. T. Harvey-Samuel, X. Xu, M. A. E. Anderson, L. Z. Carabajal Paladino, D. Purusothaman, V. C. Norman, C. M. Reitmayer, M. You, L. Alphey, Silencing RNAs expressed from W-linked PxyMasc “retrocopies” target that gene during female sex determination in Plutella xylostella. *Proc Natl Acad Sci U S A* **119**, e2206025119 (2022).

129. W. Tang, M. Seth, S. Tu, E.-Z. Shen, Q. Li, M. Shirayama, Z. Weng, C. C. Mello, A Sex Chromosome piRNA Promotes Robust Dosage Compensation and Sex Determination in C. elegans. *Developmental Cell* **44**, 762-770.e3 (2018).

130. I. V. Kim, E. M. Duncan, E. J. Ross, V. Gorbovytska, S. H. Nowotarski, S. A. Elliott, A. Sánchez Alvarado, C.-D. Kuhn, Planarians recruit piRNAs for mRNA turnover in adult stem cells. *Genes Dev* **33**, 1575–1590 (2019).

131. R. Halbach, P. Miesen, J. Joosten, E. Taşköprü, I. Rondeel, B. Pennings, C. B. F. Vogels, S. H. Merkling, C. J. Koenraadt, L. Lambrechts, R. P. van Rij, A satellite repeat-derived piRNA controls embryonic development of Aedes. *Nature* **580**, 274–277 (2020).

132. H. Wang, F. Jiang, X. Liu, Q. Liu, Y. Fu, R. Li, L. Hou, J. Zhang, J. He, L. Kang, Piwi/piRNAs control food intake by promoting neuropeptide F expression in locusts. *EMBO Rep* **23**, e50851 (2022).

133. E. M. Morazzani, M. R. Wiley, M. G. Murreddu, Z. N. Adelman, K. M. Myles, Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. *PLoS Pathog* **8**, e1002470 (2012).

134. E. Schnettler, C. L. Donald, S. Human, M. Watson, R. W. C. Siu, M. McFarlane, J. K. Fazakerley, A. Kohl, R. Fragkoudis, Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. *J Gen Virol* **94**, 1680–1689 (2013).

135. Y. H. Sun, L. H. Xie, X. Zhuo, Q. Chen, D. Ghoneim, B. Zhang, J. Jagne, C. Yang, X. Z. Li, Domestic chickens activate a piRNA defense against avian leukosis virus. *Elife* **6**, e24695 (2017).

136. T. Yu, B. S. Koppetsch, S. Pagliarani, S. Johnston, N. J. Silverstein, J. Luban, K. Chappell, Z. Weng, W. E. Theurkauf, The piRNA Response to Retroviral Invasion of the Koala Genome. *Cell* **179**, 632-643.e12 (2019).

137. P. Chen, Y. Luo, A. A. Aravin, RDC complex executes a dynamic piRNA program during Drosophila spermatogenesis to safeguard male fertility. *PLoS Genet* **17**, e1009591 (2021).

138. X. Wei, D. G. Eickbush, I. Speece, A. M. Larracuente, Heterochromatin-dependent transcription of satellite DNAs in the Drosophila melanogaster female germline. *Elife* **10**, e62375 (2021).

139. P. Genzor, P. Konstantinidou, D. Stoyko, A. Manzourolajdad, C. Marlin Andrews, A. R. Elchert, C. Stathopoulos, A. D. Haase, Cellular abundance shapes function in piRNA-guided genome defense. *Genome Res* **31**, 2058–2068 (2021).

140. J. P. Blumenstiel, A. A. Erwin, L. W. Hemmer, What Drives Positive Selection in the Drosophila piRNA Machinery? The Genomic Autoimmunity Hypothesis. *Yale J Biol Med* **89**, 499–512 (2016).

141. S. Houwing, L. M. Kamminga, E. Berezikov, D. Cronembold, A. Girard, H. van den Elst, D. V. Filippov, H. Blaser, E. Raz, C. B. Moens, R. H. A. Plasterk, G. J. Hannon, B. W. Draper, R. F. Ketting, A Role for Piwi and piRNAs in Germ Cell Maintenance and Transposon Silencing in Zebrafish. *Cell* **129**, 69–82 (2007).

142. S. Kawaoka, K. Minami, S. Katsuma, K. Mita, T. Shimada, Developmentally synchronized expression of two Bombyx mori Piwi subfamily genes, SIWI and BmAGO3 in germ-line cells. *Biochem Biophys Res Commun* **367**, 755–760 (2008).

143. C. E. Juliano, A. Reich, N. Liu, J. Götzfried, M. Zhong, S. Uman, R. A. Reenan, G. M. Wessel, R. E. Steele, H. Lin, PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells. *Proc Natl Acad Sci U S A* **111**, 337–342 (2014).

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

145. S. H. Lewis, H. Salmela, D. J. Obbard, Duplication and Diversification of Dipteran Argonaute Genes, and the Evolutionary Divergence of Piwi and Aubergine. *Genome Biol Evol* **8**, 507–518 (2016).

146. J. Gutierrez, R. Platt, J. C. Opazo, D. A. Ray, F. Hoffmann, M. Vandewege, Evolutionary history of the vertebrate Piwi gene family. *PeerJ* **9**, e12451 (2021).

147. D. Li, D. H. Taylor, J. C. van Wolfswinkel, PIWI-mediated control of tissue-specific transposons is essential for somatic cell differentiation. *Cell Rep* **37**, 109776 (2021).

148. E. F. Roovers, D. Rosenkranz, M. Mahdipour, C.-T. Han, N. He, S. M. Chuva de Sousa Lopes, L. A. J. van der Westerlaken, H. Zischler, F. Butter, B. A. J. Roelen, R. F. Ketting, Piwi proteins and piRNAs in mammalian oocytes and early embryos. *Cell Rep* **10**, 2069–2082 (2015).

149. Q. Yang, R. Li, Q. Lyu, L. Hou, Z. Liu, Q. Sun, M. Liu, H. Shi, B. Xu, M. Yin, Z. Yan, Y. Huang, M. Liu, Y. Li, L. Wu, Single-cell CAS-seq reveals a class of short PIWI-interacting RNAs in human oocytes. *Nat Commun* **10**, 3389 (2019).

150. H. Hasuwa, Y. W. Iwasaki, W. K. Au Yeung, K. Ishino, H. Masuda, H. Sasaki, H. Siomi, Production of functional oocytes requires maternally expressed PIWI genes and piRNAs in golden hamsters. *Nat Cell Biol* **23**, 1002–1012 (2021).

151. K. Ishino, H. Hasuwa, J. Yoshimura, Y. W. Iwasaki, H. Nishihara, N. M. Seki, T. Hirano, M. Tsuchiya, H. Ishizaki, H. Masuda, T. Kuramoto, K. Saito, Y. Sakakibara, A. Toyoda, T. Itoh, M. C. Siomi, S. Morishita, H. Siomi, Hamster PIWI proteins bind to piRNAs with stage-specific size variations during oocyte maturation. *Nucleic Acids Res* **49**, 2700–2720 (2021).

152. Z. Loubalova, H. Fulka, F. Horvat, J. Pasulka, R. Malik, M. Hirose, A. Ogura, P. Svoboda, Formation of spermatogonia and fertile oocytes in golden hamsters requires piRNAs. *Nat Cell Biol* **23**, 992–1001 (2021).

153. H. Zhang, F. Zhang, Q. Chen, M. Li, X. Lv, Y. Xiao, Z. Zhang, L. Hou, Y. Lai, Y. Zhang, A. Zhang, S. Gao, H. Fu, W. Xiao, J. Zhou, F. Diao, A. Shi, Y.-Q. Su, W. Zeng, L. Wu, J. Li, The piRNA pathway is essential for generating functional oocytes in golden hamsters. *Nat Cell Biol* **23**, 1013–1022 (2021).

154. O. S. Akbari, I. Antoshechkin, H. Amrhein, B. Williams, R. Diloreto, J. Sandler, B. A. Hay, The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. *G3 (Bethesda)* **3**, 1493–1509 (2013).

155. A. Wilczynska, N. Minshall, J. Armisen, E. A. Miska, N. Standart, Two Piwi proteins, Xiwi and Xili, are expressed in the Xenopus female germline. *RNA* **15**, 337–345 (2009).

156. P. J. Batista, J. G. Ruby, J. M. Claycomb, R. Chiang, N. Fahlgren, K. D. Kasschau, D. A. Chaves, W. Gu, J. J. Vasale, S. Duan, D. Conte, S. Luo, G. P. Schroth, J. C. Carrington, D. P. Bartel, C. C. Mello, PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol Cell* **31**, 67–78 (2008).

157. P. P. Das, M. P. Bagijn, L. D. Goldstein, J. R. Woolford, N. J. Lehrbach, A. Sapetschnig, H. R. Buhecha, M. J. Gilchrist, K. L. Howe, R. Stark, N. Matthews, E. Berezikov, R. F. Ketting, S. Tavaré, E. A. Miska, Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. *Mol Cell* **31**, 79–90 (2008).

158. G. Wang, V. Reinke, A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Curr Biol* **18**, 861–867 (2008).

159. M. V. Almeida, M. A. Andrade-Navarro, R. F. Ketting, Function and Evolution of Nematode RNAi Pathways. *Noncoding RNA* **5**, 8 (2019).

160. C. Li, V. V. Vagin, S. Lee, J. Xu, S. Ma, H. Xi, H. Seitz, M. D. Horwich, M. Syrzycka, B. M. Honda, E. L. W. Kittler, M. L. Zapp, C. Klattenhoff, N. Schulz, W. E. Theurkauf, Z. Weng, P. D. Zamore, Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. *Cell* **137**, 509–521 (2009).

Figure 1. Makeup of the piRNA pathway: four modules.

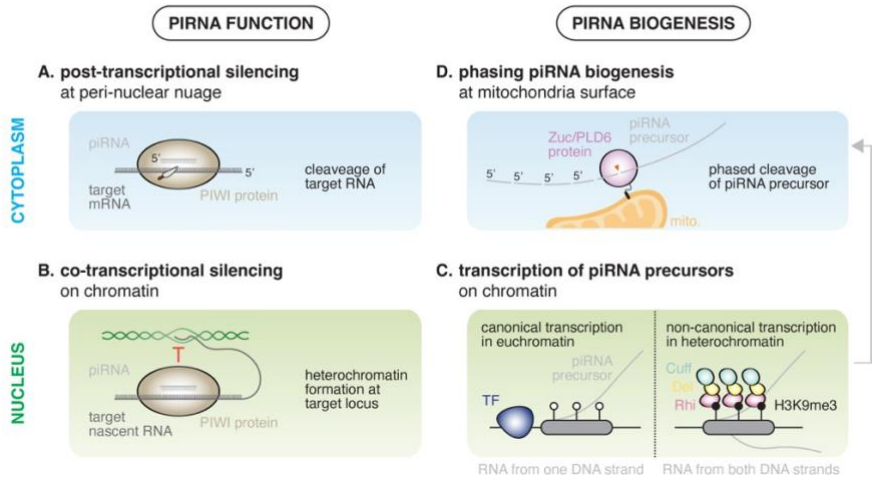
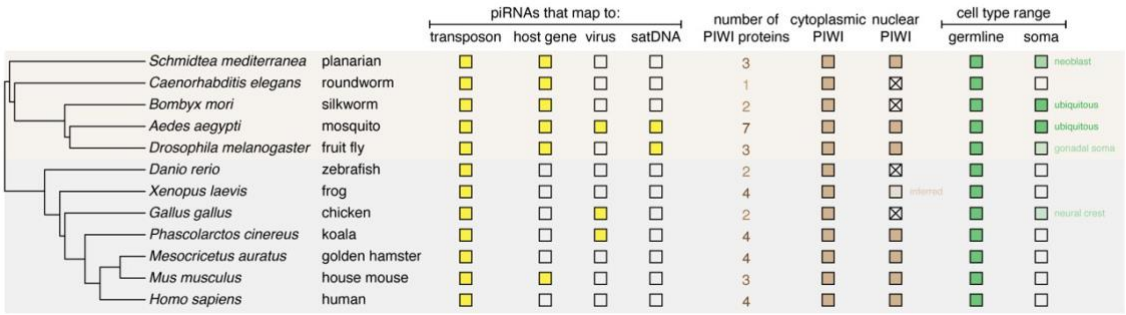


Illustration of four key modules of the piRNA pathway.

Blue backgrounds mark cytoplasmic events, while green backgrounds mark nuclear events.

The underlying processes are simplified to highlight the key aspects of each module.

Figure 2. Diversity and rapid evolution of the piRNA pathway.



Diversity and rapid evolution of the piRNA pathway components.

Colored squares represent the presence of the molecule of interest, while the empty ones represent the absence. The species included are some of those whose piRNAs and/or piRNA pathway machinery genes have been studied.

Chapter 3

PIRNA-MEDIATED GENE REGULATION AND ADAPTATION TO SEX-SPECIFIC TRANSPOSON EXPRESSION IN D. MELANOGASTER MALE GERMLINE

P. Chen, A. A. Kotov, B. K. Godneeva, S. S. Bazylev, L. V. Olenina, A. A. Aravin (2021). *Genes Dev* **35**, 914–935. https://doi.org/10.1101/gad.345041.120

ABSTRACT

Small non-coding piRNAs act as sequence-specific guides to repress complementary targets in Metazoa. Prior studies in *Drosophila* ovaries have demonstrated the function of piRNA pathway in transposon silencing and therefore genome defense. However, the ability of piRNA program to respond to different transposon landscape and the role of piRNAs in regulating host gene expression remain poorly understood. Here, we comprehensively analyzed piRNA expression and defined the repertoire of their targets in *Drosophila melanogaster* testes. Comparison of piRNA programs between sexes revealed sexual dimorphism in piRNA programs that parallel sex-specific transposon expression. Using a novel bioinformatic pipeline, we identified new piRNA clusters and established complex satellites as dual-strand piRNA clusters. While sharing most piRNA clusters, two sexes employ them differentially to combat sex-specific transposon landscape. We found two piRNA clusters that produce piRNAs antisense to four host genes in testis, including *CG12717/pirate*, a SUMO protease gene. piRNAs encoded on Y chromosome silence *pirate*, but not its paralog, to exert sex- and paralog-specific gene regulation. Interestingly, *pirate* is targeted by endogenous siRNAs in a sibling species, *Drosophila mauritiana*, suggesting distinct but related silencing strategies invented in recent evolution to regulate a conserved protein-coding gene.

Main Text

INTRODUCTION

PIWI-interacting (pi)RNA is a class of small non-coding RNAs named after their interaction with PIWI-clade Argounate proteins. piRNAs guide PIWI proteins to complementary RNAs, thereby specifying the target of PIWI silencing. Unlike miRNAs and siRNAs that are ubiquitously expressed, the expression of piRNAs is restricted to gonads in many animals. As a result, perturbation of the piRNA program often compromises reproductive functions with no obvious defects in soma. *Drosophila melanogaster* is one of the most used model organisms to study piRNA biogenesis and function. In fact, piRNAs were first described in fly testes (*1*, *2*). However, most subsequent studies were performed using ovaries as a model system. Work on female gonads has shown that most piRNAs have homology to transposable elements (TEs), suggesting TEs as major targets of piRNAs (*3*). Studies on fly ovaries also identified large intergenic regions dubbed piRNA clusters that harbor nested TE fragments, which act as genomic source loci of piRNAs. A peri-centromeric region on chr2R called *42AB* was found to be the most active piRNA cluster in ovaries. It remains largely unexplored to what extent these findings from ovaries are applicable to the male counterpart. To date, we still know very little about how sexually dimorphic the *Drosophila* piRNA program is, besides that there is a single locus on Y chromosome called *Suppressor of Stellate* (*Su(Ste)*) that produces piRNAs only in males.

Importantly, *Drosophila* as an animal model offers unique value to studying sexual dimorphism of the piRNA program in general. In zebrafish, piRNA pathway mutants are

always phenotypically males (*4–6*), rendering it nearly impossible to probe the impact of piRNA loss in females. In mice, an intact piRNA program is only required for male fertility, while murine females are insensitive to piRNA loss (*7–9*). Contrary to fish and mouse, fly fertility is dependent on a functional piRNA pathway in both sexes (*1–3, 10*). Therefore, *Drosophila* provides an unparalleled opportunity to study whether, and if so how, the piRNA program can be modified in each sex to safeguard reproductive functions.

In this study, we comprehensively analyzed the piRNA profile in *Drosophila melanogaster* testis and compared it to the female counterpart. Besides TEs, we found complex satellites as another class of selfish genetic elements targeted by the piRNA pathway in gonads of both sexes. Our analysis showed that TE-silencing piRNA program is sexually dimorphic, and it shows evidence of adaptation to sex-specific TE landscape. To understand the genomic origins of differentially produced piRNAs, we sought to *de novo* define genome-wide piRNA clusters in testis. However, we noticed that the standard pipeline used for ovary piRNAs failed to detect known piRNA clusters in testis, so we developed a new bioinformatic algorithm to tackle this problem. Using the new algorithm, we were able to identify novel piRNA clusters and to quantify their activities more accurately in both sexes. Notably, piRNA source loci are employed differentially in males and females, and the sex bias of piRNA cluster expression appears to match that of their TE contents. We also found two loci producing piRNAs with the potential to repress host protein-coding genes, including a newly identified locus on Y that we named *petrel*, which produces piRNAs against *CG12717/pirate*. Expression of *pirate*, but not its close paralog *verloren*, is de-repressed in

multiple piRNA pathway mutants, indicating that piRNAs silence its expression and can distinguish paralogs with sequence similarities. Finally, we explored the evolutionary history of *pirate* and found it to be a young gene conserved in the *melanogaster* subgroup. Intriguingly, *pirate* is targeted by another class of small non-coding RNAs, endogenous siRNAs, in the sibling species *Drosophila mauritiana*, suggesting distinct small RNA-based silencing strategies invented in recent evolution to regulate a young yet conserved gene.

RESULTS

Drosophila piRNA program is sexually dimorphic

To characterize the piRNA profile in male gonads, we sequenced 18–30nt small RNAs from testes and compared them with published ovary small RNA datasets (*11*). Mapping and annotation of small RNA reads using the pipeline shown in Supplementary Fig. S1 revealed large differences in the expression of major classes of small RNAs between testes and ovaries. In agreement with previous findings (*12*), TE-mapping 23–29nt piRNAs are the most abundant class of small RNAs in ovaries, while 21–23nt microRNAs constitute a minor fraction and an even smaller one for 21nt endogenous (endo-) siRNAs (Fig. 1A). In contrast, miRNAs constitute a larger fraction in testes, so do endo-siRNAs that map to protein-coding genes, consistent with a previous report (*13*). To define the piRNA population, we eliminated reads mapping to other types of non-coding RNA (rRNA, miRNA, snRNA, snoRNA and tRNA) from 23–29nt small RNAs. Remaining reads show a strong bias for U at the first nucleotide ("1U bias": 70.9%), the feature of *bona fide* piRNAs (Fig. 1B). The piRNA-to-miRNA ratio is distinct between sexes: ~10 in ovary and ~2 in testis. In both sexes, piRNAs

mapping to TEs take up the largest fraction of total piRNAs. However, whereas 66% of piRNAs mapped to TEs in ovaries, only 40% mapped to TEs in testes (Fig. 1B). Meanwhile, larger fractions of total piRNAs mapped to protein-coding genes (including introns) and intergenic regions in testes (24.6% and 30.0%, respectively) than ovaries (19.6% and 10.7%, respectively). These results suggest that distinct piRNA programs operate in male and female gonads.

Testis piRNAs also map to several known complex satellites: *HETRP*/*TAS* (a sub-telomeric satellite repeat), *Responder (Rsp)* and *SAR* (related to *1.688* repeat family) (Fig. 1C; Supplementary Fig. S2A). Complex satellite-mapping small RNAs in testis exhibit 1U bias and size distribution that peaks around 24–26nt, consistent with their piRNA identities. Both strands of complex satellites produce piRNAs, and their production depends on Rhino (*14*), a protein that marks dual-strand piRNA clusters and is required for their expression (*15–17*). Similarly, ovary small RNAs also map to complex satellites and show features of *bone fide* piRNAs, including 1U bias, size distribution that peaks around 24–26nt, small RNA production from both strands and dependency on Rhino. Moreover, piRNAs from complex satellites show ping-pong signature, an enrichment for 10nt overlap between the 5' ends of complementary piRNA pairs, except for *Rsp* in testis (Fig. 1C; Supplementary Fig. S2C). Finally, we examined the phasing pattern, the presence of piRNAs arranged tail-to-head one after another as a result of phased processing of piRNA precursors (*18*, *19*). We found such a phasing signature for two complex satellites in ovary, but not in testis (Supplementary Fig. S2B). These results show that complex satellites are sources of piRNAs in both sexes,

pointing to a possible role of piRNAs in regulating satellite DNA and associated heterochromatin in the gonad.

We next analyzed piRNAs targeting different TE families. Comparison of small RNA profiles in testis and ovary showed that piRNAs targeting different TEs are expressed at different levels in two sexes (Fig. 2A). Top 3 TEs targeted by piRNA are all different in testis and ovary, and, among top 10, only three are shared between sexes (Supplementary Fig. S2D). The most differentially targeted TEs are two telomere-associated TEs, *HeT-A* and *TAHRE*, which ovary makes 106 and 74 times more antisense piRNAs, respectively, than testis. In contrast, several elements such as *baggins1*, *invader3* and *copia* are targeted by more piRNAs in testis. piRNAs targeting all but one (*copia*) TE families show stronger ping-pong signature in ovary, as measured by ping-pong z-score (Fig. 2A). In conclusion, different TE families are targeted by piRNAs differentially in two sexes.

Distinct piRNA programs in two sexes parallel sex-specific TE expression

To explore if sex differences in TE-targeting piRNA programs are accompanied by differential expression of TEs themselves, we set out to compare expression levels of different TE families in two sexes. Since piRNA pathway efficiently represses TEs, their expression in wild-type animals does not reflect their full expression potentials that can be achieved when piRNA silencing is removed. Hence, we analyzed TE expression in testes

and ovaries of *rhi* mutants that lose piRNA production from dual-strand clusters in both sexes (*14*) and controls.

Profiling TE expression in two sexes by polyA-selected (polyA+) RNA-seq demonstrated clear sexual dimorphism. Overall, TE expression in piRNA pathway mutant testes and ovaries is weakly correlated (Spearman's ρ: 0.18; Supplementary Fig. S3A). Among the ten most expressed TE families in two sexes, only four overlap, though the same element, *copia*, has the highest expression in both ovary and testis (Supplementary Fig. S3B). There are more TE families expressed above each of the three expression cutoffs (1000, 100 and 10 RPKM) in ovaries than testes (Supplementary Fig. S3A). The most ovary-biased TEs include *Blood*, *Gypsy12*, *Burdock* and two telomere-associated TEs, *HeT-A* and *TART* (Fig. 2B). Only a few TE families are expressed higher in testis than ovary (Fig. 2B; Supplementary Fig. S3A). In this group, *Transib2* and *doc2* show the strongest bias for expression in testis (14- and 6.5-fold higher in testis than ovary, respectively). Overall, the majority of TE families demonstrate strong differences in their expression between sexes.

To quantify the effect of piRNA pathway in suppressing TEs in two sexes, we calculated levels of TE de-repression upon disruption of piRNA pathway. Few TE families remained unaffected by *rhi* mutation, often accompanied by unperturbed antisense piRNA production (e.g., *gypsy*, *gypsy10* and *tabor*). There are 9 TE families up-regulated more than 100-fold in ovary. In contrast, no TE is up-regulated that strongly in testis (Fig. 2C). Overall, the vast majority of TEs show stronger de-repression in ovaries, with *gypsy12* (389-fold), *Burdock*

(317-fold), *HeT-A* (239-fold) and *TART* (80-fold) being the most prominent examples, as all of them exhibited no or mild de-repression (<4-fold) in testes (Fig. 2B). We found only 6 TEs that show stronger (at least 4-fold) de-repression in testis than ovary (*Transib2, BS2, baggins1, Dm297, invader3, invader6*). Altogether, our results show that piRNAs regulate the expression of different TE families to distinct extents in two sexes, with many TEs silenced more in ovary and only a few silenced more in testis.

To explore the link between TE expression and piRNA programs in two sexes, we identified a set of 36 TE families repressed by piRNA pathway in at least one sex (see methods). For these TE families, there is a positive correlation between sex bias of piRNA production and sex bias of TE de-repression (Pearson's ρ: 0.53, P<0.001; Fig. 2D). For example, disruption of piRNA pathway by *rhi* mutations dramatically increases expression of three telomere-associated TEs (*HeT-A, TAHRE* and *TART*) in ovaries, where there are abundant piRNAs targeting these elements. On the contrary, much fewer piRNAs target these telomeric TEs in testes and expression of these TEs remained very low in *rhi* mutant males (Fig. 2A, B, E). This result indicates that telomeric TEs have a strong, intrinsic bias in their expression towards the female germline, and that piRNA pathway appears to have adapted to this bias generating respective antisense piRNAs in female, but not male, gonads. In contrast to ovary-biased TEs like telomeric elements, testis-biased TEs such as *Transib2* and *baggins1* are targeted by more antisense piRNAs in testis than ovary (Fig. 2A, B, E). Some TEs, such as *copia*, *mdg3* and *I-element* are strongly repressed by piRNAs in both sexes. For such elements, the sex bias in piRNA production does not always match that of TE repression

(Fig. 2D). Taken together, these findings suggest that, for most TEs, piRNA programs in males and females have adapted to differential TE activities between sexes.

To further explore whether differential expression of piRNAs between sexes has functional consequences, we studied *Burdock*, an long terminal repeat (LTR) retro-transposon targeted by 53 times more piRNAs in ovary (3,756 RPM) than testis (70 RPM) (Fig. 2A). We used a reporter composed of a fragment of *Burdock* expressed under the control of heterologous *nanos* promoter that drives expression in germline of both sexes (*20*). While reporter was efficiently silenced in ovaries of wild-type flies, it was strongly de-repressed in piRNA pathway mutants ($rhi^{-/-}$) (Fig. 2F), indicating that the piRNA program efficiently silences *Burdock* in female germline. In contrast, we observed strong reporter expression in testes of wild-type males, and the disruption of piRNA pathway in *rhi* mutants did not lead to an observable increase in its expression (Fig. 2F). This finding shows that *Burdock* is not silenced in testes, likely as a result of very few *Burdock*-targeting piRNAs in males (Fig. 2A). Notably, expression of endogenous *Burdock* is high in ovary (when piRNA production is disrupted) but low in both wild-type and mutant testis (Fig. 2B, E). Thus, similar to telomeric TEs, the ability of piRNA pathway to repress *Burdock* in female but not male germline correlates with an intrinsic bias for its expression in females. We conclude that differential expression of TE-targeting piRNAs in male and female gonads can have functional consequences in their abilities to silence TEs.

Definition of piRNA clusters in testis with a new algorithm

To get deeper understandings of the piRNA program in male gonads, we sought to define the genomic origin of piRNAs and compare it between two sexes. Since genome-wide identification of piRNA clusters has only been done in ovary, we decided to systematically search for genomic loci that generate piRNA in testis. We noticed that two major clusters in testis identified to date, *Su(Ste)* and *AT-chX*, both contain internal tandem repeats, i.e., they are made of many copies of almost identical sequences (*1*, *21*). As a result, most piRNAs produced by these two loci mapped to the genome at multiple positions. However, the algorithm employed in previous studies to systematically define piRNA clusters in ovary only uses piRNAs that map to the genome at single unique positions (*3*, *16*, *22*), raising the question of whether it is an appropriate approach to detect clusters like *Su(Ste)* composed primarily of internal tandem repeats. In fact, both *Su(Ste)* and *AT-chX* clusters were initially identified by different approaches (*1*, *23*).

Even though piRNAs produced from *Su(Ste)* and *AT-chX* cannot be mapped to single unique genomic loci, most of them mapped to several local repeats inside the respective clusters but nowhere else in the genome (Fig. 3A). Taking advantage of this property, we developed a new algorithm that takes into account local repeats to define piRNA clusters (Supplementary Fig. S4A, B). Briefly, in addition to uniquely mapped piRNAs, the algorithm searches for piRNA sequences that map to multiple positions within a single genomic region but nowhere else in the genome. This approach ensures that the identified region as a whole generates piRNAs, though the exact origin within the region remains unknown. Unlike the previous approach that uses exclusively uniquely mapped piRNAs, this algorithm successfully

identified *Su(Ste)* and *AT-chX*, two major piRNA clusters in testis that contain local repeats (Fig. 3E, F).

We applied this new algorithm to systematically identify piRNA clusters active in testes. We recovered piRNA clusters known to be active in testes as well as piRNA clusters previously defined in ovaries (e.g., *42AB, 38C, 20A* and *flam*) (Fig. 3C; Supplementary Table S1). Furthermore, our search identified several novel piRNA loci. One of the novel piRNA clusters is located on Y chromosome flanked by *FDY* and *Mst77Y* genes (Fig. 3C, F) around heterochromatin band *h17* (*24*). We named this locus *petrel* for "proximal to fertility regions on YL". Another novel locus is *h52-1*, flanked by *eIF4B* and *CG17514* genes on chr3L. *h52-1* harbors tandem local repeats composed of nested TE fragments that cannot be found elsewhere in the genome. Similar to piRNA clusters identified in ovaries, only a few clusters active in testes produce piRNAs from one genomic strand (e.g., *flam* and *20A*, so-called "uni-strand clusters"), and the majority are dual-strand clusters that generate piRNAs from both genomic strands (Fig. 3E). In sum, our algorithm successfully found previously known piRNAs clusters and identified novel ones in *Drosophila* testes.

To compare new algorithm with the approach that considers only uniquely mapped piRNAs, we applied both techniques to analyze the same testis piRNA dataset. This comparison showed that major piRNA clusters in testis can be divided into two groups (Fig. 3B). The first group (*42AB, 38C, 20A* and *flam*) contains piRNA clusters that harbor many unique sequences, so including local repeats does not substantially change their identification and

quantification. On the other hand, the second group of genomic loci (*Su(Ste), AT-chX*, *petrel*, *Hsp70B* and *h52-1*) is composed of piRNA clusters that contain few unique sequences but many local repeats, and, accordingly, our new algorithm identified more than 10-fold more piRNAs produced from these loci (Fig. 3B). Thus, this algorithm is not only useful for finding new piRNA source loci but also provides a more accurate quantification of piRNA production from previously known clusters.

Sex difference in piRNA cluster expression

To compare the expression of piRNA clusters between sexes, we first applied our algorithm to published ovary piRNA datasets (Fig. 3D; Supplementary Table S1) (*11*). Thus, piRNA clusters were defined and their activities were quantified in both sexes using the same algorithm, allowing for fair comparison. Surprisingly, our analysis revealed that *AT-chX*, originally described as a piRNA cluster in testes, is also highly active in ovaries. *AT-chX* locus consists of local repeats (*21*), so piRNAs produced from this locus were excluded in previous studies that analyzed only uniquely mapped reads. In fact, *AT-chX* is the second most active piRNA cluster in ovary, producing ~7% of total piRNAs.

Comparison between piRNA clusters in males and females revealed a clear sex difference: a small number of loci produce the majority of piRNAs in testis, which is not the case for ovary (Fig. 4A). The two most active piRNA clusters in testes, *Su(Ste)* on Y chromosome and *AT-chX* on X chromosome, produce ~43% and ~31% of total piRNAs in testes, respectively

(Fig. 3C). They are followed by the novel piRNA cluster on Y chromosome, *petrel*, that produces ~4% piRNAs. Along with another 6 loci, the top 9 piRNA clusters in testis account for 81.8% of total piRNAs. In comparison, only 30.4% of total piRNAs are made from the top 9 clusters in ovary, with the most active locus *42AB* producing ~11% of total piRNAs (Fig. 3D). Whereas a few loci dominate the global piRNA population in testis, the ovary piRNA profile is shaped by many loci producing piRNAs in comparable amounts.

Next, we compared expression levels of different piRNA clusters in male and female gonads. Females lack Y chromosome, so they do not have piRNAs produced by Y-linked *Su(Ste)* and *petrel* clusters. For major clusters present in both male and female genomes, we observed pronounced sex differences (Spearman's ρ: 0.07; Fig. 4B). For instance, *38C* produces more piRNAs than *42AB*, *80EF* and *40F7* in testes, but the opposite trend is found in ovaries. Some loci such as *Sox102F* on chr4 (*16*, *17*) appear to be active only in ovaries but not in testes (Fig. 4F). These differentially expressed piRNA clusters located on autosomes, which both males and females have two copies, exemplify the sex-specific usage of piRNA loci. Moreover, we examined expression levels of major piRNA clusters on chrX (*AT-chX*, *flam* and *20A*), which females have two copies (XX) and males have only one (XY). We found that a larger fraction of piRNAs originate from *AT-chX* in testes than ovaries, but the reverse was found for *flam* and *20A*, suggesting that copy numbers of piRNA clusters do not correlate well with their expression. Altogether, these findings illustrate a sexually dimorphic employment of piRNA clusters, where different loci are engaged differentially in a sex-specific manner.

Different piRNA clusters have distinct TE contents, so their differential expression might sculpt sex-specific piRNA programs with distinct TE-silencing capacities in males and females. To explore a link between the expression of piRNA cluster and its TE content, we computed cumulative sex bias of the TE content of each major piRNA cluster (Fig. 4C). This was done by summing sex biases of individual TEs in the piRNA cluster weighted by their length contributions to the cluster (see example in Fig. 4C). The sex bias of cluster TE content matches the sex bias in piRNA cluster expression, suggesting a link between the expression of piRNA clusters and TEs they control. To substantiate this finding, we analyzed sequence compositions of three differentially expressed piRNA clusters: *42AB* (ovary-biased), *38C* (testis-biased) and *Sox102F* (ovary-specific). The top 4 TEs most enriched by length in ovary-biased *42AB* (*batumi*, *gypsy12*, *FW* and *DMRT1b*) are all ovary-biased in their expression (Fig. 4D; Supplementary Fig. S3A). Importantly, these 4 TEs are completely absent in testis-biased *38C* cluster. In contrast, three testis-biased TE families, *hobo*, *BS2* and *Transib2*, are more enriched in *38C* than in *42AB* (Fig. 4E; Supplementary Fig. S3A). Moreover, ovary-specific *Sox102F* cluster harbors a single autonomous transposon, *Tc1-2,* which has higher activity in ovary (Fig. 4F; Supplementary Fig. S3A). These examples show that differential expression of piRNA clusters in two sexes often matches the differential activities of TEs they control, supporting the notion that piRNA clusters are employed in a sex-specific fashion to cope with distinct TE landscape in male and female gonads.

piRNA clusters composed of local repeats produce piRNAs that target host genes

Our analysis indicated that 13.8% of testis piRNAs might potentially be involved in targeting host genes as they can be mapped to protein-coding genes in antisense orientation with a small number (0 to 3) of mismatches between piRNA and gene sequences (Fig. 1B). To understand the genomic origin of these piRNAs, we further analyzed sequence compositions of piRNA clusters. We found that two clusters, *Hsp70B* and *petrel*, both of which contain local repeats, generate piRNAs that have the potential to target host genes.

The *Hsp70B* cluster spans ~35Kb between two paralogous *Hsp70B* genes on chr3R, and it is active in both ovary and testis (Fig. 5A). The body of *Hsp70B* cluster contains several TEs. Even though there are piRNAs mapping to these TEs, they can be mapped elsewhere in the genome as well, rendering it impossible to be certain that they originate from *Hsp70B* locus. In fact, this cluster was previously identified through the presence of uniquely mapped piRNAs from flanking non-repetitive genes (*16*). However, our algorithm that takes into account local repeats revealed piRNAs generated from a ~354bp local repeat at *Hsp70B* locus, which occupies nearly all inter-transposon space within this cluster. Importantly, these piRNAs mapped exclusively to this local repeat at *Hsp70B* cluster but nowhere else in the genome. Every copy of this local repeat is flanked by sequences of *copia2* retrotransposon and corresponds to a tandem repeat at *Hsp70B* described ~40 years ago (*25*, *26*). The entire repeat-rich region between two *Hsp70B* genes is present in *D. melanogaster* but not in either of its sibling species *D. simulans* or *D. mauritiana* (*27*, *28*), suggesting a recent evolutionary origin. Intriguingly, these repeats have a ~92% sequence identity to an exon of the *nod* gene, which encodes a kinesin-like protein necessary for chromosome segregation during meiosis

(*29–31*). *Hsp70B* cluster generates piRNAs that are antisense to *nod* with a 91.3% averaged nucleotide identity to it. This level of sequence similarity is close to that between *Suppressor of Stellate* piRNAs and their *Stellate* targets, the first known case of piRNA repression (*1, 2*), suggesting that piRNAs produced from *Hsp70B* locus might be able to repress the *nod* gene.

The second locus producing piRNAs that might target host genes is the novel piRNA cluster *petrel* on Y chromosome, which is only present in XY males (Fig. 5B). This cluster spans more than 200Kb and includes two loci duplicated from chr2L and chrX, respectively, that contain almost the entire *CG12717* gene (which encodes a SUMO protease) and small parts of *Paics* (which encodes an enzyme involved in purine biogenesis) and *ProtA* (which encodes protamine, a sperm chromatin protein) (*32*). These gene-homologous sequences are further duplicated locally on Y to over 20 copies and take up nearly all space in between TEs at *petrel* locus (Fig. 5B; Supplementary Fig. S5B). However, these gene-related sequences likely do not retain coding potentials as they are frequently interrupted by TE sequences. *petrel* locus produces piRNAs antisense to *CG12717, Paics and ProtA* genes, with averaged levels of nucleotide identity 92.5%, 93.9%, and 91.0%, respectively. Together, two piRNA clusters, *Hsp70B* and *petrel*, encode piRNAs with the potential to target both TEs and host genes.

We quantified expression of piRNAs antisense to *nod*, *CG12717, Paics* and *ProtA* genes from these two clusters. Even though these piRNAs all possess over 90% identity to their

putative targets, their abundances differ dramatically (Fig. 5C). *CG12717* gene is targeted by abundant piRNAs (4,310 RPM), comparable to the 15th most targeted TE family in testis. piRNAs against *nod* are expressed at 813 RPM (~fivefold less compared to *CG12717),* while the levels of piRNA against *Paics* or *ProtA* are low (both ~50 RPM). In addition, nearly the entire length of *CG12717* gene is targeted by piRNAs, whereas only small parts of *nod, Paics* and *ProtA* are targeted. These findings suggest that *CG12717* and *nod* might be regulated by piRNAs in testis.

piRNA-guided repression of SUMO protease *CG12717*/*pirate* during spermatogenesis

To examine the role of piRNAs in gene regulation, we employed RNA-seq to analyze expression of host genes in testes of three different piRNA pathway mutants: *aub*, *zuc* and *spn-E* (*23, 33–35*). Transcriptome profiling revealed that only two genes, *CG12717* and *frtz,* exhibited ≥2-fold up-regulation in all three piRNA pathway mutants (Fig. 6A). Unlike *CG12717*, there are very few, if any, antisense piRNAs targeting *frtz*, so its up-regulation likely reflects a secondary phenotype following TE de-repression. Strikingly, expression of *CG12717* increased more than 10-fold in all three mutants (Fig. 6B), indicating that it is indeed strongly repressed by the piRNA pathway. Meanwhile, we observed no statistically significant up-regulation of *nod, Paics* or *ProtA* in these three mutants (Fig. 6B), correlating with fewer piRNAs against these genes than *CG12717* (Fig. 5C). Transcriptome profiling thus identifies *CG12717* as a target of piRNA silencing and suggests that abundant antisense piRNAs with high target coverage might be required for efficient silencing.

To further examine *CG12717* expression, we performed RNA *in situ* hybridization chain reaction (*in situ* HCR). Expression of *CG12717* is very low in control testis, but it was significantly increased in testes of *aub*, *zuc* and *spn-E* mutants, establishing this gene as a *bona fide* gene target of the piRNA pathway (Fig. 6C). We also found strongly elevated *CG12717* expression in testes of XO males that have an intact piRNA pathway but lack Y chromosome (Fig. 6D), confirming that *CG12717*-silencing piRNAs are encoded on Y chromosome. Consistent with the Y-linkage of piRNAs against *CG12717*, it is silenced in testes but expressed in ovaries (Fig. 6E). When de-repressed, *CG12717* is specifically expressed in differentiating spermatocytes, but not in germline stem cells or mitotic spermatogonia. Interestingly, *Stellate* is expressed at the same stage when the silencing by *Su(Ste)* piRNAs is removed (*36*).

piRNA-guided cleavage of target RNAs often triggers the production of secondary piRNAs from target RNAs in a process dubbed ping-pong cycle (*3*). Examination of piRNA sequences revealed abundant piRNAs derived from the entire length of *CG12717* mRNAs (Fig. 5C). In contrast, we found few piRNAs processed from transcripts of *nod*, *Paics* or *ProtA.* Furthermore, sense piRNAs derived from *CG12717* mRNAs and antisense piRNAs produced from *petrel* locus demonstrated a strong ping-pong signature (Z_{10}=16.8; Fig. 6F), characteristic of active ping-pong cycle. This finding suggests the direct cleavage of *CG12717* transcripts guided by *petrel* piRNAs. Following the generation of secondary piRNAs by ping-pong, some target RNAs continue to be processed into tail-to-head strings of phased piRNAs dubbed trailing piRNAs (*18, 19*). We observed statistically significant

phasing signature among *CG12717*-derived sense piRNAs (Z_1=3.2; Supplementary Fig. S5C), and for a quarter of the secondary piRNAs, we could identify trailing piRNAs following the ping-pong sites, consistent with *CG12717* being a *bona fide* piRNA target.

To gain further confidence in piRNA-guided cleavage of *CG12717* transcripts, we performed degradome-seq to profile cellular RNAs that bear 5' monophosphate, which include 3' products of piRNA-guided cleavage. Analysis of testis degradome revealed that both *Stellate* and *CG12717* are enriched among degradome fragments relative to their expression levels measured by polyA+ RNA-seq, consistent with both genes being subject to piRNA-guided cleavage (Fig. 6G). In contrast, fragments from *nod*, *Paics* and *ProtA* were not enriched in the degradome library. In total, we obtained 354 degradome reads from two replicates which correspond to 19 unique 5' ends that are derived from *CG12717* mRNAs (Fig. 6H). Importantly, 39% (138/354) of these reads have corresponding antisense piRNAs that overlap 10nt 5'-to-5' and thus might be responsible for cleavage at these sites (Fig. 6H). Together, the presence of *CG12717*-derived sense piRNAs and identification of piRNA-guided cleavage products place *CG12717* mRNA as a direct target of piRNAs in testis. As our results indicate that expression of *CG12717*, a SUMO protease gene related to Ulp2 in yeast (*37*), is strongly repressed by piRNAs in testis, we propose to name it *pirate* (*pira*) for "piRNA target in testis".

Evolution of *pirate* and *pirate*-targeting small RNAs

To understand how piRNA-mediated gene regulation of *pira* has evolved, we performed a tblastn search using *Drosophila melanogaster pira* gene against genomes of other *Drosophila* species. We found multiple copies of *pira*-related sequences in genomes of *Drosophila simulans* species complex (*D. simulans*, *D. sechellia* and *D. mauritiana*) (Fig. 7A), but not in more distantly related species like *D. erecta* or *D. yakuba*. Similar to *petrel* locus in *D. melanogaster*, these *pira*-related sequences reside in TE-rich regions (either peri-centromeric heterochromatin or unassigned scaffolds) in *D. simulans* species complex. While all *pira*-related sequences are exclusively located at *petrel* on Y chromosome of *D. melanogaster*, *pira*-homologous sequences can be found on different chromosomes in genomes of *D. simulans* species complex. For instance, in *D. mauritiana*, *pira*-homologous sequences can be found on at least chrY, chrX, chr3L and chr3R (Fig. 7B). Therefore, duplications of *pira*-related sequences into heterochromatin have occurred in all four species.

To investigate whether heterochromatic *pira*-homologous sequences produce small RNAs in testes of other species, we analyzed published small RNA datasets from testes of *D. simulans* and *D. mauritiana* (*21*, *38*). We found no small RNAs mapping to the orthologous *pira* gene in *D. simulans* testes, but abundant ones in *D. mauritiana* testes (Fig. 7C). Unexpectedly, unlike 23–29nt *pira*-mapping piRNAs in *D. melanogaster*, *pira*-mapping small RNAs in *D. mauritiana* are mostly 21nt long, indicating that they are endo-siRNAs. These endo-siRNAs have on average 93.5% identity with the *D. mauritiana pira* gene. Notably, similar to other dual-strand piRNA clusters described in *D. melanogaster* ovaries (*12*, *39*, *40*), *petrel* in *D. melanogaster* testes also generates *pira*-mapping endo-siRNAs, though much less abundant

than 23–29nt piRNAs (Fig. 7C). Examination of heterochromatic, *pira*-homologous sequences in *D. mauritiana* genome revealed that most of them are arranged head-to-tail (Fig. 7B). However, there are four instances where *pira*-homologous sequences are arranged head-to-head (Fig. 7B, D), which could potentially generate hairpin RNAs (hpRNAs), the preferred substrate for processing into endo-siRNAs by Dicer. Thus, targeting of *pira* by small RNAs in testis seems to be conserved in two *Drosophila* species. While *pira* is repressed mostly by piRNAs in *D. melanogaster*, it is targeted nearly exclusively by endo-siRNAs in *D. mauritiana*, suggesting two related but distinct regulation strategies employed in sibling species that diverged less than 3 million years ago.

In addition to *pira,* there is another Ulp2-like SUMO protease gene, *verloren* (*velo*) in *D. melanogaster* genome. According to modENCODE data, both genes are expressed throughout the body across development, except that *pira* has a very low expression level in testis (*41*). *pira* and *velo* are paralogs whose homologous domains share 75% nucleotide identity (Supplementary Fig. S5A). In agreement with the sequence similarity, functions of Pira and Velo in SUMO deconjugation pathway were shown to be partially redundant (*37*). Phylogenetic analysis showed that, while *velo* is found at syntenic locations throughout the *Drosophila* genus, *pira* is much younger and was only born after the split of *D. melanogaster* and *ananassae* species subgroups (Fig. 7A). These results indicate that *pira* and *velo* have evolved from a common ancestor gene, via inter-chromosomal duplication.

Considering the 75% nucleotide identity between the parts of *pira* and *velo* genes in *D. melanogaster*, *pira*-targeting *petrel* piRNAs have a potential to target *velo* transcripts. However, we found that none of the *pira*-antisense piRNAs can be mapped to *velo* transcript perfectly. Moreover, ~200-fold fewer piRNAs have a potential to target *velo* with one to three mismatches. Transcriptome profiling in testes of *aub, zuc* and *spn-E* mutants showed that, unlike *pira*, *velo* is not repressed by piRNAs (Fig. 6B). In addition, while *pira* is only expressed in ovaries, *velo* is expressed in both testes and ovaries and, in fact, has a higher expression level in testes (Fig. 6E). These results show that Y-linked *petrel* piRNAs repress specifically *pira*, but not its paralog, *velo*, suggesting that a high degree of complementarity is required for efficient piRNA silencing. Therefore, piRNAs distinguish closely related paralogs with high sequence similarity to achieve sex- and paralog-specific gene regulation.

Taken together, our results allowed us to reconstruct the evolutionary history of two paralogous, Ulp2-like SUMO protease genes. First, the *pira* gene was born via inter-chromosomal duplication after the split of *D. melanogaster* and *ananassae* species subgroups. This then permitted the differentiation of *velo* and *pira* functions, though these two genes remain in part functionally redundant in *D. melanogaster* (*37*). Next, divergence between *pira* and *velo* sequences created an opportunity for paralog-selective gene regulation by small RNA-guided mechanisms. This was achieved by duplications of *pira* sequences into heterochromatin in genomes of *D. melanogaster* and *D. simulans* species complex. It is plausible that, initially, heterochromatic, *pira*-homologous sequences did not play a role in gene regulation, as illustrated by the absence of *pira*-mapping small RNAs in *D. simulans.*

However, subsequent expansion and interaction with TE sequences might have enabled the evolution of two distinct repression mechanisms, via production of *pira*-targeting piRNAs and endo-siRNAs, that dominated in *D. melanogaster* and *D. mauritiana*, respectively. Repression of *pira* by Y-linked piRNAs led to its specific repression in *D. melanogaster* testis, implicating the piRNA pathway in establishing distinct expression patterns of closely related paralogs after gene duplication.

DISCUSSION

Previous studies systematically analyzed piRNA profiles in female gonads of *D. melanogaster,* revealing an essential role of piRNAs in regulation of many TEs (*3*, *42*, *43*). However, these studies only provided a single snapshot of the relationship between TE and piRNA defense system, as they are insufficient to understand how the piRNA program might adapt to changing TE repertoire and different levels of their expression. To this end, several studies explored the piRNA pathway in other species of *Drosophila* (*43–45*). These studies revealed that piRNA profiles are different across species, suggesting an adaptation of the defense mechanism to distinct challenges. However, drastic differences in both TE contents and piRNA cluster sequences even among closely related *Drosophila* species (*43*, *46*, *47*) make it difficult to disentangle different factors that sculpt species-specific piRNA programs. Here, we examined TE expression in males and females of the same species, revealing strong differences in TE activities between sexes. This allowed us to compare piRNA programs in two sexes with similar genomic contents (except Y chromosome).

Another obstacle to understanding responses of the piRNA program to TEs is properly assessing TE expression. *D. melanogaster* genome includes over 100 different TE families whose expression levels can be measured by standard methods such as RNA-seq. However, TE expression in wild-type animals is greatly suppressed by the piRNA pathway (>100-fold for some families) (*11*). Therefore, in order to understand true expression potentials of TEs, it is necessary to study their expression upon removal of piRNA silencing, which is difficult to do in species other than model organisms like *D. melanogaster*. In this work, we examined the TE expression in piRNA pathway mutants, revealing genuine potentials of TE expression in both sexes. Combined analysis of TE and piRNA expression showed responses of the piRNA program to distinct TE expression profiles in two sexes.

Analysis of the genomic origin of piRNAs represents an important but challenging task. As piRNA sequences are short (23–29nt) and often derive from repetitive genomic regions, a large fraction of sequenced piRNA reads can be mapped to multiple genomic loci, preventing an unambiguous assignment of their origin. Accordingly, algorithms employed in previous studies only used the small fraction of piRNA reads that can be mapped to the genome at single unique positions to identify genomic regions that generate piRNAs. We took advantage of the fact that some genomic repeats are local, i.e., they reside within one genomic region and are absent in the rest of the genome, to develop a new algorithm for piRNA cluster definition and analysis (Fig. 3A and Supplementary Fig. S4). This approach was successful in identifying new piRNA clusters. Furthermore, it also provided a more accurate quantification of the piRNA cluster expression. We found that *Hsp70B* cluster generates

piRNAs against the *nod* gene. In addition, we discovered a novel cluster, *petrel*, on Y chromosome that generates piRNAs against three host genes and ensures the strong silencing of SUMO protease, *CG12717/pira,* during spermatocyte differentiation.

Our identification of the novel *petrel* locus on Y expanded known functions of entirely heterochromatic Y chromosome (Fig. 3F and Fig. 5B). Three functionalities have been assigned to Y by the early 1980s (*24*). First, together with X chromosome, Y encodes rDNA loci that express rRNAs and mediates homolog pairing. Second, Y encodes six protein-coding genes, so-called "fertility factors", whose protein products are required for completion of spermatogenesis. Finally, Y chromosome harbors the *Su(Ste)* locus that generates piRNAs to suppress *Stellate* genes to safeguard normal spermatogenesis (*1*, *2*). A handful of new protein-coding genes were discovered on Y in the past two decades (*48, 49*), however, many of them appeared dispensable. Our finding that Y chromosome encodes a novel piRNA cluster and produces piRNAs to regulate expression of the *pira* gene assigns a new function to Y chromosome.

Sexual dimorphism of TE expression and TE-silencing piRNA programs

D. melanogaster is an excellent model to study TE regulations and host-TE interactions, as its genome harbors many TE families that are transcriptionally and transpositionally active, generating new insertions in the population (*47*). As ovaries and testes have complex and distinct tissue compositions, expression levels measured by RNA-seq and small RNA-seq

cannot be used directly for comparison of cellular concentrations of transposon transcripts and piRNAs in male and female germline. Therefore, we have compared rank orders of transposon and piRNA expression in two sexes as well as fold-changes in their levels upon disruption of the piRNA pathway between sexes. Our results indicate that expression of both TEs and piRNAs is sexually dimorphic. The majority of TE families are strongly expressed in ovaries, though some TEs are more active in testes. In line with this, our results indicate a stronger TE-silencing piRNA program in female gonads (Fig. 2).

For TEs to be evolutionarily successful, they need to evolve strategies to maximize their chance to be inherited and expanded through generations. For example, TEs often hijack germline gene expression programs to be preferentially active in germ cells. Germline-biased expression leaves the choice of expression to either female or male germline, or both. Importantly, the two sexes employ distinct evolutionary strategies and have different contributions towards the zygote. While the major contribution of sperm is its genome, oocyte contributes large amounts of yolk, various protein factors, RNAs and organelles such as mitochondria, in addition to its genome. This sexual asymmetry in their contributions to the next generation has important implications for reproduction strategies of TEs. TEs active in the male germline need to complete the entire life cycle from transcription to genomic insertion before sperm maturation, in order to propagate. In contrast, once transcribed, TEs active during oogenesis could finish their life cycle in the zygote after fertilization, as long as transcribed TE transcripts are deposited into the oocyte. The latter strategy is also used by mammalian L1 retrotransposon that is expressed during gametogenesis, but genomic

insertions might occur later during early embryogenesis (*50*). Thus, the expression bias towards ovaries observed for most TEs can be explained by an advantage for their proliferation, specifically, the extended window to finish their life cycle, in female germline.

There are a few TEs that bias testis for expression, suggesting that there are likely male-specific vulnerabilities exploitable by these elements. For example, male germ cells use a testis-specific gene expression machinery (e.g., tTAF and tMAC) to transcribe meiotic and post-meiotic genes (*51*, *52*). TEs might exploit this tissue-specific transcriptional machinery to enable their sex-biased expression. It will be important in the future to uncover molecular mechanisms underlying differentially expressed TEs between sexes.

Analysis of piRNA profiles in testis and ovary indicates that piRNA programs have adapted to sex-biased TE expression (Fig. 2). The most striking example is the nearly exclusive expression of telomeric TEs and corresponding antisense piRNAs in the female germline. Our results suggest that differential expression of piRNA clusters in two sexes together with differential TE-targeting capacity of each cluster contributes to the sex-specific, TE-targeting piRNA program. We found that piRNA cluster expression is sexually dimorphic. Besides the *Su(Ste)* locus, we identified another major cluster on Y chromosome that is only active in XY males. However, sex-biased expression is not restricted to Y-linked clusters, as many X-linked and autosomal clusters have differential activities between sexes as well. Besides differential expression, genomic analysis showed differences in piRNA cluster TE contents, suggesting that different piRNA clusters are, to some extent, specialized to target different

sets of TEs. Importantly, sex bias in cluster expression and their TE-targeting potentials are linked: clusters preferentially targeting ovary-biased TEs are more active in ovary, while testis-biased clusters tend to target testis-biased TEs (Fig. 4). Hence, piRNA clusters appear to be employed differentially by two sexes to counteract specific TE threats they face. What determines the differential expression of piRNA clusters between sexes awaits future studies. Previous work suggests that TE promoters embedded in piRNA clusters retain their activities (*16*). Contribution of TE promoters to piRNA precursor transcription from piRNA clusters might explain the correlation between expression of clusters and their TE targets.

Satellite DNA as target of piRNA silencing

Satellite DNAs can be classified as either simple or complex satellites based on the length of repeating units, and they occupy large portions of *Drosophila* genome, particularly at peri-centromeric and sub-telomeric regions (*53–56*). We found piRNAs expressed from three major families of complex satellites: sub-telomeric *HETRP*/*TAS*, *Responder (Rsp)*, and *SAR/1.688* (including 359bp). In fact, piRNAs can be mapped to both strands of complex satellites in gonads of both sexes, and they often possess ping-pong signature (Fig. 1C). Thus, our results expand the previous observation of piRNAs mapping to one strand of *Rsp* (*57*) and establish complex satellites as dual-strand piRNA clusters and potential targets of piRNA silencing in *Drosophila* germline of both sexes. Our analysis was focused on complex satellites, as simple satellite repeats are still largely intractable to sequencing technologies today (*58*). However, a recent study reported that transcripts from AAGAG simple satellite repeats regulate heterochromatin in male germline and are required for male fertility (*59*). It

will be interesting to determine whether simple satellites produce piRNAs and, if so, whether their piRNA production is required for male fertility.

piRNAs loaded onto the nuclear Piwi protein guide heterochromatin assembly (*60–63*). For this reason, satellite piRNAs might play a role in establishing germline heterochromatin, similar to heterochromatin formation guided by siRNAs in fission yeast (*64*, *65*). While the function of complex satellites remains mostly elusive, *Rsp* has been implicated in a meiotic drive system called segregation distortion (*56*, *66*, *67*). During male meiosis, the *Segregation Distorter (SD)* allele enhances its own transmission to haploid cells at the cost of wild-type (SD^+) allele in SD/SD^+ heterozygous males, violating Mendelian law of inheritance. Importantly, segregation distortion requires the presence of a sufficient number of *Rsp* satellite repeats *in trans*. Though described more than 60 years ago (*68*), the molecular mechanism of segregation distortion remains unknown. Intriguingly, mutations of *aubergine (aub)*, a PIWI protein, were found to be enhancers of *SD* (*69*). Together with our data, these data suggest that piRNA pathway may play a role in segregation distortion during spermatogenesis.

Regulation of host genes by piRNAs

Though the central and conserved function of piRNA pathway seems to be TE repression, other functions were also described in several organisms (reviewed in Ozata et al., 2019). The role of piRNAs in regulating host gene expression is particularly intriguing and remains

somewhat controversial. The first described piRNAs, *Su(Ste)* piRNAs, silence the expression of *Stellate* genes (*1*, *2*). However, *Stellate* genes and their piRNA suppressors appear to resemble selfish toxin-antitoxin systems rather than representing an example of host gene regulation (*71*). Since the discovery of piRNA pathway, there have been several studies reporting host protein-coding genes regulated by *Drosophila* piRNAs (reviewed in Rojas-Ríos and Simonelig, 2018). In this work, we analyzed the ability of *Drosophila* piRNAs to regulate host genes in testes, by examining gene-targeting piRNAs and changes in host gene expression across three piRNA pathway mutants. We found piRNAs targeting four host genes: *nod* (a kinesin-like protein), *CG12717/pira* (a SUMO protease), *Paics* (a metabolic enzyme) and *ProtA* (a sperm chromatin protein). These four genes are targeted by antisense piRNAs generated from two piRNA clusters that contain sequence homology to them. However, only one of the four, *CG12717/pira*, is substantially repressed (over 10-fold) by piRNAs (Fig. 6). As *pira*-silencing piRNAs are encoded on Y chromosome and thus only expressed in males, they are responsible for differential expression of *pira* in two sexes. Indeed, in wild-type files, *pira* is specifically silenced in male gonads while highly expressed in female counterparts. Thus, our results establish the ability of piRNAs to repress host protein-coding genes, and, at the same time, suggest that this role is likely restricted to a small number of genes.

Our results indicate that piRNA-guided repression of host genes requires a sufficient number of targeting piRNAs. While all four genes are targeted by piRNAs with similar levels of sequence identity (91–94%, i.e., about two mismatches per piRNA), the abundance of piRNAs

against each gene differs drastically. There are much more *pira*-targeting piRNAs than the other three gene targets, at a level comparable to the 15th most targeted TE. Furthermore, while *pira* is targeted along almost the entire length, only small regions of other three genes are targeted by piRNAs. These differences in piRNA abundance and distribution of target sites could explain strong repression of *pira*, in contrast to the other three genes. It is possible that these genes are still regulated by piRNAs at specific stages, the question that remains to be further investigated. Importantly, abundant *pira*-silencing piRNAs do not repress the *pira* paralog, *velo*, that has a 75% sequence identity with *pira*, indicating that a high complementarity between piRNA and target may be important for efficient silencing. In agreement with these, a previous report indicated that a similar level of sequence identity (~76%) is insufficient for the silencing of *vasa* by *AT-chX* piRNAs (*21*). Therefore, both high expression and high complementarity with targets might be required for efficient piRNA silencing in *D. melanogaster*.

This conclusion is important for analyzing the potential of piRNAs to repress host protein-coding genes. Unlike miRNAs, sequences of piRNAs are extremely diverse. Accordingly, if mismatches between piRNA and its target are well tolerated, a large number of cellular mRNAs should be targeted and repressed by piRNAs. Indeed, some host genes were proposed to be repressed by a few piRNA species that have multiple mismatches to mRNA sequences (*73–76*). Our results suggest that such a spurious targeting by individual piRNAs is unlikely to cause repression. In fact, a high threshold for efficient target repression might

permit production of diverse piRNA sequences against genuine targets such as TEs, without unintended interference with host gene expression.

The role of piRNA in evolution

Analysis of *pira* repression revealed a remarkable picture of evolutionary innovation (Fig. 7). piRNA-dependent repression of *pira* occurs in *D. melanogaster* but not in its sibling species, suggesting its rather recent origin. Efficient silencing of *pira* is linked to the presence of multiple copies of *pira*-homologous sequences in a piRNA cluster inside heterochromatin. Interestingly, duplications of *pira* sequences into, and their expansion within, heterochromatin can be found in three closely related species of *D. simulans* complex, in addition to *D. melanogaster*. However, distribution and copy number of *pira*-related sequences differ among these four species. In fact, both the *petrel* locus that generates *pira*-silencing piRNAs and its two flanking protein-coding genes, *FDY* and *Mst77Y*, evolved after the split of *D. melanogaster* and *D. simulans* species complex (*32*, *49*, *77*), suggesting that the entire locus is unique to *D. melanogaster*. Furthermore, no small RNAs are generated from heterochromatic *pira* sequences in *D. simulans*, while endo-siRNAs are made against *pira* in *D. mauritiana*. The neutral theory of molecular evolution provides the most parsimonious interpretation of these results. This theory suggests that the initial duplication of *pira* sequences into heterochromatin might have been a random event that did not play a role in regulating the ancestral *pira* gene. However, subsequent evolution of *pira*-related sequences inside heterochromatin gave rise to two different modes of regulations, piRNA and endo-siRNA, in two different but closely related species. Emergence of small RNA-

mediated repression was probably facilitated by the fact that *pira* itself was recently evolved and retains partially redundant functions with its paralog, *velo* (*37*), allowing independent regulation of two paralogs.

The evolutionarily innovative role of piRNAs in regulating host genes in *Drosophila* has interesting parallels in other organisms. piRNA pathway was shown to regulate expression of *xol-1* gene involved in sex determination and dosage compensation in two different worm species, *C. elegans* and *C. briggsae*, that diverged more than 50 million years ago (*78*). On the other hand, pachytene piRNAs expressed during spermatogenesis in mammals evolved very fast and are generally poorly conserved (*79*). The function of pachytene piRNAs is under active debate as no obvious targets can be easily discerned by analysis of their sequences (*80–82*). Recently, knock-out of one pachytene piRNA cluster led to unexpected conclusion that a small fraction of piRNAs promote biogenesis from other piRNA clusters and regulate the expression of a few host genes, while the vast majority do not target any transcripts (*83*). Thus, mammalian pachytene piRNAs can be considered a selfish system that occasionally involves in regulation of the host gene expression. Species-specific regulation of host genes by piRNAs in both *Drosophila* and mouse suggests that piRNA pathway is used in evolution to create innovation in gene regulatory networks that might contribute to speciation. More generally, piRNAs might promote the evolvability of animal species. Though it is difficult to establish the function of any molecular mechanism in evolution, this proposal makes a testable prediction that host genes repressed by piRNAs

differ even among closely related species. Future studies in non-model organisms will shed light on the role of piRNAs in evolution and speciation.

REFERENCES

1. A. A. Aravin, N. M. Naumova, A. V. Tulin, V. V. Vagin, Y. M. Rozovsky, V. A. Gvozdev, Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Current Biology* **11**, 1017–1027 (2001).

2. V. V. Vagin, A. Sigova, C. Li, H. Seitz, V. Gvozdev, P. D. Zamore, A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).

3. J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G. J. Hannon, Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128**, 1089–1103 (2007).

4. S. Houwing, L. M. Kamminga, E. Berezikov, D. Cronembold, A. Girard, H. van den Elst, D. V. Filippov, H. Blaser, E. Raz, C. B. Moens, R. H. A. Plasterk, G. J. Hannon, B. W. Draper, R. F. Ketting, A Role for Piwi and piRNAs in Germ Cell Maintenance and Transposon Silencing in Zebrafish. *Cell* **129**, 69–82 (2007).

5. S. Houwing, E. Berezikov, R. F. Ketting, Zili is required for germ cell differentiation and meiosis in zebrafish. *EMBO J* **27**, 2702–2711 (2008).

6. L. M. Kamminga, M. J. Luteijn, M. J. den Broeder, S. Redl, L. J. T. Kaaij, E. F. Roovers, P. Ladurner, E. Berezikov, R. F. Ketting, Hen1 is required for oocyte development and piRNA stability in zebrafish. *EMBO J* **29**, 3688–3700 (2010).

7. W. Deng, H. Lin, Miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* **2**, 819–830 (2002).

8. S. Kuramochi-Miyagawa, T. Kimura, T. W. Ijiri, T. Isobe, N. Asada, Y. Fujita, M. Ikawa, N. Iwai, M. Okabe, W. Deng, H. Lin, Y. Matsuda, T. Nakano, Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**, 839–849 (2004).

9. M. A. Carmell, A. Girard, H. J. G. van de Kant, D. Bourc'his, T. H. Bestor, D. G. de Rooij, G. J. Hannon, MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* **12**, 503–514 (2007).

10. H. Lin, A. C. Spradling, A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. *Development* **124**, 2463–2476 (1997).

11. M. F. ElMaghraby, P. R. Andersen, F. Pühringer, U. Hohmann, K. Meixner, T. Lendl, L. Tirian, J. Brennecke, A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production. *Cell* **178**, 964–979.e20 (2019).

12. B. Czech, C. D. Malone, R. Zhou, A. Stark, C. Schlingeheyde, M. Dus, N. Perrimon, M. Kellis, J. A. Wohlschlegel, R. Sachidanandam, G. J. Hannon, J. Brennecke, An endogenous small interfering RNA pathway in Drosophila. *Nature* **453**, 798–802 (2008).

13. J. Wen, H. Duan, F. Bejarano, K. Okamura, L. Fabian, J. A. Brill, D. Bortolamiol-Becet, R. Martin, J. G. Ruby, E. C. Lai, Adaptive Regulation of Testis Gene Expression

and Control of Male Fertility by the Drosophila Hairpin RNA Pathway. *Molecular Cell* **57**, 165–178 (2015).

14. P. Chen, Y. Luo, A. A. Aravin. 2020. RDC complex executes a dynamic piRNA program during *Drosophila* spermatogenesis to safeguard male fertility. bioRxiv doi: 10.1101/2020.08.25.266643

15. C. Klattenhoff, H. Xi, C. Li, S. Lee, J. Xu, J. S. Khurana, F. Zhang, N. Schultz, B. S. Koppetsch, A. Nowosielska, H. Seitz, P. D. Zamore, Z. Weng, W. E. Theurkauf, The Drosophila HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters. *Cell* **138**, 1137–1149 (2009).

16. F. Mohn, G. Sienski, D. Handler, J. Brennecke, The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in Drosophila. *Cell* **157**, 1364–1379 (2014).

17. Z. Zhang, J. Wang, N. Schultz, F. Zhang, S. S. Parhad, S. Tu, T. Vreven, P. D. Zamore, Z. Weng, W. E. Theurkauf, The HP1 Homolog Rhino Anchors a Nuclear Complex that Suppresses piRNA Precursor Splicing. *Cell* **157**, 1353–1363 (2014).

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

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

20. D. Handler, K. Meixner, M. Pizka, K. Lauss, C. Schmied, F. S. Gruber, J. Brennecke, The Genetic Makeup of the Drosophila piRNA Pathway. *Molecular Cell* **50**, 762–777 (2013).

21. A. A. Kotov, V. E. Adashev, B. K. Godneeva, M. Ninova, A. S. Shatskikh, S. S. Bazylev, A. A. Aravin, L. V. Olenina, piRNA silencing contributes to interspecies hybrid sterility and reproductive isolation in Drosophila melanogaster. *Nucleic Acids Research* **47**, 4255–4271 (2019).

22. P. George, S. Jensen, R. Pogorelcnik, J. Lee, Y. Xing, E. Brasset, C. Vaury, I. V. Sharakhov, Increased production of piRNAs from euchromatic clusters and genes in Anopheles gambiae compared with Drosophila melanogaster. *Epigenetics Chromatin* **8**, 50 (2015).

23. K. M. Nishida, K. Saito, T. Mori, Y. Kawamura, T. Nagami-Okada, S. Inagaki, H. Siomi, M. C. Siomi, Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad. *RNA* **13**, 1911–1922 (2007).

24. M. Gatti, S. Pimpinelli, Cytological and genetic analysis of the Y chromosome of Drosophila melanogaster: I. Organization of the fertility factors. *Chromosoma* **88**, 349–373 (1983).

25. J. T. Lis, L. Prestidge, D. S. Hogness, A novel arrangement of tandemly repeated genes at a major heat shock site in D. melanogaster. *Cell* **14**, 901–919 (1978).

26. R. W. Hackett, J. T. LIs, DNA sequence analysis reveals extensive homologies of regions preceding hsp7O and af8 heat shock genes in Drosophila melanogaster. 5 (1981).

27. K. J. Livak, R. Freund, M. Schweber, P. C. Wensink, M. Meselson, Sequence organization and transcription at two heat shock loci in Drosophila. *Proceedings of the National Academy of Sciences* **75**, 5613–5617 (1978).

28. A. J. L. Brown, D. Ish-Horowicz, Evolution of the 87A and 87C heat-shock loci in Drosophila. *Nature* **290**, 677–682 (1981).

29. A. T. Carpenter, A meiotic mutant defective in distributive disjunction in Drosophila melanogaster. *Genetics* **73**, 393–428 (1973).

30. P. Zhang, B. A. Knowles, L. S. Goldstein, R. S. Hawley, A kinesin-like protein required for distributive chromosome segregation in Drosophila. *Cell* **62**, 1053–1062 (1990).

31. R. S. Hawley, W. E. Theurkauf, Requiem for distributive segregation: achiasmate segregation in Drosophila females. *Trends Genet.* **9**, 310–317 (1993).

32. M. Mendez-Lago, C. M. Bergman, B. de Pablos, A. Tracey, S. L. Whitehead, A. Villasante, A Large Palindrome With Interchromosomal Gene Duplications in the Pericentromeric Region of the D. melanogaster Y Chromosome. *Molecular Biology and Evolution* **28**, 1967–1971 (2011).

33. A. Schmidt, G. Palumbo, M. P. Bozzetti, P. Tritto, S. Pimpinelli, U. Schäfer, Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of Drosophila melanogaster. *Genetics* **151**, 749–760 (1999).

34. W. Stapleton, S. Das, B. D. McKee, A role of the Drosophila homeless gene in repression of Stellate in male meiosis. *Chromosoma* **110**, 228–240 (2001).

35. A. Pane, K. Wehr, T. Schüpbach, zucchini and squash Encode Two Putative Nucleases Required for rasiRNA Production in the Drosophila Germline. *Developmental Cell* **12**, 851–862 (2007).

36. A. A. Aravin, M. S. Klenov, V. V. Vagin, F. Bantignies, G. Cavalli, V. A. Gvozdev, Dissection of a Natural RNA Silencing Process in the Drosophila melanogaster Germ Line. *MCB* **24**, 6742–6750 (2004).

37. D. Berdnik, V. Favaloro, L. Luo, The SUMO protease Verloren regulates dendrite and axon targeting in olfactory projection neurons. *J. Neurosci.* **32**, 8331–8340 (2012).

38. C.-J. Lin, F. Hu, R. Dubruille, J. Vedanayagam, J. Wen, P. Smibert, B. Loppin, E. C. Lai, The hpRNA/RNAi Pathway Is Essential to Resolve Intragenomic Conflict in the Drosophila Male Germline. *Developmental Cell* **46**, 316–326.e5 (2018).

39. M. Ghildiyal, H. Seitz, M. D. Horwich, C. Li, T. Du, S. Lee, J. Xu, E. L. W. Kittler, M. L. Zapp, Z. Weng, P. D. Zamore, Endogenous siRNAs Derived from Transposons and mRNAs in Drosophila Somatic Cells. *Science* **320**, 1077–1081 (2008).

40. A. Le Thomas, E. Stuwe, S. Li, J. Du, G. Marinov, N. Rozhkov, Y.-C. A. Chen, Y. Luo, R. Sachidanandam, K. F. Toth, D. Patel, A. A. Aravin, Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev.* **28**, 1667–1680 (2014).

41. J. B. Brown, N. Boley, R. Eisman, G. E. May, M. H. Stoiber, M. O. Duff, B. W. Booth, J. Wen, S. Park, A. M. Suzuki, K. H. Wan, C. Yu, D. Zhang, J. W. Carlson, L. Cherbas, B. D. Eads, D. Miller, K. Mockaitis, J. Roberts, C. A. Davis, E. Frise, A. S. Hammonds, S. Olson, S. Shenker, D. Sturgill, A. A. Samsonova, R. Weiszmann, G. Robinson, J. Hernandez, J. Andrews, P. J. Bickel, P. Carninci, P. Cherbas, T. R. Gingeras, R. A. Hoskins, T. C. Kaufman, E. C. Lai, B. Oliver, N. Perrimon, B. R. Graveley, S. E. Celniker, Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**, 393–399 (2014).

42. C. Li, V. V. Vagin, S. Lee, J. Xu, S. Ma, H. Xi, H. Seitz, M. D. Horwich, M. Syrzycka, B. M. Honda, E. L. W. Kittler, M. L. Zapp, C. Klattenhoff, N. Schulz, W. E. Theurkauf, Z. Weng, P. D. Zamore, Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. *Cell* **137**, 509–521 (2009).

43. C. D. Malone, J. Brennecke, M. Dus, A. Stark, W. R. McCombie, R. Sachidanandam, G. J. Hannon, Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **137**, 522–535 (2009).

44. N. V. Rozhkov, A. A. Aravin, E. S. Zelentsova, N. G. Schostak, R. Sachidanandam, W. R. McCombie, G. J. Hannon, M. B. Evgen'ev, Small RNA-based silencing strategies for transposons in the process of invading Drosophila species. *RNA* **16**, 1634–1645 (2010).

45. B. Saint-Leandre, P. Capy, A. Hua-Van, J. Filée, piRNA and Transposon Dynamics in Drosophila: A Female Story. *Genome Biology and Evolution* **12**, 931–947 (2020).

46. E. Lerat, N. Burlet, C. Biémont, C. Vieira, Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. *Gene* **473**, 100–109 (2011).

47. R. Kofler, V. Nolte, C. Schlötterer, Tempo and Mode of Transposable Element Activity in Drosophila. *PLoS Genet.* **11**, e1005406 (2015).

48. A. Bernardo Carvalho, L. B. Koerich, A. G. Clark, Origin and evolution of Y chromosomes: Drosophila tales. *Trends in Genetics* **25**, 270–277 (2009).

49. F. J. Krsticevic, H. L. Santos, S. Januário, C. G. Schrago, A. B. Carvalho, Functional Copies of the *Mst77F* Gene on the Y Chromosome of *Drosophila melanogaster*. *Genetics* **184**, 295–307 (2010).

50. H. Kano, I. Godoy, C. Courtney, M. R. Vetter, G. L. Gerton, E. M. Ostertag, H. H. Kazazian, L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* **23**, 1303–1312 (2009).

51. M. Hiller, X. Chen, M. J. Pringle, M. Suchorolski, Y. Sancak, S. Viswanathan, B. Bolival, T.-Y. Lin, S. Marino, M. T. Fuller, Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development* **131**, 5297–5308 (2004).

52. E. L. Beall, P. W. Lewis, M. Bell, M. Rocha, D. L. Jones, M. R. Botchan, Discovery of tMAC: a Drosophila testis-specific meiotic arrest complex paralogous to Myb-Muv B. *Genes Dev.* **21**, 904–919 (2007).

53. T. Hsieh, D. Brutlag, Sequence and sequence variation within the 1.688 g/cm3 satellite DNA of Drosophila melanogaster. *J. Mol. Biol.* **135**, 465–481 (1979).

54. G. H. Karpen, A. C. Spradling, Analysis of subtelomeric heterochromatin in the Drosophila minichromosome Dp1187 by single P element insertional mutagenesis. *Genetics* **132**, 737–753 (1992).

55. A. R. Lohe, A. J. Hilliker, P. A. Roberts, Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. *Trends in Genetics* **9**, 379 (1993).

56. A. M. Larracuente, D. C. Presgraves, The Selfish *Segregation Distorter* Gene Complex of *Drosophila melanogaster*. *Genetics* **192**, 33–53 (2012).

57. K. Saito, K. M. Nishida, T. Mori, Y. Kawamura, K. Miyoshi, T. Nagami, H. Siomi, M. C. Siomi, Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev.* **20**, 2214–2222 (2006).

58. D. E. Khost, D. G. Eickbush, A. M. Larracuente, Single-molecule sequencing resolves the detailed structure of complex satellite DNA loci in *Drosophila melanogaster*. *Genome Res.* **27**, 709–721 (2017).

59. W. K. Mills, Y. C. G. Lee, A. M. Kochendoerfer, E. M. Dunleavy, G. H. Karpen, RNA from a simple-tandem repeat is required for sperm maturation and male fertility in Drosophila melanogaster. *eLife* **8**, e48940 (2019).

60. S. H. Wang, S. C. R. Elgin, Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 21164–21169 (2011).

61. G. Sienski, D. Dönertas, J. Brennecke, Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**, 964–980 (2012).

62. A. Le Thomas, A. K. Rogers, A. Webster, G. K. Marinov, S. E. Liao, E. M. Perkins, J. K. Hur, A. A. Aravin, K. F. Tóth, Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* **27**, 390–399 (2013).

63. N. V. Rozhkov, M. Hammell, G. J. Hannon, Multiple roles for Piwi in silencing Drosophila transposons. *Genes Dev.* **27**, 400–412 (2013).

64. I. M. Hall, G. D. Shankaranarayana, K.-I. Noma, N. Ayoub, A. Cohen, S. I. S. Grewal, Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232–2237 (2002).

65. T. A. Volpe, C. Kidner, I. M. Hall, G. Teng, S. I. S. Grewal, R. A. Martienssen, Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).

66. D. L. Hartl, Complementation analysis of male fertility among the segregation distorter chromosomes of Drosophila melanogaster. *Genetics* **73**, 613–629 (1973).

67. C. I. Wu, T. W. Lyttle, M. L. Wu, G. F. Lin, Association between a satellite DNA sequence and the Responder of Segregation Distorter in D. melanogaster. *Cell* **54**, 179–189 (1988).

68. L. Sandler, Y. Hiraizumi, I. Sandler, Meiotic Drive in Natural Populations of Drosophila Melanogaster. I. the Cytogenetic Basis of Segregation-Distortion. *Genetics* **44**, 233–250 (1959).

69. S. L. Gell, R. A. Reenan, Mutations to the piRNA Pathway Component *Aubergine* Enhance Meiotic Drive of Segregation Distorter in *Drosophila melanogaster*. *Genetics* **193**, 771–784 (2013).

70. D. M. Ozata, I. Gainetdinov, A. Zoch, D. O'Carroll, P. D. Zamore, PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet* **20**, 89–108 (2019).

71. A. A. Aravin, Pachytene piRNAs as beneficial regulators or a defense system gone rogue. *Nat. Genet.* **52**, 644–645 (2020).

72. P. Rojas-Ríos, M. Simonelig, piRNAs and PIWI proteins: regulators of gene expression in development and stem cells. *Development* **145**, dev161786 (2018).

73. K. Saito, S. Inagaki, T. Mituyama, Y. Kawamura, Y. Ono, E. Sakota, H. Kotani, K. Asai, H. Siomi, M. C. Siomi, A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. *Nature* **461**, 1296–1299 (2009).

74. J. Gonzalez, H. Qi, N. Liu, H. Lin, Piwi Is a Key Regulator of Both Somatic and Germline Stem Cells in the Drosophila Testis. *Cell Reports* **12**, 150–161 (2015).

75. J. D. Klein, C. Qu, X. Yang, Y. Fan, C. Tang, J. C. Peng, c-Fos Repression by Piwi Regulates Drosophila Ovarian Germline Formation and Tissue Morphogenesis. *PLoS Genet* **12**, e1006281 (2016).

76. P. Rojas-Ríos, A. Chartier, S. Pierson, M. Simonelig, Aubergine and pi RNA s promote germline stem cell self-renewal by repressing the proto-oncogene *Cbl*. *EMBO J* **36**, 3194–3211 (2017).

77. A. B. Carvalho, B. Vicoso, C. A. M. Russo, B. Swenor, A. G. Clark, Birth of a new gene on the Y chromosome of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **112**, 12450–12455 (2015).

78. W. Tang, M. Seth, S. Tu, E.-Z. Shen, Q. Li, M. Shirayama, Z. Weng, C. C. Mello, A Sex Chromosome piRNA Promotes Robust Dosage Compensation and Sex Determination in C. elegans. *Developmental Cell* **44**, 762–770.e3 (2018).

79. D. M. Özata, T. Yu, H. Mou, I. Gainetdinov, C. Colpan, K. Cecchini, Y. Kaymaz, P.-H. Wu, K. Fan, A. Kucukural, Z. Weng, P. D. Zamore, Evolutionarily conserved pachytene piRNA loci are highly divergent among modern humans. *Nat Ecol Evol* **4**, 156–168 (2020).

80. A. Aravin, D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino, P. Morris, M. J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J. J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan, T. Tuschl, A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207 (2006).

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

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

83. P.-H. Wu, Y. Fu, K. Cecchini, D. M. Özata, A. Arif, T. Yu, C. Colpan, I. Gainetdinov, Z. Weng, P. D. Zamore, The evolutionarily conserved piRNA-producing locus pi6 is required for male mouse fertility. *Nat Genet* **52**, 728–739 (2020).

84. H. M. T. Choi, M. Schwarzkopf, M. E. Fornace, A. Acharya, G. Artavanis, J. Stegmaier, A. Cunha, N. A. Pierce, Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145** (2018).

85. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

86. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525–527 (2016).

87. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

88. J. Thurmond, J. L. Goodman, V. B. Strelets, H. Attrill, L. S. Gramates, S. J. Marygold, B. B. Matthews, G. Millburn, G. Antonazzo, V. Trovisco, T. C. Kaufman, B. R. Calvi, the FlyBase Consortium, N. Perrimon, S. R. Gelbart, J. Agapite, K. Broll, L. Crosby, G. dos Santos, D. Emmert, L. S. Gramates, K. Falls, V. Jenkins, B. Matthews, C.

Sutherland, C. Tabone, P. Zhou, M. Zytkovicz, N. Brown, G. Antonazzo, H. Attrill, P. Garapati, A. Holmes, A. Larkin, S. Marygold, G. Millburn, C. Pilgrim, V. Trovisco, P. Urbano, T. Kaufman, B. Calvi, B. Czoch, J. Goodman, V. Strelets, J. Thurmond, R. Cripps, P. Baker, FlyBase 2.0: the next generation. *Nucleic Acids Research* **47**, D759–D765 (2019).

89. W. Wang, M. Yoshikawa, B. W. Han, N. Izumi, Y. Tomari, Z. Weng, P. D. Zamore, The Initial Uridine of Primary piRNAs Does Not Create the Tenth Adenine that Is the Hallmark of Secondary piRNAs. *Molecular Cell* **56**, 708–716 (2014).

90. A. D. Yates, P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, J. C. Marugán, C. Cummins, C. Davidson, K. Dodiya, R. Fatima, A. Gall, C. G. Giron, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, I. Lavidas, T. Le, D. Lemos, J. G. Martinez, T. Maurel, M. McDowall, A. McMahon, S. Mohanan, B. Moore, M. Nuhn, D. N. Oheh, A. Parker, A. Parton, M. Patricio, M. P. Sakthivel, A. I. Abdul Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, M. Sycheva, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, B. Flint, A. Frankish, S. E. Hunt, G. IIsley, M. Kostadima, N. Langridge, J. E. Loveland, F. J. Martin, J. Morales, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, S. J. Trevanion, F. Cunningham, K. L. Howe, D. R. Zerbino, P. Flicek, Ensembl 2020. *Nucleic Acids Research*, gkz966 (2019).

91. D. Karolchik, A. S. Hinrichs, T. S. Furey, K. M. Roskin, C. W. Sugnet, D. Haussler, W. J. Kent, The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–496 (2004).

92. W. J. Kent, BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).

93. Drosophila 12 Genomes Consortium, Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**, 203–218 (2007).

94. T. E. Oliphant, *Guide to NumPy* (Continuum Press, Austin, Tex., 2015).

95. W. McKinney, "Data Structures for Statistical Computing in Python" (Austin, Texas, 2010; https://conference.scipy.org/proceedings/scipy2010/mckinney.html), pp. 56–61.

96. J. VanderPlas, B. Granger, J. Heer, D. Moritz, K. Wongsuphasawat, A. Satyanarayan, E. Lees, I. Timofeev, B. Welsh, S. Sievert, Altair: Interactive Statistical Visualizations for Python. *JOSS* **3**, 1057 (2018).

97. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

98. J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, J. P. Mesirov, Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).

99. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics* **14**, 178–192 (2013).

Figure 1. Chen et al.

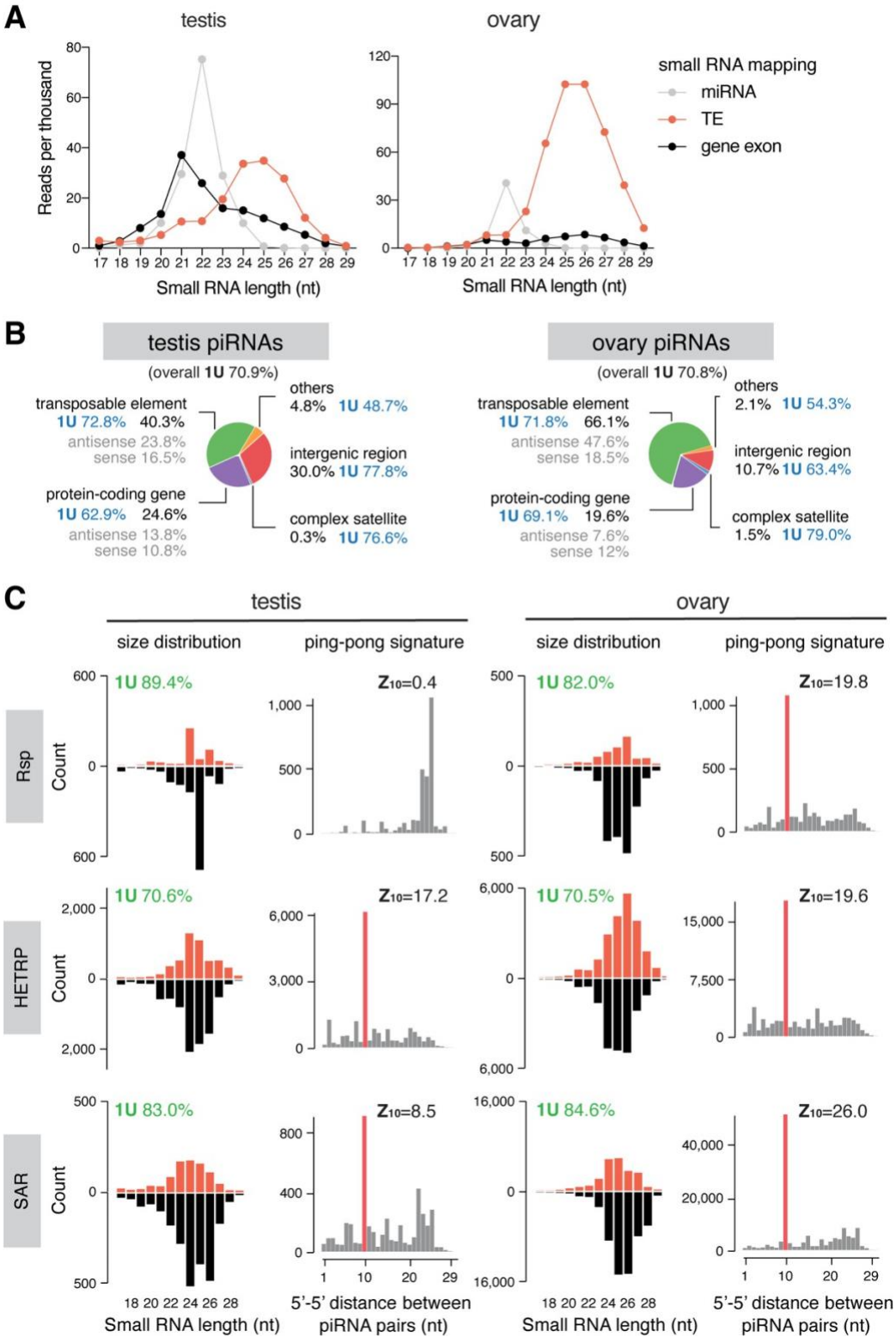


Figure 1. Analysis of small RNA profiles in testis and ovary.

(A) Size distribution plots of microRNAs (gray), remaining small RNAs that map to TE consensus (red) and protein-coding gene exons (black), in testis (left) and ovary (right).

(B) Annotation of piRNA reads in testis (left) and ovary (right). 1U nucleotide bias (%) for overall piRNA population and each category is shown next to labels. See also Supplementary Fig. S1.

(C) Characterization of piRNAs mapping to three known complex satellites in two sexes. Left panels of each sex are size distribution of piRNAs mapping to consensus sequences of each complex satellite. Right panels are distributions of 5'-to-5' distances of piRNA pairs, showing an enrichment for 10nt (i.e., ping-pong signature), except for *Rsp* in testis ($P<0.05$ for $z>1.96$). 1U nucleotide bias (%) and ping-pong z-score are shown above plots. See also Supplementary Fig. S2.

Figure 2. Chen et al.

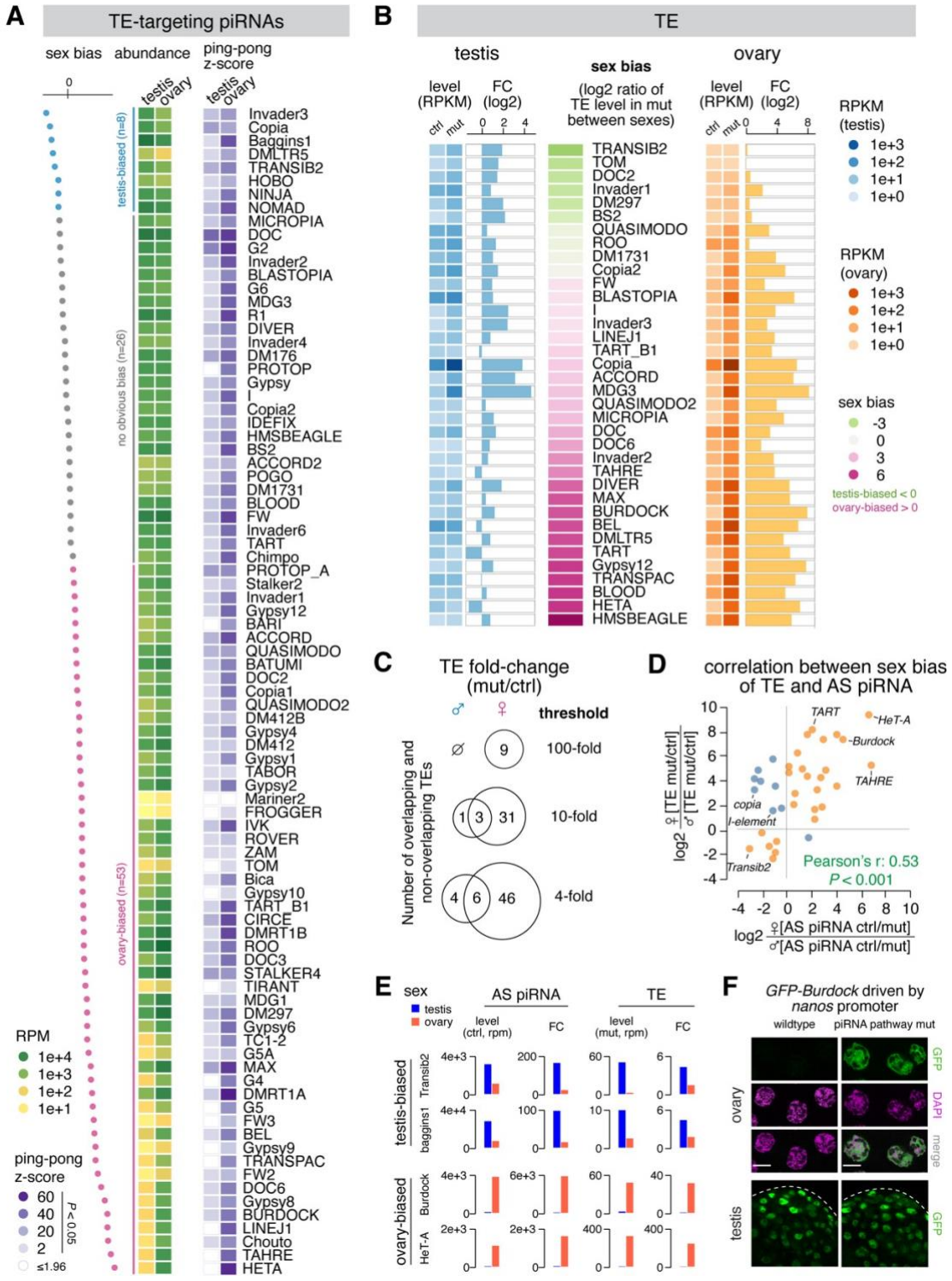


Figure 2. Sexually dimorphic piRNA programs parallel sex-specific TE expression.

(A) Heatmaps showing the abundance of antisense piRNA (left) and ping-pong z-score (right) for each TE family in two sexes. Statistically significant ping-pong z-scores ($z>1.96$, equivalent to $P<0.05$) are color coded, while the remaining are marked as blank. TE families are sorted by sex bias of piRNA expression, defined as the log2 ratio of antisense piRNA abundance in testis over ovary. TEs with more than 2-fold differences in antisense piRNAs are colored as testis-biased (blue) and ovary-biased (pink), respectively, with the remaining having no obvious bias (gray).

(B) Expression of 36 TE families that are regulated by *rhi* (see methods) in testis (left) and ovary (right). TE families are sorted by sex bias of their expression in piRNA pathway mutant ($rhi^{-/-}$), defined as the log2 ratio between sexes. Heatmaps display TE levels in control and mutant, while bar graphs show the fold change of expression in mutant over control.

(C) Venn diagrams of the number of TEs showing 100-, 10- and 4-fold de-repression in *rhi* mutant over control of two sexes.

(D) Scatter plot displaying the correlation between sex biases of TE and TE-antisense piRNA. For each TE family, the loss of antisense piRNAs in *rhi* mutants was calculated in each sex (ctrl over mut). The sex bias of piRNAs was defined as the log2 ratio of piRNA loss in female over male. Similarly, TE de-repression in *rhi* mutants was calculated in each sex (mut over ctrl), and the sex bias was defined as the log2 ratio of TE de-repression in female

over male. TE families that show a correlation between the sex bias of antisense piRNA and that of TE de-repression are colored as orange, with the rest as blue.

(E) Histograms showing profiles of two sex-biased TEs for each sex. Antisense piRNA levels refer to those in control gonads, TE levels refer to those in piRNA pathway mutants (*rhi*$^{-/-}$), and the fold change is calculated as mutant over control for TEs and the reverse for antisense piRNAs.

(F) Confocal images of the apical tip of stage 7–8 nurse cells in ovary (top) and testis (bottom) that express a *Burdock*-fused GFP reporter in wild-type and piRNA pathway mutant (*rhi*$^{-/-}$) background, respectively. The reporter is expressed by *nanos* promoter that drives germline expression in both sexes, thus enabling the examination of piRNA silencing of *Burdock* sequences independent of natural expression patterns of *Burdock* transposon. Scale bars: 20μm.

Figure 3. Chen et al.

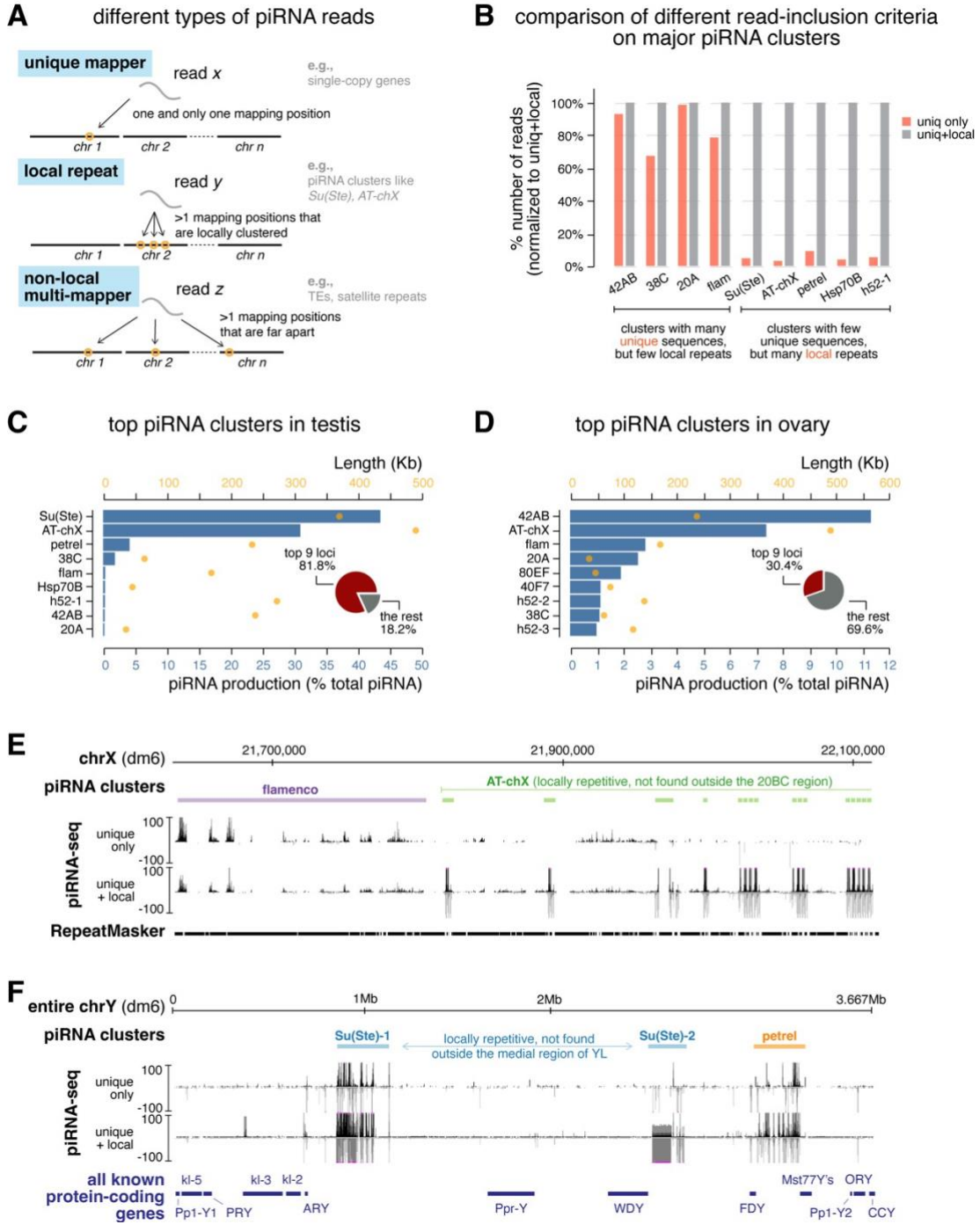


Figure 3. Definition of piRNA clusters in testis and ovary using a new algorithm.

(A) Three types of piRNA reads, defined based on their mapping positions. Uniquely mapped reads can be mapped to only one position in the genome and their origin is unambiguous. Reads derived from local repeats can be mapped to several positions in the genome; however, all of these mapping positions are locally clustered in a single genomic region. On the other hand, non-local multi-mappers can be mapped to multiple positions that are not restricted to one genomic region (typically mapped to more than one chromosome). Previously, only uniquely mapped reads were used to define piRNA clusters and quantify their expression, as the genomic origin of multi-mappers is ambiguous. Inclusion of multi-mappers derived from local repeats, as shown in this study, allows identification of new piRNA clusters as well as a more accurate quantification of piRNA production from known clusters. At the same time, it preserves the certainty that reads are generated from genomic loci in question. See Supplementary Fig. S4 for detailed pipeline.

(B) Histogram comparing numbers of mapped reads for major piRNA clusters using different read-inclusion criteria as defined in (A). For each cluster, the number of mapped reads generated by different methods is normalized to the method that includes both unique and local repeat reads (the right column). See also Supplementary Fig. S4 and methods.

(C) Expression of the top 9 most active piRNA clusters in testis. Blue bars depict the contribution of each cluster to total piRNAs (%) and orange dots show cluster lengths according to dm6 genome assembly. Insert is a pie chart of the contribution of top 9 loci to total piRNAs in testis.

(D) Same as in (C) but for ovary.

(E) UCSC genome browser view of a peri-centromeric region (chrX) encompassing the entire *flamenco* locus (purple) and the distal part of *AT-chX* piRNA cluster (green). Below genomic coordinates (dm6) are piRNA coverage tracks using different read-inclusion criteria. Note that, whereas *flamenco* produce piRNAs that can be mostly mapped to unique genomic positions, *AT-chX* generates piRNAs that map to local repeats in this cluster, but nowhere else in the genome.

(F) UCSC genome browser view of the entire Y chromosome that harbors two *Su(Ste)* loci (blue) and the novel *petrel* piRNA cluster (orange). piRNA coverage tracks using different read-inclusion criteria are shown below genomic coordinates (dm6). At the bottom, all known Y-linked protein-coding genes are drawn for reference (not to exact scale). Note that piRNA profiles of *Su(Ste)* and *petrel* clusters collapse if piRNAs derived from local repeats are excluded.

Figure 4. Chen et al.

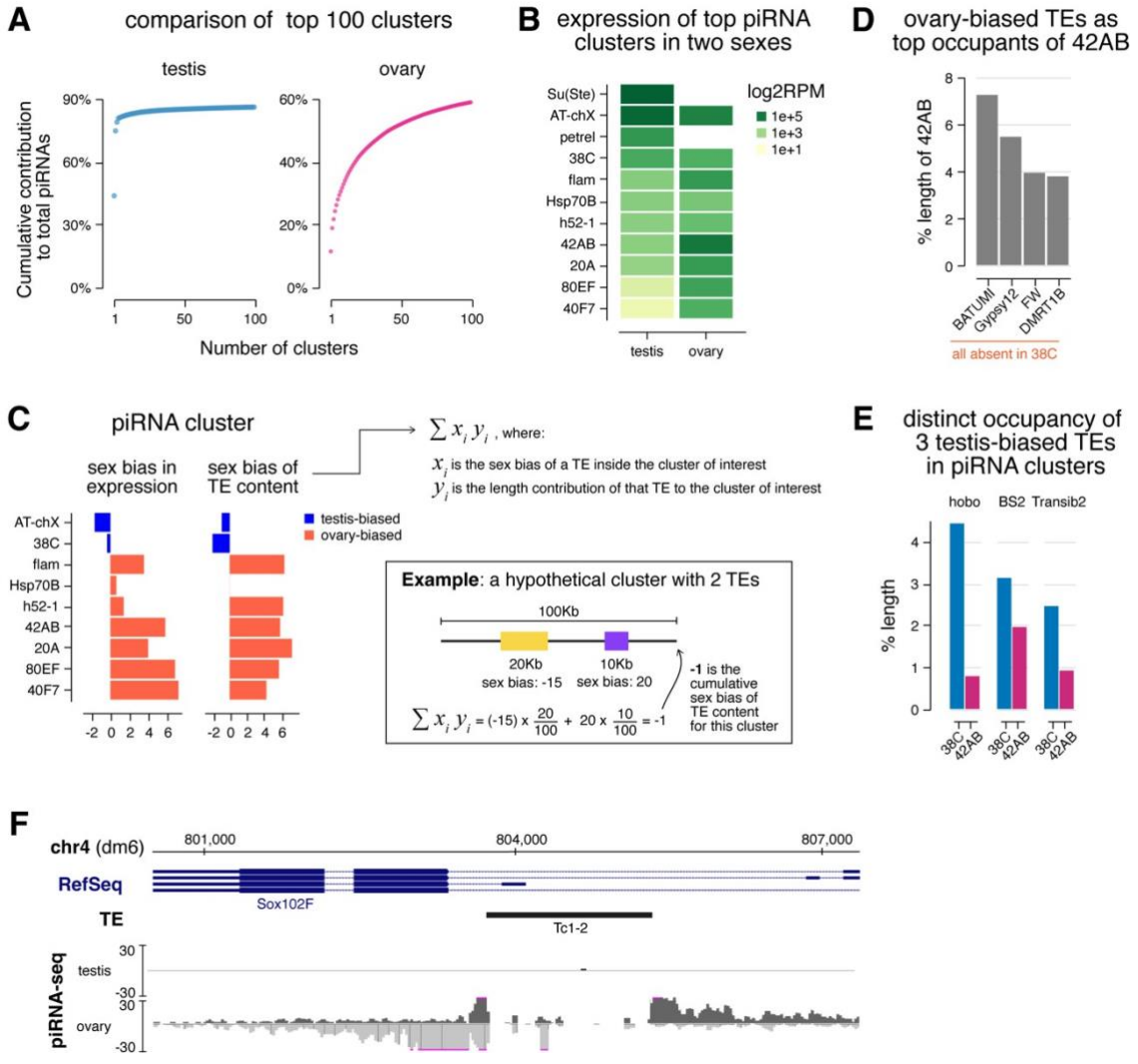


Figure 4. piRNA clusters are differentially employed to tame sex-specific TE expression.

(A) Plot showing the cumulative contribution of top piRNA clusters to the total piRNA populations in testis (left) and ovary (right), up to 100 clusters.

(B) Heatmaps showing piRNA production from major piRNA clusters. Note that *Su(Ste)* and *petrel* clusters are Y-linked so there are no piRNAs from these loci in females that lack Y chromosome.

(C) Bar graphs displaying the sex bias of piRNA cluster expression (left) and cumulative sex bias of the TE context for each cluster (right). Sex bias of piRNA cluster expression is defined as the log2 ratio of piRNA cluster expression in ovary over testis shown in (B), so ovary-biased ones are positive in value. Cumulative sex bias of cluster TE content is calculated by summing the sex bias of TEs (as described for Fig. 2B) weighted by their length contributions to the cluster (equation shown on the right). An example is shown on the bottom right for a hypothetical cluster composed of two TEs with lengths and sex biases labeled accordingly for illustration. Only TEs showing strong sex biases were used in calculation. See also methods.

(D) TE composition of ovary-biased *42AB* cluster. Shown are fractions of *42AB* cluster occupied by sequences from top 4 TE families. These 4 TEs are completely absent in *38C*, a testis-biased piRNA cluster. Expression of these 4 TEs is all ovary-biased (Supplementary Fig. S3A).

(E) Contributions of three testis-biased TEs (Supplementary Fig. S3A) to the ovary-biased *42AB* cluster and testis-biased *38C* cluster. These TEs were selected as the most enriched by length in *38C* compared to *42AB*.

(F) The *Sox102F* gene generates piRNAs in ovary, but not in testis. This locus harbors a single autonomous TE, *Tc1-2*, that has ovary-biased expression (Supplementary Fig. S3A). piRNA coverage tracks show both uniquely-mapped and local repeat-derived reads.

Figure 5. Chen et al.

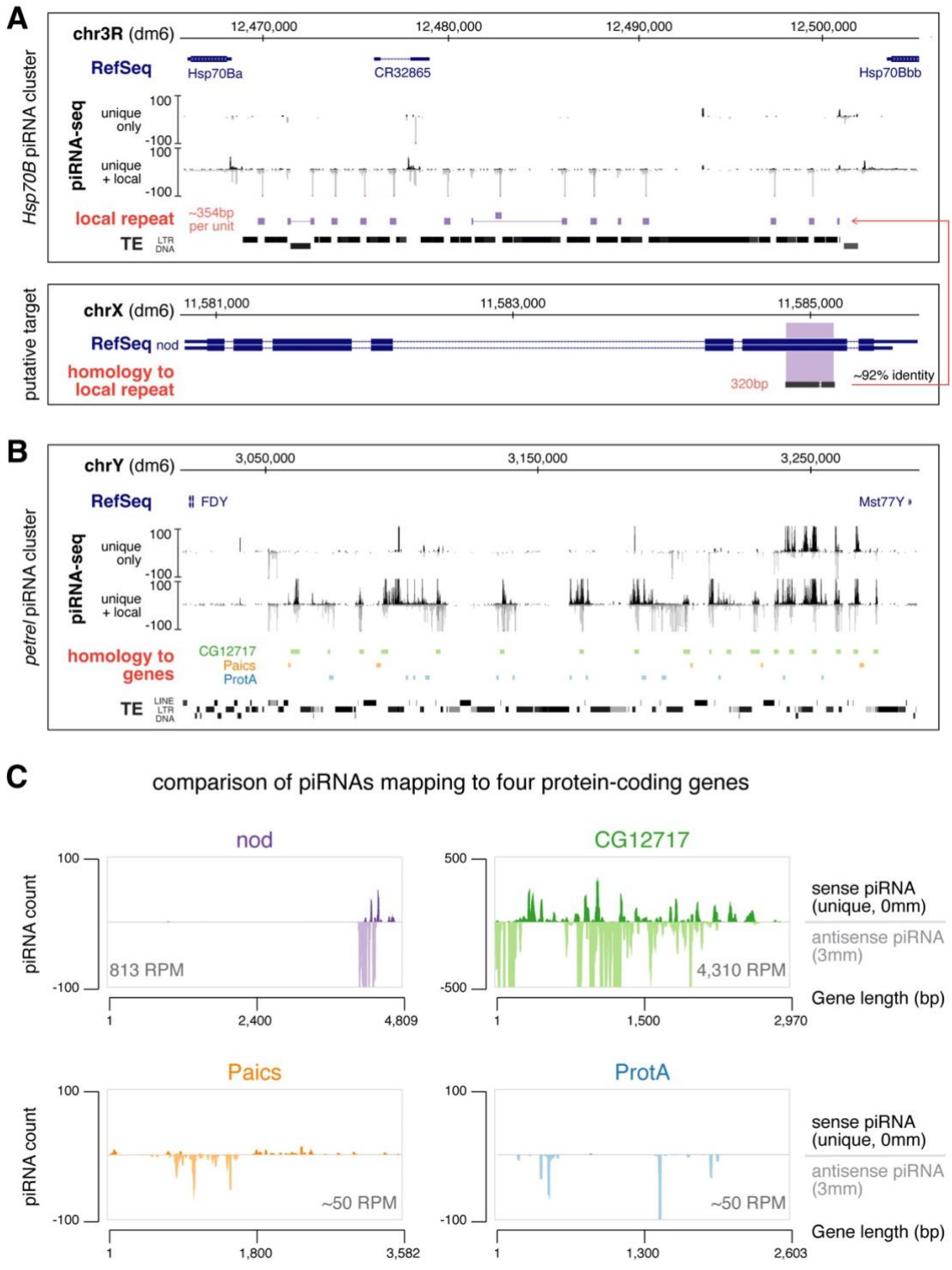


Figure 5. *Hsp70B* and *petrel* piRNA clusters encode piRNAs that target host genes.

(A) *Hsp70B* piRNA cluster (top) and the putative target, *nod* (bottom). piRNA coverage tracks using different read-inclusion criteria are shown below RefSeq and genomic coordinates (dm6) for *Hsp70B* cluster. ~354bp local repeats homologous to a 320bp exonic region of *nod* are depicted as solid blocks, which fill up most inter-TE space at this locus. Note that "unique+local" piRNA track does not include TE-derived piRNAs that map outside this locus, but it picks up *bona fide* local repeats that are homologous, but not identical, to *nod*.

(B) *petrel* piRNA cluster on Y chromosome. piRNA coverage tracks using different read-inclusion criteria are shown. Sequences with high levels of sequence similarity to protein-coding genes are depicted as colored blocks (not to exact scale): *CG12717* (green), *Paics* (orange), *ProtA* (blue). Note that gene-homologous islands fill up most inter-TE space at this locus. Genomic coordinates are based on dm6 genome assembly.

(C) Coverage of sense (genome-unique, 0 mismatch) and antisense piRNAs (with up to three mismatches) over four putative, protein-coding gene targets of testis piRNAs. Antisense piRNA abundance is shown for each gene.

Figure 6. Chen et al.

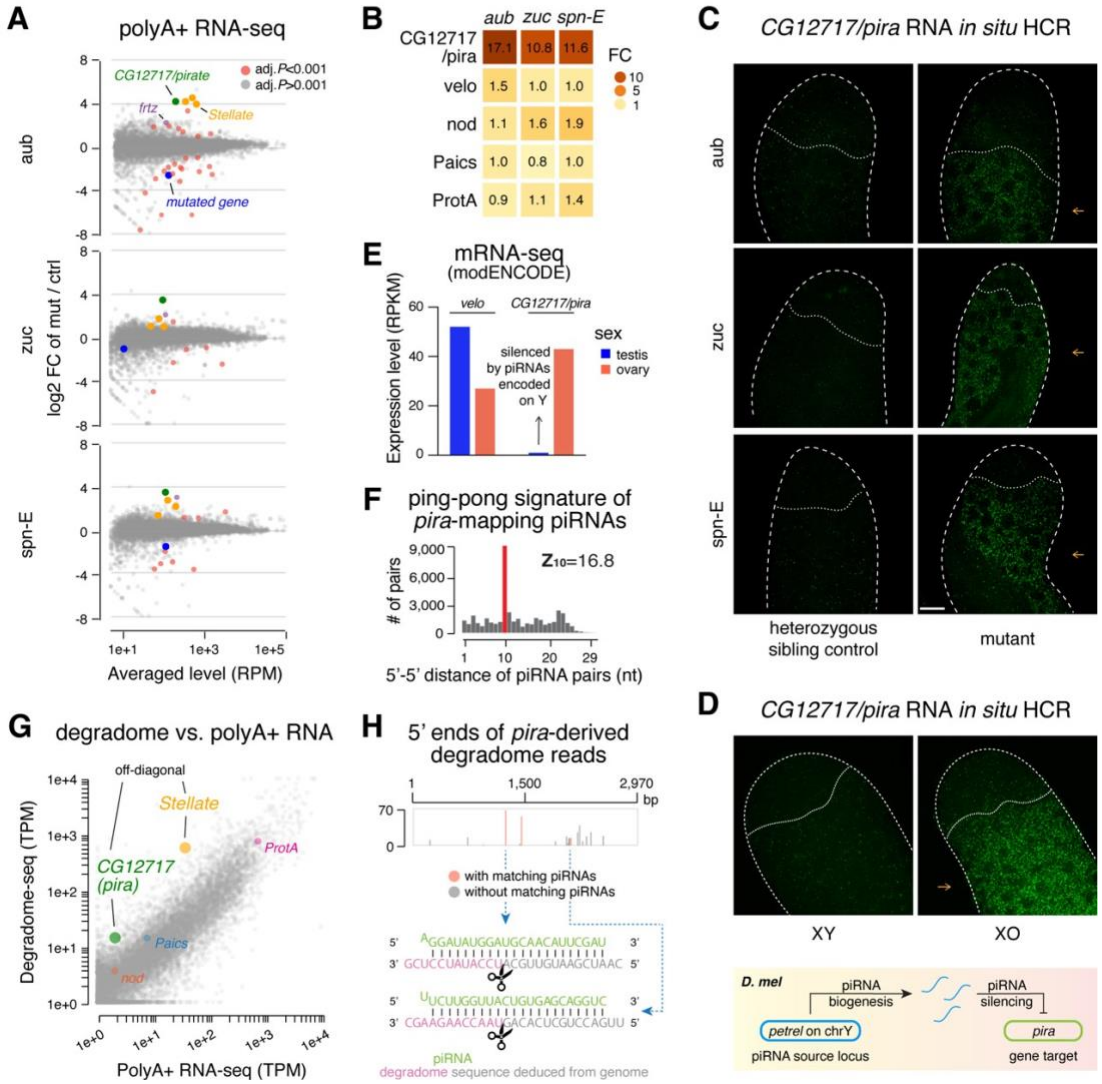


Figure 6. Regulation of *CG12717/pira* by the piRNA pathway.

(A) MA plots showing gene expression changes from polyA+ RNA-seq of *aub* (top), *zuc* (middle) and *spn-E* (bottom) mutant testes versus heterozygous sibling controls. Genes are marked red when passing a stringent statistical cutoff (adjusted $P<0.001$, from DESeq2).

Additional coloring includes: *CG12717/pira* (green), annotated *Stellate* transcripts (orange), *frtz* (purple), and the mutated gene in each mutant (blue).

(B) Heatmaps showing fold change of five protein-coding genes in three mutant testes according to polyA+ RNA-seq shown in (A).

(C) Confocal images of *pira* mRNAs detected by *in situ* HCR in *aub* (top), *zuc* (middle) and *spn-E* (bottom) mutant testes along with respective heterozygous sibling controls. Probes were designed against a ~400bp sequence unique to *pira* and absent on Y (Supplementary Fig. S5B), so they do not target *petrel* piRNA precursors. Note that de-repression of *pira* in piRNA pathway mutants is observed specifically in differentiating spermatocytes (pointed to by orange arrows). Scale bar: 20µm.

(D) Confocal images of *pira* transcripts detected by *in situ* HCR in XY and XO testes. Same scale as in (C). A schematic of Y chromosome- and piRNA-dependent silencing of *pira* is shown at the bottom.

(E) Bar graphs displaying modENCODE data of *pira* and its paralog *velo* expression in *D. melanogaster* gonads of both sexes.

(F) Analysis of ping-pong processing of *pira*-mapping piRNAs. Histogram shows distribution of 5'-to-5' distances of complementary piRNA pairs with an enrichment for 10nt (i.e., ping-pong signature). To select secondary piRNAs processed from *pira* transcripts, only reads that map perfectly to *pira* mRNAs in sense orientation and do not map perfectly to *petrel* cluster were used in this analysis. Antisense piRNAs were selected allowing up to three mismatches.

(G) Analysis of cellular transcripts enriched in degradome-seq library. Scatter plot shows the number of degradome-seq reads for each gene relative to its expression measured by polyA+ RNA-seq. Transcripts enriched in degradome-seq library relative to their expression are located above diagonal. These include *Stellate*, a known target of piRNA repression, and *CG12717/pira*, while *nod*, *Paics* and *ProtA* transcripts are not enriched in degradome-seq. Different annotated copies of *Stellate* genes were merged.

(H) Analysis of *pira*-derived degradome reads. The abundance of 5' ends of *pira*-derived degradome reads is plotted, with the ones that have 10nt 5'-5' overlap with antisense piRNAs marked red. Examples of two such degradome and piRNA pairings are shown at the bottom.

Figure 7. Chen et al.

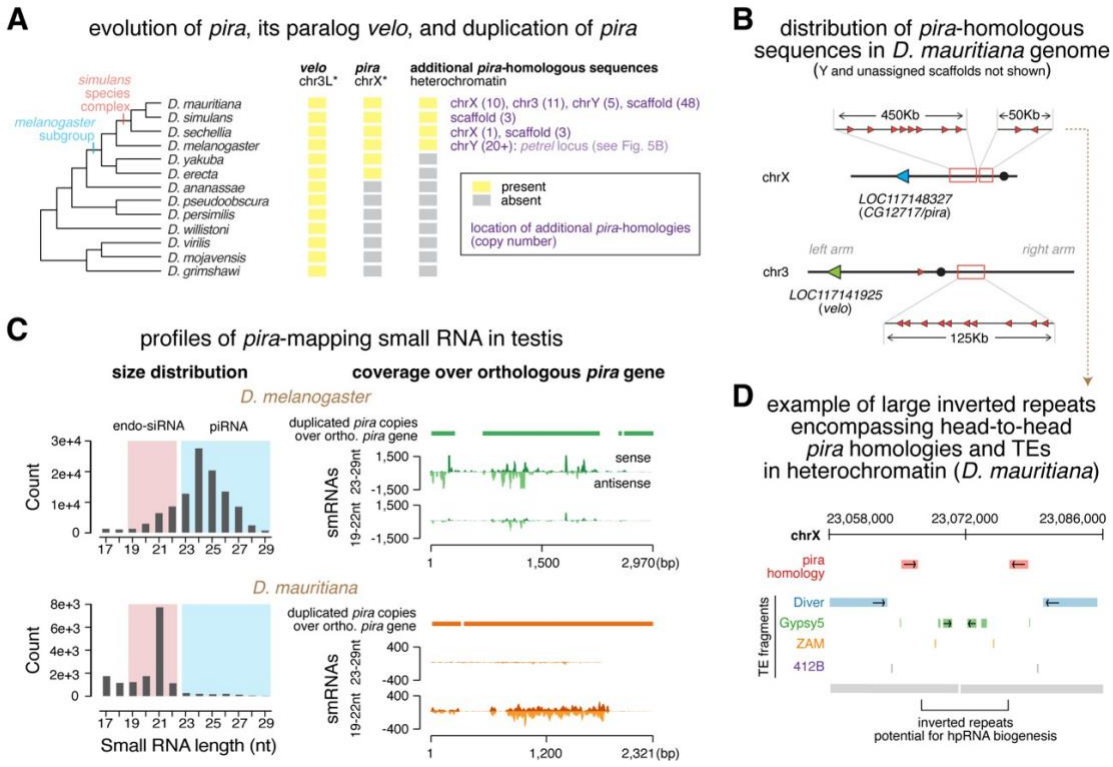


Figure 7. Evolution of *pira* and *pira*-targeting small RNAs.

(A) Cladogram of major species in *Drosophila* genus (left) and the evolutionary history of *velo*, *pira* and *pira*-related sequences in genomes of these species (right). Orthologs were identified based on sequence homology and synteny. Shown in purple are locations of additional *pira* copies in each species and copy numbers in parenthesis. Asterisk marks the chromosome name in the *melanogaster* subgroup, as karyotype differs in more distantly related species.

(B) Cartoon depicting distribution of *pira*-homologous sequences in *D. mauritiana* genome. Orthologous *pira* is marked blue, orthologous *velo* is marked green, and the duplicated,

candidate sources of *pira*-targeting endo-siRNAs are marked red. Note that they scatter across peri-centromeric heterochromatin of chrX and chr3, as well as chrY and scaffolds (not shown).

(C) Profiles of *pira*-mapping small RNAs in testes of *D. melanogaster* (top) and *D. mauritiana* (bottom). Size distributions are shown on the left. Coverage plots over the orthologous *pira* gene in each species are shown on the right, including: cumulative alignment of heterochromatic, duplicated copies of *pira* over the syntenic, orthologous *pira* (top, solid bar), stranded coverage of 23–29nt piRNAs (middle, histogram) and 19–22nt endo-siRNAs (bottom, histogram) over the orthologous *pira* gene.

(D) Illustration showing two representative head-to-head copies of *pira* homology (red) in the peri-centromeric heterochromatin of *D. mauritiana* X chromosome. *pira*-related sequences are flanked by TEs and are part of a large inverted repeat that could potentially permit hpRNA biogenesis.

MATERIALS AND METHODS

Fly stocks

Stocks and crosses were raised at 25 °C. The following stocks were used: *aub*^{QC42} (BDSC4968), *aub*^{HN2} (BDSC8517), *zuc*^{Df} (BSDC3079), *spn-E*^{hls3987} (BDSC24853) and *spn-E*^{1} (BDSC3327) were obtained from Bloomington Drosophila Stock Center; *rhi*^{2} and *rhi*^{KG} were gifts of William Theurkauf; *zuc*^{HM27} was a gift from Trudi Schüpbach; *nosP-GFP-Burdock* was a gift from Julius Brennecke. Heterozygous siblings were used as controls for all experiments, unless noted otherwise. XO male was generated by crossing XY males to C(1)RM females (BDSC9460).

RNA *in situ* hybridization chain reaction (HCR)

A kit containing a DNA probe set, a DNA probe amplifier and hybridization, amplification and wash buffers were purchased from Molecular Instruments (molecularinstruments.org) for *CG12717* transcripts. To avoid targeting the *petrel* locus on Y, we designed probes against a ~400bp unique region present in *CG12717* on X but absent on Y chromosome. The *CG12717* probe set (unique identifier: 3916/E064) initiated B3 (Alexa546) amplifier. *In situ* HCR v3.0 (*84*) was performed according to manufacturer's recommendations for generic samples in solution.

Image acquisition and analysis

Confocal images were acquired with Zeiss LSM 800 using a 63x oil immersion objective (NA=1.4) and processed using Fiji (*85*). Single focal planes were shown in all images, where dotted outlines were drawn for illustration purposes.

RNA-seq

RNA was extracted from 160–200 pairs of dissected testes of $aub^{QC42/HN2}$, $spn\text{-}E^{1/hls3987}$, $zuc^{HM27/Df}$ and respective heterozygous sibling controls in TRIzol (Invitrogen). PolyA+ selection was done using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB E7490), followed by strand-specific library prep with NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB E7760) according to manufacturer's instructions. Libraries were sequenced on Illumina HiSeq 2500 yielding 11–17 million 50bp single-end reads. PolyA-selected RNA-seq of *rhi* mutants and controls were downloaded from NCBI SRA (see Chen et al. 2020 for testis and GSE126578 for ovary, two biological replicates per sex per genotype).

RNA-seq analysis

To quantify expression levels of protein-coding genes across different piRNA pathway mutants (*aub*, *zuc* and *spn-E*), we used kallisto 0.46.1 (*86*). Three heterozygous controls were pooled as triplicates of controls to be analyzed against duplicates of each of the three piRNA pathway mutants. Transcript-level quantification was pooled to obtain gene-level quantification. Differential gene expression was done with DESeq2 (*87*). Expression of

CG12717 and *velo* in ovary and testis from modENCODE (*41*) was extracted from FlyBase (*88*).

For analysis of TE expression and TE fold change in piRNA pathway mutants of both sexes, *rhi* mutants were used where piRNA production from germline-specific dual-strand clusters was abolished. Reads mapped to rRNA were discarded using bowtie 1.2.2 allowing 3 mismatches. Reads were then mapped to TE consensus from RepBase17.08 using bowtie 1.2.2 with -v 3 -k 1 and normalized to the total number of reads mapped to dm6 genome. For simplicity, reads mapped to LTR and internal sequences were merged for each LTR TE given their well-correlative behaviors. Only TEs that have ≥5 RPM and ≥2.5 RPKM expression in piRNA pathway mutants of either sex were kept for the analysis (n=87). A pseudo-count of 1 was added before calculating TE fold change in piRNA pathway mutants.

Degradome-seq and analysis

Degradome-seq was done as previously described (*89*). Briefly, RNA was extracted from 50 pairs of wild-type testes (w^{1118}) in TRIzol, followed by DNase treatment (TURBO DNase) and rRNA depletion (ribo-zero). Next, RNA bearing 5’ monophosphate was enriched by ligating 5’ adaptor (5’-GUUCAGAGUUCUACAGUCCGACGAUC) with T4 RNA ligase, followed by size selection for RNA >200nt (RNA Clean & Concentrator-5). Reverse transcription was performed with SuperScript III and a primer containing a degenerate sequence at its 3’ end (5’-GCACCCGAGAATTCCANNNNNNNN), which also introduced the 3’ adaptor. PCR was done to amplify cDNA and to introduce sequencing primer and index sequences. Two replicates were sequenced on Illumina HiSeq 2500 for 150bp single-

end reads. To compare degradome with polyA+ RNA-seq, we used kallisto to assign reads to transcripts as described above. To identify degradome sequences that are 3' piRNA-guided cleavage products of *CG12717* transcripts, we mapped degradome reads (trimmed to 50bp) to dm6 genome with bowtie 1.2.2 -v 0 -m 1 and extracted those mapping to the coding strand of *CG12717* (note that it is an intron-less gene). *CG12717*-derived degradome reads were extended upstream of their 5' ends based on dm6 genome sequence to examine their overlap with antisense piRNAs and extent of complementarity.

Identification of TEs regulated by *rhi*

To identify a set of TEs regulated by *rhi* in at least one sex, we looked for TEs that have at least 100 RPM in *rhi* mutant ovaries or at least 25 RPM in *rhi* mutant testes. Next, we filter out TEs that show fewer than threefold de-repression in both sexes. From the initial 87 TEs defined above, these led to a total of 36 TEs regulated by *rhi* in at least one sex shown in Fig. 2B and Fig. 2D. See Supplementary Fig. S3C for detailed profiles of these 36 TEs.

piRNA-seq

RNA extraction was done as above for RNA-seq. 18–30nt small RNAs were purified by PAGE (15% polyacrylamide gel) from ~1µg total RNA. Purified small RNA was subject to library prep using NEBNext Multiplex Small RNA Sample Prep Set for Illumina (NEB E7330) according to manufacturer's instructions. Adaptor-ligated, reverse-transcribed, PCR-amplified samples were purified again by PAGE (6% polyacrylamide gel). Two biological

replicates per genotype were sequenced on Illumina HiSeq 2500 yielding 15–20 million 50bp single-end reads.

piRNA-seq analysis of TEs, complex satellites and genes

To isolate piRNAs, adaptor-trimmed total small RNAs were size-selected for 23–29nt (cutadapt 2.5) and those mapped to rRNA, miRNA, snRNA, snoRNA and tRNA were discarded (bowtie 1.2.2 with -v 3). piRNAs were first mapped to RepBase17.08 to obtain the portion mapping to TEs and complex satellites; the rest was then mapped to gene sequences derived from the gtf file downloaded from Ensembl (BDGP6.28.99) (*90*); reads unmapped to repeats and genes were then mapped to dm6 to infer the portion mapping to inter-genic regions, and the unmapped ones were listed under "others" category. A pipeline is also drawn in Supplementary Fig. S1. For TE-antisense piRNA analysis, piRNA reads were mapped, normalized and processed as done for polyA+ RNA-seq (see above). For complex satellite-mapping small RNAs, we plotted size distribution, analyzed nucleotide bias at position 1 and calculated coverage along consensus sequences using bedtools v2.28.0. Analysis of ping-pong signature (i.e., 5'-to-5' distances between complementary piRNA pairs) and phasing signature (i.e., 3'-to-5' distances on the same strand) were done with custom scripts. Ping-pong z-score was calculated using 1–9nt and 11–23nt as background distribution for an enrichment of 10nt, whereas phasing z-score, Z_1 as defined in (*18*), was calculated using 2–50nt as background distribution for an enrichment of 1nt. For piRNAs antisense to protein-coding genes of interest, we downloaded gene sequences from FlyBase (*88*) and mapped piRNAs to them using bowtie 1.2.2. For mRNA-derived sense piRNAs, we mapped piRNAs

to genome and kept ones with unique mapping and zero mismatch (bowtie 1.2.2 with -v 0 -m 1) to the gene regions and orientations of interest.

A pipeline tolerating local repeats for piRNA cluster analysis

We first separated rRNA-depleted 23–29nt small RNA reads that map to one unique location in the genome and others that have multiple mapping positions ("multi-mappers"). For all multi-mappers, we filtered out those who map to more than one chromosome arm, retaining only ones with multi-mapping positions on a single chromosome arm ("intra-chromosomal repeats"). Then, for each of the reads we kept as intra-chromosomal repeats, we calculated the maximum distance ("max distance") of all mapping positions. In order to enforce the local requirement, we hoped to identify a cutoff distance for max distances, which is large enough to contain known piRNA loci but small enough to allow certain resolution of neighboring loci. To this end, we analyzed a pool of 50bp DNA fragments tiling the entire dm6 genome and plotted a histogram of max distances for all intra-chromosomal repeats (Supplementary Fig. S4B). This revealed a density of intra-chromosomal repeats having max distances smaller than ~500Kb, as well as four pronounced peaks with larger max distances. Sequence analysis uncovered the identities of these peaks: the peak with ~600Kb max distance corresponds to *AT-chX*, the peak with ~1.8Mb max distance represents *Su(Ste)*, and the other two peaks mostly contain Y-specific simple repeats. We thus set a 2Mb tolerance threshold of max distances to allow local repeats in piRNA cluster analysis. In other words, we defined local repeats as repeats that have all copies contained within a window smaller than 2Mb and merged their normalized counts with unique sequences for piRNA cluster

analysis. Alignment was done using bowtie2 to dm6 genome. To compare this new pipeline with the convention that uses only unique mappers, we calculated the number of reads mapped to major piRNA clusters using both methods (Fig. 3B). A summary of this pipeline is shown on Supplementary Fig. S4A.

Definition of piRNA clusters

23–29nt small RNAs were mapped to dm6 genome using the above-mentioned pipeline tolerating local repeats and generated coverage profiles across 1Kb windows that tile the genome. 1Kb windows including highly expressed miRNA, snRNA, snoRNA, hpRNA or 7SL SRP RNA were excluded. 1Kb windows with low read-coverage (≤100bp) were also excluded. Then, 1Kb windows that produce at least certain amounts of piRNAs were extracted for cluster definition (≥10RPM for testis, ≥50RPM for ovary). Neighboring 1Kb widows within 3Kb were merged. If merged windows were ≥5Kb, they were merged again within 15Kb. This yields 844 piRNA clusters in testis and 525 piRNA clusters in ovary, after manual curation. Major piRNA clusters described before in ovaries (*3, 16*) were all recovered with similar resolution. To compare expression levels of major piRNA clusters between sexes, cluster boundaries were manually curated to guarantee identical regions being compared. piRNA clusters defined in this study for both sexes are listed in Supplementary Table S1.

TE content of piRNA clusters

TE annotation in dm6 genome was downloaded from UCSC Table Browser (*91*). piRNA cluster boundaries were defined as described above. For piRNA cluster of interest, the TE content is calculated as length contribution to the entire cluster length by individual TEs. TE contents add up to less than 100%, as TEs do not fill completely the cluster length.

Sex bias of piRNA cluster TE content

Sex bias of individual TEs was first computed as log2 ratio of expression levels in piRNA pathway mutants (*rhi*) between sex (ovary over testis). Sex bias of piRNA cluster TE content was then computed as the cumulative sex bias of individual TEs inside the cluster, weighted by their length contribution to the cluster. Using all expressed TEs or only ones that show pronounced sex bias generated comparable results. To eliminate noise, we only used TEs that exhibit strong, ≥10-fold sexual difference in expression (n=24). An equation and an example are shown in Fig. 4C.

BLAT and BLAST analysis

To characterize the unannotated sequence between annotated repeats in piRNA clusters, inter-repeat sequences were analyzed using BLAT on UCSC Genome browser (*92*). For example, an inter-TE sequence at *Hsp70B* locus was used to BLAT against dm6 genome, which revealed the homology with an exon of *nod* gene (Fig. 5A). Homology between *CG12717* and *velo* was done with both BLAT and BLAST, which yielded similar results. Characterization of *CG12717*-homologous sequences at *petrel* locus (Supplementary Fig.

S5B) was done by multiple sequence alignment with the Needle program (ebi.ac.uk/Tools/psa/emboss_needle/).

Phylogenetic analysis

The longest transcripts of *velo* and *CG12717* in *D. melanogaster* genome were used to BLAST against nucleotide collection with tblastn program. Orthologs of these two genes in other *Drosophila* species were identified based on high nucleotide similarity and synteny. In all orthologs identified for both genes, we found the same flanking protein-coding genes, confirming their ortholog identities. Occasionally, BLAST with *CG12717* revealed the *velo* ortholog in that species as well; but only in *D. mauritiana*, *D. simulans* and *D. sechellia* genomes are there additional hits with high sequence homology to *CG12717*, other than the orthologous *CG12717* and *velo*. These additional *CG12717*-related sequences are in some cases annotated as predicted genes, but all buried in TE-rich heterochromatin (close to centromere or in highly repetitive unassigned scaffolds). To examine the organization of *CG12717*-related sequences in *D. mauritiana* genome in detail, we ran BLAST using *D. mauritiana CG12717* gene against its genome (assembly: GCA_004382145.1), which revealed additional unannotated regions with high sequence similarity to *CG12717*. Those located on chrX and chr3 were drawn in Fig. 7B. The instance where two adjacent *CG12717*-related sequences are arranged head-to-head on chrX is illustrated in Fig. 7D, and the other three such instances are found in unassigned scaffolds. To uncover the identity of flanking unannotated sequences, we BLAST the 50Kb region encompassing *CG12717*-related

sequences against TE consensus (RepBase17.08). The cladogram was drawn for illustration (*93*).

Analysis of testis small RNAs in non-*D. melanogaster* species

Testis small RNA libraries from non-*D. melanogaster* species was downloaded from NCBI SRA: *D. simulans* SRR7410589 (*38*) and *D. mauritiana* SRR7961897 (*21*). Adaptor-trimmed reads were mapped to the orthologous *CG12717* gene, *D. simulans GD15918* and *D. mauritiana LOC117148327*, respectively (bowtie 1.2.2 with -v 3 -k 1). Coverage was plotted along the orthologous *CG12717* gene.

Data visualization and statistical analysis

Most data visualization and statistical analysis were done in Python 3 via JupyterLab with the following software packages: numpy (*94*), pandas (*95*) and altair (*96*). The UCSC Genome Brower (*97*) and IGV (*98, 99*) were used to explore sequencing data and to prepare browser track panels shown.

Data availability

Sequencing data can be accessed via NCBI SRA with accession numbers listed below.

PRJNA646006 (*rhi*), PRJNA646216 (*aub, zuc and spn-E*), PRJNA719671 (degradome).

Figure S1. Chen et al.

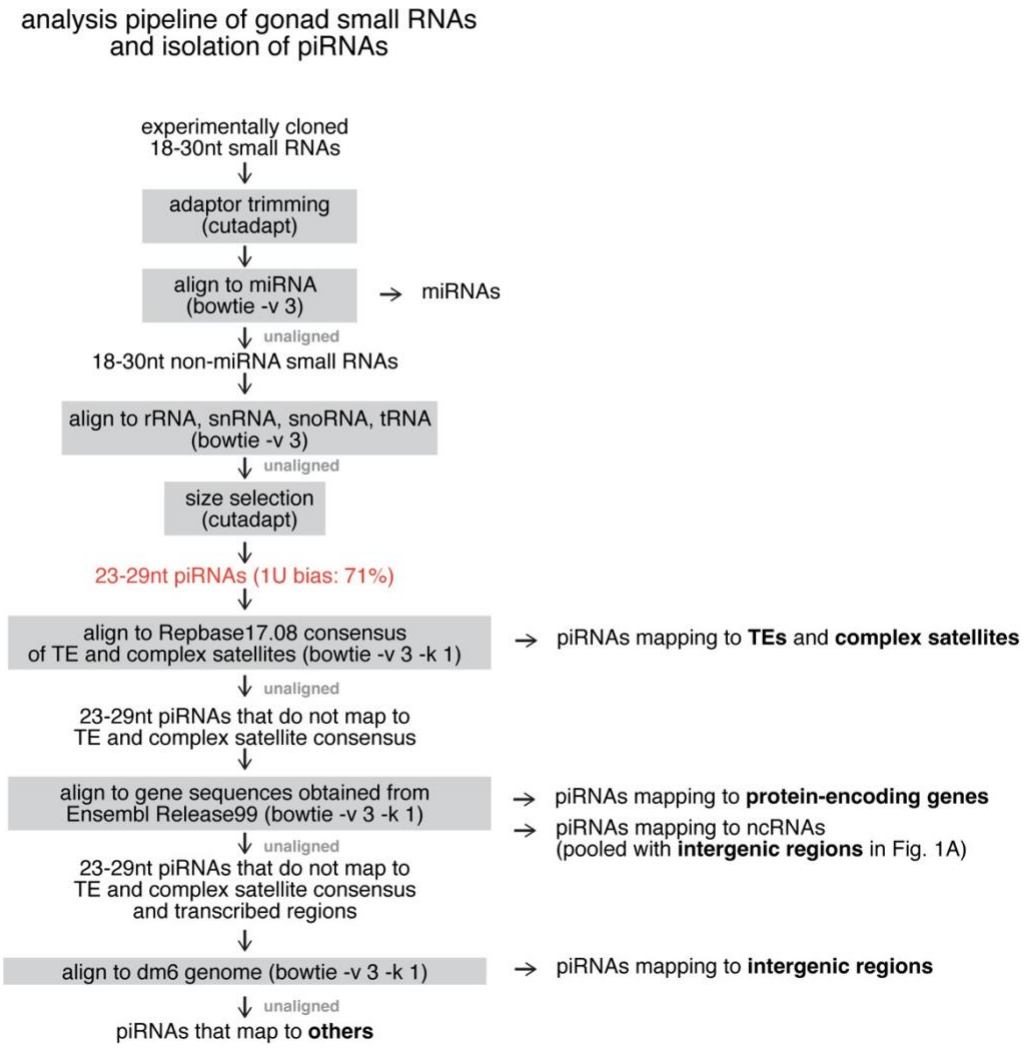


Figure S1. Analysis pipeline of gonad small RNAs.

Flow chart showing step-wise isolation of piRNAs from total small RNAs and subsequent mappings to different annotations (repeats, protein-coding genes and genome).

Figure S2. Chen et al.

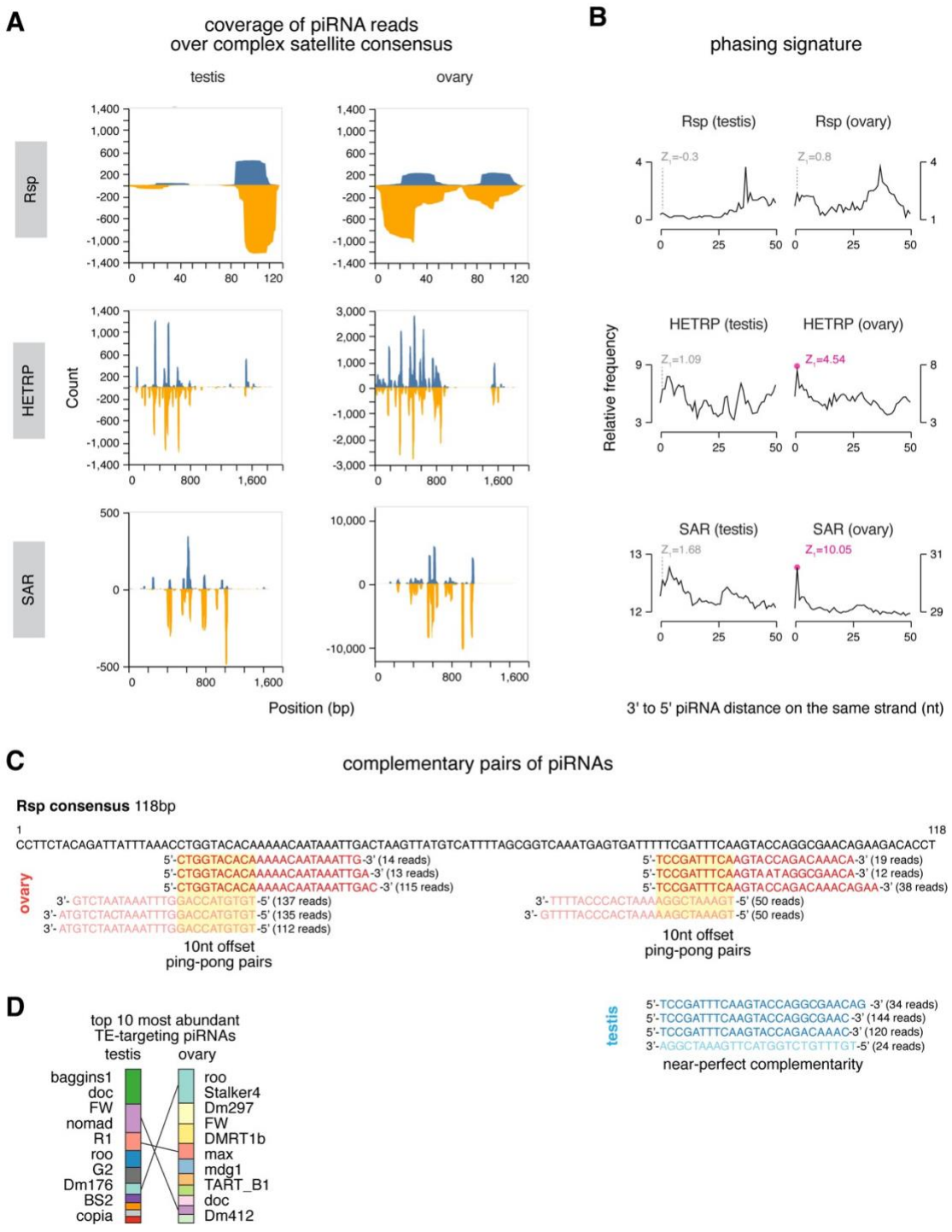


Figure S2. Profiles of complex satellite- and TE-mapping piRNAs in two sexes.

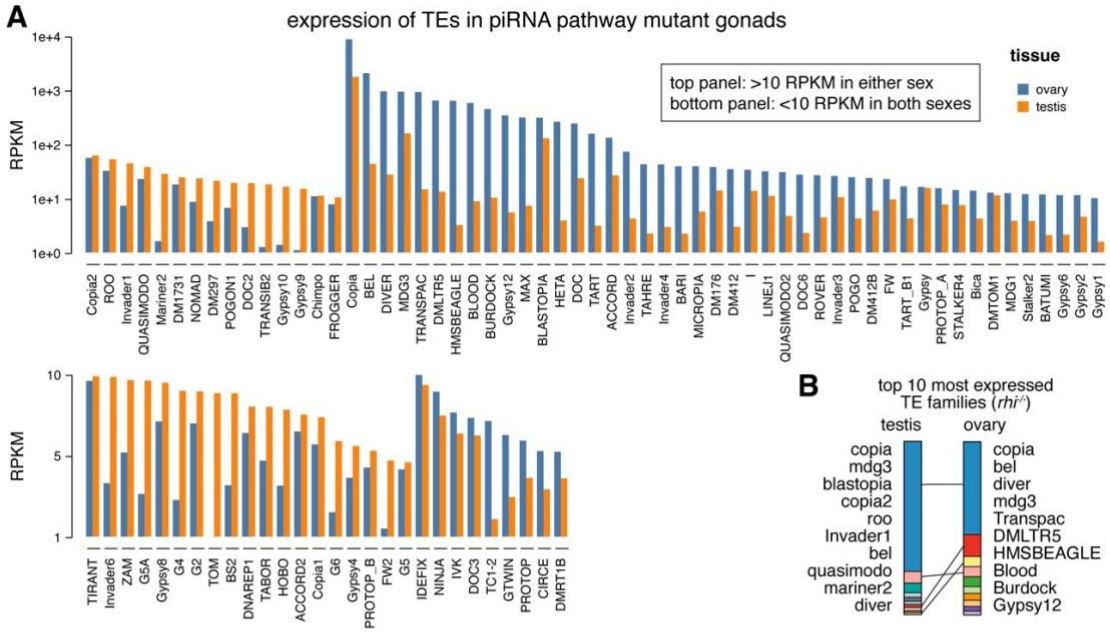
(A) Coverage plots of piRNAs over *Rsp* (top), *HETRP* (middle) and *SAR* (bottom), in testis (left) and ovary (right).

(B) Distributions of 3'-to-5' piRNA distance on the same strand for each of the three complex satellites in testis and ovary. Z_1 score marks the statistical significance of a putative enrichment of 1nt ($P<0.05$ for $z>1.96$). Significant ones are marked pink with the rest in gray.

(C) Examples of complementary pairs of *Rsp*-mapping piRNAs. Note that in ovary (red) they show an enrichment for 10-nt overlap, i.e., ping-pong signature, but in testis (blue) they show near perfect-complementarity with no evidence for ping-pong signature.

(D) Top 10 TEs targeted by the most abundant antisense piRNAs in testis (left) and ovary (right). Heights of slices correspond to relative abundance in each sex, and the sum of top 10 TEs is then scaled to the same height between sexes. Each TE family is given a unique color, and the same TE family is connected by a line to help visualize distinct rank orders between sexes. Names of TE families are shown following the same order, though not directly next to respective slices.

Figure S3. Chen et al.



C profiles of 36 TEs regulated by Rhi in at least one sex

TE_merge	ovary_TE_fc	ovary_silencing potential	ovary_pi	ovary_pi_FC	testis_TE_fc	testis_silencing potential	testis_pi	testis_pi_FC	Derepressed in	CAN EXPLAIN	Note
MDG3	382.2	TRUE	2098.3	4.3	33.2	TRUE	2944.7	8	Both sex	YES	
ACCORD	109.6	TRUE	1282.2	15.4	14.4	TRUE	496.8	9.6	Both sex	YES	
Copia	114.4	TRUE	1110.3	4.4	12.5	TRUE	5141.1	26.4	Both sex	YES	
Invader3	9.7	TRUE	731.9	20.4	12.1	TRUE	4857.3	81.8	Both sex	YES	
I	22.2	TRUE	2632.6	9.3	7.9	TRUE	2586.3	19.6	Both sex	YES	
DIVER	70.7	TRUE	1563.2	9.2	3.6	TRUE	1985.5	2.9	Both sex	YES	
MICROPIA	100	TRUE	1357.9	25.7	3.5	TRUE	2531.6	124	Both sex	YES	
BS2	2.9	TRUE	7253.5	39.6	7.8	TRUE	5274.9	105.7	Testis	YES	testis biased TE
TRANSIB2	1.4	TRUE	1062.1	20.1	4.2	TRUE	3090.7	161.3	Testis	YES	testis biased TE
DOC2	1.9	TRUE	2090.2	13	3	TRUE	725.3	3.6	Testis	YES	testis biased TE
DM297	0.9	FALSE	27748.2	0.3	4.8	FALSE	3626.7	0.7	Testis	?	TE derepression w/o tissue-wide piRNA loss
TOM	1	FALSE	137.2	0.7	3.8	FALSE	28.7	1.3	Testis	?	TE derepression w/o tissue-wide piRNA loss
ROO	1.6	FALSE	48306.1	0.5	3.1	FALSE	8654.7	0.8	Testis	?	TE derepression w/o tissue-wide piRNA loss
Gypsy12	389.1	TRUE	2051.2	29.3	2.5	TRUE	854.4	3.6	Ovary	YES	ovary biased TE
BLASTOPIA	109.2	TRUE	2319.5	7.7	2.1	TRUE	3376	16.5	Ovary	YES	ovary biased TE
MAX	93.7	TRUE	16354	38.9	1.2	TRUE	1505.1	20.5	Ovary	YES	ovary biased TE
Copia2	42.8	TRUE	1856.3	25.6	1.8	TRUE	1648.4	23	Ovary	YES	ovary biased TE
DM1731	40.3	TRUE	1047.1	23.9	2.1	TRUE	713.9	3.4	Ovary	YES	ovary biased TE
DMLTR5	27	TRUE	108.3	2.1	1.5	TRUE	377.9	12.8	Ovary	YES	ovary biased TE
Invader1	5.2	TRUE	2021.1	32.1	1.6	TRUE	872.7	41.5	Ovary	YES	ovary biased TE
BURDOCK	317	TRUE	3755.7	47.1	2	FALSE	69.9	2	Ovary	YES	ovary biased TE
HETA	238.5	TRUE	5725.4	30.9	0.4	FALSE	53.3	0.3	Ovary	YES	ovary biased TE
TRANSPAC	192.6	TRUE	751.1	6.3	0.9	FALSE	43.6	2	Ovary	YES	ovary biased TE
BEL	130.6	TRUE	2245.7	8.7	0.6	FALSE	145.4	0.5	Ovary	YES	ovary biased TE
TART	80.1	TRUE	4354.2	2.6	0.3	FALSE	2756.7	0.6	Ovary	YES	ovary biased TE
HMSBEAGLE	69.2	TRUE	2461	4.2	2.3	FALSE	1871.8	1.7	Ovary	YES	ovary biased TE
TART_B1	23.2	TRUE	14384.2	15	0.8	FALSE	2709.2	1.7	Ovary	YES	ovary biased TE
LINEJ1	20.5	TRUE	2184.4	17	1.9	FALSE	40.4	1	Ovary	YES	ovary biased TE
Invader2	19.6	TRUE	2859.8	7.1	2.1	FALSE	4728.2	1.2	Ovary	YES	ovary biased TE
TAHRE	18.9	TRUE	5589	53.1	0.5	FALSE	74.4	0.5	Ovary	YES	ovary biased TE
DOC	9.3	TRUE	12492.9	14.8	2.6	FALSE	22434.2	2	Ovary	YES	ovary biased TE
QUASIMODO	8.2	TRUE	5956.2	2.2	2.1	FALSE	2290	1.4	Ovary	YES	ovary biased TE
FW	6.7	TRUE	22035	6.5	2.2	FALSE	14250.2	1.4	Ovary	YES	ovary biased TE
DOC6	4.4	TRUE	2051.2	9.3	2.5	FALSE	69.3	1.8	Ovary	YES	ovary biased TE
QUASIMODO2	28.6	FALSE	1169.4	1.3	1.9	TRUE	351.7	5.4	Ovary	?	TE derepression w/o tissue-wide piRNA loss
BLOOD	50.9	FALSE	5419.9	1	1.5	FALSE	3524.8	0.9	Ovary	?	TE derepression w/o tissue-wide piRNA loss

"silencing potential" is TRUE when 1) there is >100 RPM piRNAs normally AND 2) there is >2-fold piRNA loss in rhi mutants.

Figure S3. TE levels in piRNA pathway mutants and curation of TEs regulated by Rhi in at least one sex.

(A) Bar graphs showing TE levels in piRNA pathway mutant (*rhi*) testes (orange) and ovaries (blue). TEs that have >10 RPKM in either sex is shown at the top, with the rest at the bottom.

(B) Top ten most expressed TE families in piRNA pathway mutant testis (left) and ovary (right). *rhi*$^{-/-}$ was used, where piRNA production from genome-wide dual-strand piRNA clusters collapses. Slice heights and colors were depicted as described in Supplementary Fig. S2D, though the same TE can be marked by a different color.

(C) Table reporting manual curation of 36 confidently affected TE families by *rhi*$^{-/-}$. Silencing potential is TRUE when there are normally >100 RPM antisense piRNAs and they show >2-fold reduction in *rhi* mutants. TEs are deemed de-repressed when having >3-fold up-regulation. Note a few unexpected cases where TE de-repression is not accompanied by piRNA loss, the ovary ones of which were described before (Klattenhoff et al. 2009).

Figure S4. Chen et al.

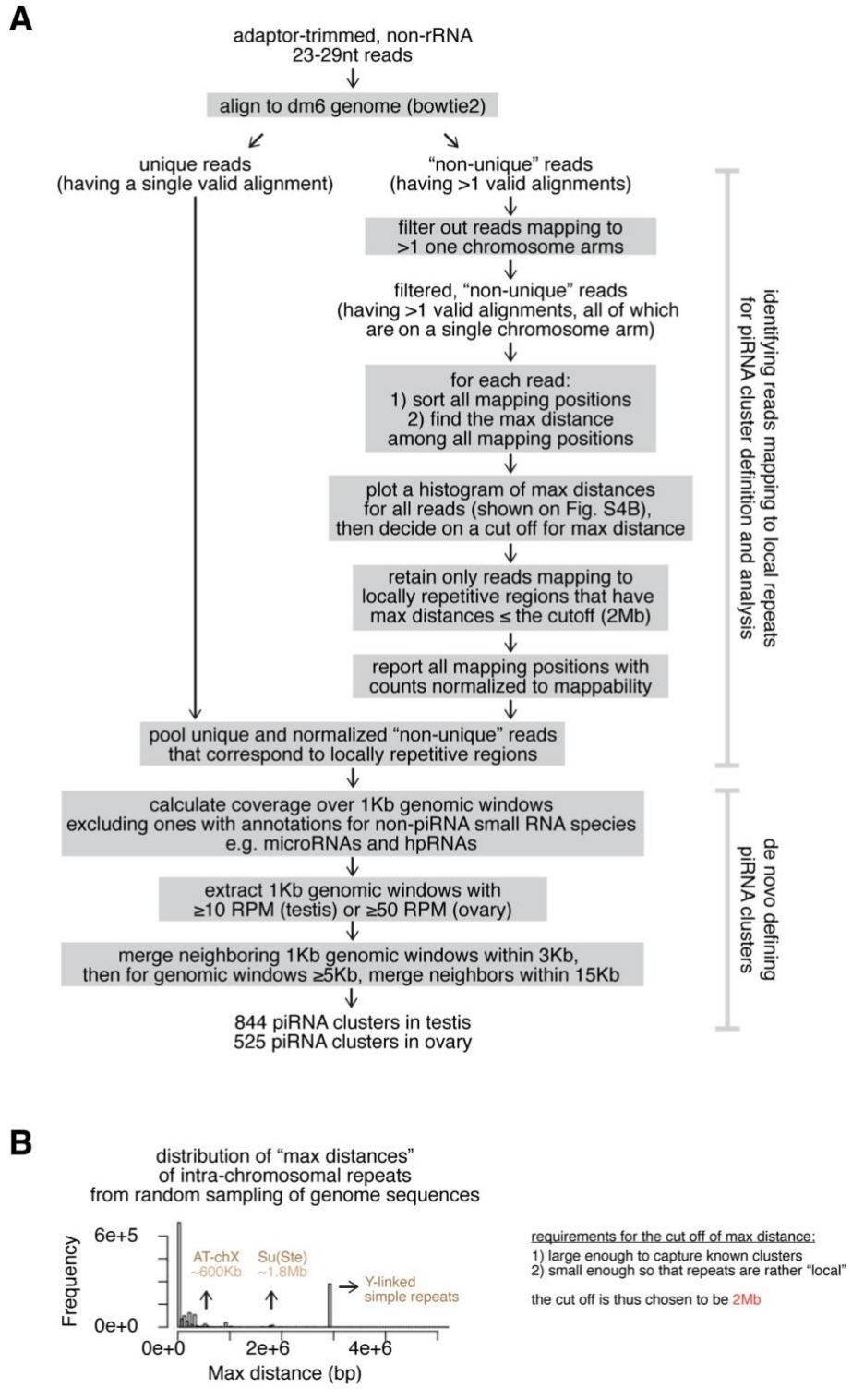


Figure S4. An algorithm that includes local repeats in piRNA cluster definition and analysis.

(A) Flow chart showing steps of the new algorithm that includes local repeats in piRNA cluster definition and analysis. See also methods.

(B) Histogram showing the distribution of “max distances” defined in (A) to identify a meaningful cutoff (2Mb) for distinguishing local from non-local repeats. See also methods.

Figure S5. Chen et al.

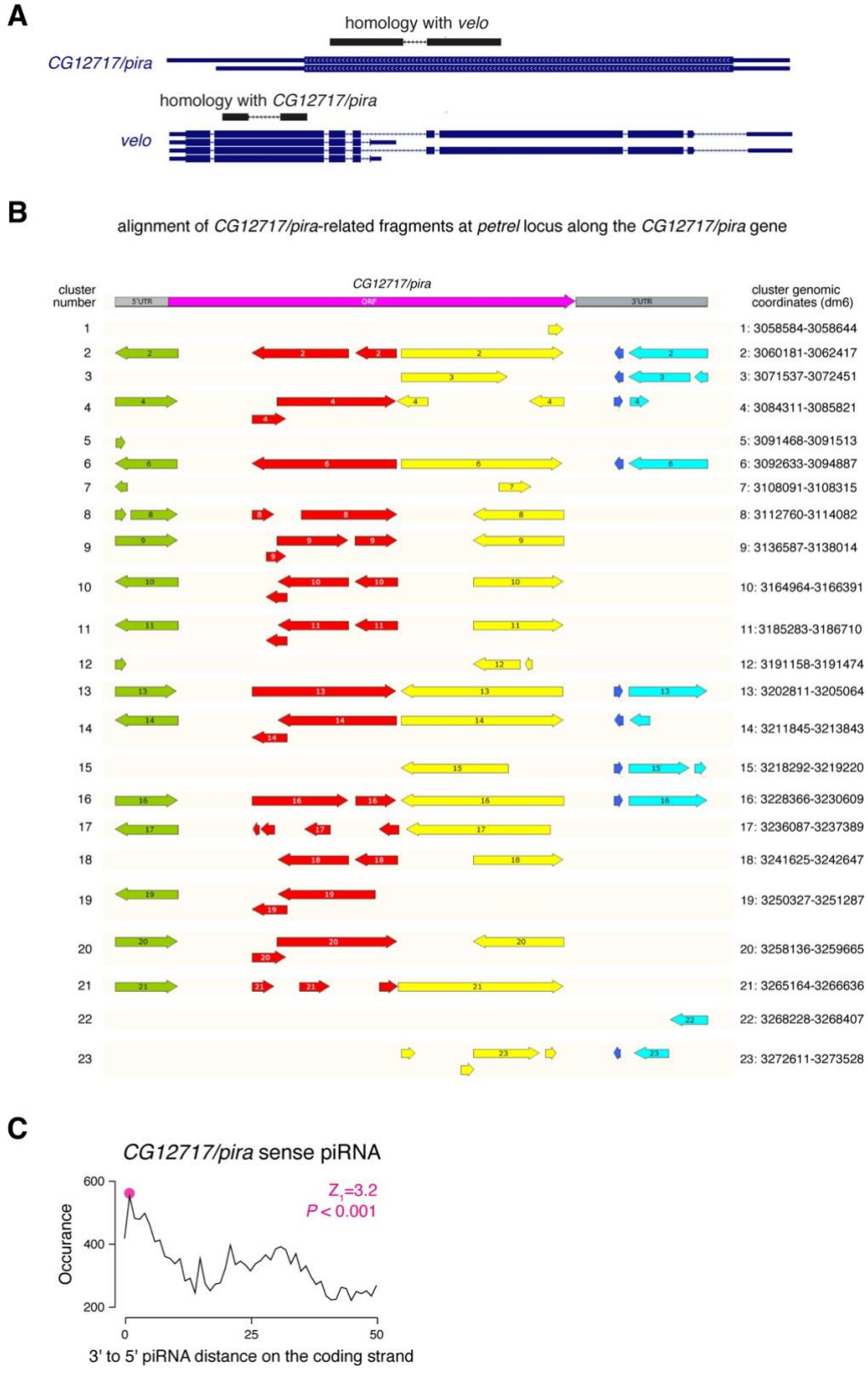


Figure S5. Characterization of *pira* homology in *D. melanogaster*, and the phasing pattern of *pira*-derived sense piRNAs.

(A) Homology between two *D. melanogaster* paralogs: *velo* and *CG12717/pira*. The homologous regions are marked using BLAT and they share 75% nucleotide sequence identity.

(B) Alignment of duplicated, partial copies of *CG12717* at *petrel* on *D. melanogaster* Y to its *CG12717* gene (left), and their genomic coordinates (right). Note that there are two small regions of *CG12717* absent on Y. RNA *in situ* HCR was targeted against the ORF region unique to *CG12717* gene.

(C) Distribution of 3'-to-5' piRNA distance on the coding strand of *CG12717/pira* in testis. Only genome-unique sense piRNAs with zero mismatch are used. Z_1=3.2 is equivalent to $P<$ 0.001.

Chapter 4

RDC COMPLEX EXECUTES A DYNAMIC PIRNA PROGRAM DURING DROSOPHILA SPERMATOGENESIS TO SAFEGUARD MALE FERTILITY

P. Chen, Y. Luo, A. A. Aravin (2021).
PLoS Genet **17**, e1009591. https://doi.org/10.1371/journal.pgen.1009591

ABSTRACT

piRNAs are small non-coding RNAs that guide the silencing of transposons and other targets in animal gonads. In *Drosophila* female germline, many piRNA source loci dubbed "piRNA clusters" lack hallmarks of active genes and exploit an alternative path for transcription, which relies on the Rhino-Deadlock-Cutoff (RDC) complex. RDC was thought to be absent in testis, so it remains to date unknown how piRNA cluster transcription is regulated in the male germline. We found that components of RDC complex are expressed in male germ cells during early spermatogenesis, from germline stem cells (GSCs) to early spermatocytes. RDC is essential for expression of dual-strand piRNA clusters and transposon silencing in testis; however, it is dispensable for expression of Y-linked *Suppressor of Stellate* piRNAs and therefore *Stellate* silencing. Despite intact *Stellate* repression, males lacking RDC exhibited compromised fertility accompanied by germline DNA damage and GSC loss. Thus, piRNA-guided repression is essential for normal spermatogenesis beyond *Stellate* silencing. While RDC associates with multiple piRNA clusters in GSCs and early spermatogonia, its localization changes in later stages as RDC concentrates on a single X-linked locus, *AT-chX*. Dynamic RDC localization is paralleled by changes in piRNA cluster expression, indicating that RDC executes a fluid piRNA program during different stages of spermatogenesis. These results disprove the common belief that RDC is dispensable for piRNA biogenesis in testis and uncover the unexpected, sexually dimorphic and dynamic behavior of a core piRNA pathway machinery.

Main Text

INTRODUCTION

Transposable elements (TEs) are selfish genetic elements that have the ability to propagate in the genome. When unchecked, transposition of TEs can cause overwhelming DNA damage and, eventually, genome instability. This poses a particular threat to germ cells, and TE de-repression often leads to reproductive defects like sterility. To cope with this, a small RNA-mediated genome defense mechanism involving PIWI proteins and PIWI-interacting RNAs (piRNAs) is employed in animal gonads to silence TEs (*1*).

In ovaries of *Drosophila melanogaster*, most piRNAs are made from so-called dual-strand piRNA clusters, where both genomic strands are transcribed to give rise to piRNA precursors. Transcription of dual-strand piRNA clusters is unusual in a number of ways. First, there is no clear promoter signature for transcription initiation (*2*). Second, splicing, termination and poly-adenylation of nascent transcripts are all suppressed (*2–4*). Third, transcription occurs at the presence of H3K9me3 (*5*), a histone modification generally seen as a repressive mark for gene expression. In fact, canonical transcription must be repressed by factors like Maelstrom to allow proper piRNA production from dual-strand piRNA clusters (*6*). Prior work has shown that such non-canonical transcription and co-transcriptional processing of piRNA precursors depend on the RDC complex, composed of Rhino (Rhi), Deadlock (Del) and Cutoff (Cuff) proteins, that bind dual-strand clusters (*2*, *4*). Rhi belongs to the HP1 family and binds H3K9me3 through its chromo-domain, anchoring the RDC complex onto dual-strand clusters (*2, 4, 7*). Cuff is a homolog of the conserved cap-

binding protein Rai1 that was reported to suppresses both splicing (*4*) and transcriptional termination (*3*), in order to facilitate the production of long, unspliced piRNA precursors. Del, on the other hand, recruits a paralog of transcription initiation factor TFIIA-L, Moonshiner (Moon), to initiate transcription in hostile heterochromatin environment (*8*). Together, the RDC complex conveys transcriptional competence to dual-strand piRNA clusters, and the majority of piRNA production collapses when one of its components is missing.

While piRNA pathway is known to be active in both male and female germline of *Drosophila*, expression and functions of Rhi, Del and Cuff were studied exclusively during oogenesis. Mutations of *rhi*, *del* and *cuff* were shown to cause female sterility, however, mutant males remained fertile (*9, 10*). In addition, *rhi*, *del* or *cuff* are predominantly expressed in ovaries, with low or no expression in testes and somatic tissues (*11–13*). Similarly, Moon was also believed to be ovary-specific, given that a high expression level could only be found in ovaries (*8*). Finally, the silencing of *Stellate* by abundant *Suppressor of Stellate* piRNAs was shown to be unperturbed in testes of *rhi* and *cuff* mutants, suggesting that the RDC complex is dispensable for piRNA biogenesis in males (*14, 15*). Collectively, these findings led to the notion that RDC complex is dispensable for piRNA pathway in male germline, raising the question of how piRNA cluster expression is controlled during spermatogenesis.

Here, we describe a developmentally regulated assembly of RDC complex in testes. We found that low expression of RDC complex components can be attributed to the fact that only a small subset of cells at early stages of spermatogenesis express *rhi*, *del* and *cuff*. Loss of RDC complex in testes results in a collapse of piRNA production, TE de-repression, and, ultimately, compromised male fertility, supporting an indispensable role of RDC complex in spermatogenesis. Even though RDC complex is assembled and functional in both sexes, we found differential genome occupancies of RDC complex between two sexes, correlating with sexually dimorphic usage of genome-wide piRNA source loci. Finally, RDC complex appears to exhibit dynamic binding on different piRNA clusters during spermatogenesis, allowing different piRNAs source loci to be used at different stages of early sperm development.

RESULTS

Components of the RDC complex are required for normal male fertility

Previous studies showed that, while *rhi* is required for female fertility, it is dispensable for male fertility (*13*, *14*). In agreement with this, we found that *rhi* mutant males indeed produce progeny when crossed with wildtype females. However, careful examination of the male fertility by sperm exhaustion test (*16*) revealed significantly compromised fertility in *rhi* mutant males. Even though most *rhi* mutant males were initially fertile, the percentage of fertile males dropped as they aged and stayed low from day 3 in comparison to heterozygous sibling controls (Figure 1A). To probe male fertility more quantitatively, we repeated the test and counted numbers of progeny for each male every day. We found that even young, 1-day-

old *rhi* mutant males, which were fertile, produced fewer progeny than heterozygous sibling controls (Figure 1A). Also, *rhi* mutant males produced nearly no progeny after two-day sperm exhaustion, while heterozygous sibling controls continued to produce ~100 progeny on average throughout the sperm exhaustion process. To extend this observation to other components of the RDC complex, we repeated the sperm exhaustion tests for *del* and *cuff* mutant males. Both *del* and *cuff* mutants displayed a reduction in male fertility compared to their respective heterozygous sibling controls (n=5, $P \leq 0.001$, Figure S1) These results demonstrate that male fertility is substantially compromised at the absence of *rhi*, *del* or *cuff*, suggesting an indispensable role of all three RDC complex components in maintaining normal male fertility.

RDC complex is assembled in nuclei of germ cells from GSCs to early spermatocytes

The dependency of normal male fertility on *rhi*, *del* and *cuff* prompted us to re-examine whether RDC complex is assembled in testis. modENCODE data and previous work showed that tissue-wide mRNA levels of *rhi*, *del* and *cuff* are high in ovaries but low in testes and the soma (*11–13*), which led to the notion that RDC might be ovary-specific (*12*, *13*, *17*). To examine expression of Rhi, Del and Cuff in testis, we took an imaging-based approach that provides single-cell resolution and preserves spatial information. We examined expression of individual components of the RDC complex using GFP-tagged Rhi, Del and Cuff transgenes that are expressed under the control of their native regulatory regions (*2*). Importantly, GFP-tagged RDC components are functional, as their expression fully rescued the sterility of the respective female mutant (Figure S2). All three proteins are expressed at

the apical tip of testis that contain germ cells at early steps of spermatogenesis (Figure 1B and 1C), indicating that all three components of RDC complex are expressed in testis, though only in a small subset of the cells.

Rhi, Del and Cuff form foci in nuclei (Figure 2A). To test whether all three proteins co-localize in nuclear foci, we tagged Rhi with a different fluorophore and expressed it using the previously described regulatory region of *rhi* (*4*). After verifying that this transgene rescues female sterility of the *rhi* mutation (Figure S2), we analyzed its localization with Del and Cuff. Indeed, Rhi co-localizes with both Del and Cuff (Figure 2A), consistent with the formation of RDC complex. Next, we examined the inter-dependence of Rhi, Del and Cuff localization (Figure 2B). In either *del* or *cuff* mutants, Rhi becomes dispersed and no longer forms puncta in nuclei. Similarly, Del also disperses at the absence of Rhi or Cuff. Expression of Cuff is strongly decreased in both *rhi* and *del* mutants, indicating its destabilization. Therefore, Rhi, Del and Cuff co-localize in distinct nuclear foci that depend on the simultaneous presence of all three proteins.

We further characterized the expression of RDC complex in testis. Rhi, Del and Cuff are expressed in nuclei of germline stem cells (GSCs) that are directly adjacent to somatic hub cells labeled by Fas3, but not in hub cells (Figure 2C). Rhi expression continues beyond spermatogonia marked by Bam, until early spermatocytes that express Sa (Spermatocyte arrest) (Figure 2D and S3). In *bam* mutant testes, where spermatogenesis is arrested at the spermatogonia-to-spermatocyte transition stage, we observed an expansion of

spermatogonia and expression of Rhi throughout entire testes (Figure 2E). In addition to germ cells and hub cells, the apical tip of testes contains somatic gonadal cells (cyst stem cells and early cyst cells) that can be distinguished from germ cells by Tj expression. Rhi is not expressed in somatic cells that express Tj (Figure 2D and S3), confirming its restriction to the germline. Taken together, we conclude that RDC complex is assembled in male germ cells during spermatogenesis, from GSCs to spermatogonia and early spermatocytes.

In ovaries, RDC complex is known to promote piRNA cluster transcription by two mechanisms: 1) suppression of premature transcriptional termination, a function mediated by Cuff (*3*), and 2) licensing of non-canonical transcriptional initiation, a function that requires the recruitment of a basal transcriptional factor TFIIA-L paralog, Moonshiner (Moon) (*8*). Expression of Moon was reported to be specific to female germline, raising the question of whether RDC complex can fulfill its function in male germline, if its functional partner is missing. However, we observed Moon expression in testis using a GFP-tagged Moon transgene (expressed under its native regulatory region) that is able to rescue the female sterility caused by the *moon* mutation (*8*) (Figure 2A). Importantly, Moon is expressed at similar stages as components of RDC complex, from GSCs to spermatogonia and early spermatocytes at the apical tip of testis (Figure 2A and 2C). Furthermore, Moon co-localizes with Rhi in nuclear foci from late spermatogonia to early spermatocyte, and its focal localization is abolished in *rhi* mutants (Figure 2A and 2B). In GSCs, Moon localization is more diffused than components of the RDC complex (Figure 2A and 2C). However, Moon expression is perturbed in *rhi* mutants, even in GSCs (Figure 2B), suggesting that Moon

depends on Rhi for protein stability and proper localization in the nucleus throughout its expression window. On the contrary, Rhi localization appears normal in two different *moon* mutants, $moon^{\Delta 1}$ and $moon^{\Delta 28}$, indicating that it acts genetically downstream of RDC complex (Figure S4). These observations suggest that RDC complex can recruit Moon to license transcription initiation in the male germline.

Loss of RDC complex causes DNA damage and germ cell death in testis

To identify cellular mechanisms underlying fertility decline in males lacking RDC complex, we examined morphology of germ cells marked by Vasa-GFP in *rhi* mutants. Normally, Vasa-positive germ cells are tightly packed at the apical tip of testis, as we observed in testes of heterozygous control males (n=162; Figure 3A). However, half of *rhi* mutant testes (54.9%, n=134/244) had visibly fewer germ cells with prominent gaps in between, indicative of an elevation of germ cell death (Figure 3A). Furthermore, another quarter of *rhi* mutant testes (25.4%, n=62/244) completely lost early germ cells altogether, and only 19.7% (n=48/244) showed wildtype-like germline morphology (Figure 3A and 3B). We concluded that loss of Rhi leads to a reduction in the germ cell count in testis.

Next, we examined impacts of Rhi loss on the resident GSC population. We quantified the number of GSCs per testis by counting the number of Vasa-positive germ cells directly adjacent in 3D to the somatic GSC niche labeled by Fas3. We found a reduction of GSCs in testes of *rhi* mutants compared with heterozygous controls in two age groups (1-

4 and 9–12 days old) (P<0.0001, Mann–Whitney–Wilcoxon test, Figure 3C). About a quarter of *rhi* mutant testes did not have any GSC at all. Accordingly, we observed an expansion of Fas3-positive hub at a similar rate (24.4%, n=59/244), which usually occurs at the absence of GSCs and is never seen in control testes (Figure 3A). Hence, GSC population sizes shrink drastically in testes lacking Rhi. Similar to *rhi* mutants, testes of 10-day-old *del* and *cuff* mutants are often completely depleted of early germ cells including GSCs (Figure 3D), indicating that loss of any component of the RDC complex leads to a collapse of spermatogenesis. Notably, aged virgin mutant males that lost all early germ cells, nevertheless, harbor mature sperm in their seminal vesicles (Figure 3D), suggesting that disruption of the RDC complex does not block spermatogenesis at a specific stage in young males, but early germ cells are depleted when they age. Staining testes for the phosphorylated H2A variant (γ-H2Av), a marker for DNA double-strand breaks (DSBs), revealed massive accumulation of unrepaired DNA DSBs in early germ cells of *rhi*, *del* and *cuff* mutant testes (Figure 3E). Unrepaired DNA DSBs likely causes germ cell death in mutants of RDC complex components. Overall, our results suggest that the loss of early germ cells, including GSCs, accompanied by widespread unrepaired DNA DSBs is responsible for the compromised fertility of mutant males lacking an intact RDC complex.

RDC complex is required for TE silencing in testis

Widespread DNA DSBs can result from TE transposition. To quantify TE expression, we sequenced polyadenylated (polyA+) RNAs from *rhi* mutant and control testes. PolyA+ RNA-seq revealed 32 TE families showing significant up-regulation in testes of *rhi* mutants

(>2-fold increase, FDR < 0.05; Figure 3G). Among them, the most de-repressed ones include *mdg3* (36-fold), *invader3* (18-fold) and *copia* (13-fold). To verify TE de-repression, we employed a *copia-lacZ* reporter, where the LTR of *copia* containing *copia* promoter is fused upstream to the *lacZ* gene and its expression can be directly examined by X-gal staining (*18*). *copia* is known to be active in the male germline (*19*) and has the highest expression level among all TEs in testes (*20*). Whereas no detectable X-gal staining was seen in control testes, robust staining was observed in *rhi* mutants (Figure 3F), confirming strong de-repression of *copia* at the absence of *rhi*. Similarly, the *copia* reporter was de-repressed in *del* and *cuff* mutant testes (Figure 3F). These results show that TEs are de-repressed in testes lacking a functional RDC complex.

In ovaries, RDC complex is required for piRNA production from dual-strand clusters to ensure efficient TE silencing (*2–4*, *8*). To examine piRNA biogenesis, we sequenced and analyzed small RNAs in *rhi* mutant and control testes. We found a loss of antisense piRNAs targeting many TE families in *rhi* mutant testes (>2-fold reduction, FDR < 0.05; Figure 3H), suggesting an overall defect in piRNA production. In fact, there is a moderate correlation between the fold-derepression of TEs and the fold reduction of TE-targeting piRNAs in testis upon mutating *rhi* (log-transformed values: Spearman's $\rho = 0.41$, $P = 7.3 \times 10^{-5}$, Pearson's $\rho = 0.46$, $P = 8.4 \times 10^{-6}$). For TEs showing strong up-regulation in *rhi* mutant testes, we observed a concurrent, pronounced loss of antisense piRNAs (e.g., *mdg3, invader3* and *copia*). Notably, there are antisense piRNAs against several TE families (e.g., *BS2, Transib2, invader6*) that show over 100-fold reduction in *rhi* mutant testes, a magnitude not observed

for any TE family in *rhi* mutant ovaries (Figure 3I). Finally, sense piRNAs were also lost for many TE families (Figure S5), consistent with dual-strand clusters producing piRNAs from both genomic strands in a Rhi-dependent manner. These results show that efficient production of TE-silencing piRNAs in testis depends on the RDC complex, without which many TEs are de-repressed, causing DNA damage and germ cell death in testis.

RDC complex is required for piRNA production from dual-strand clusters in early male germ cells

To understand the role of RDC complex in piRNA cluster expression in testis, we analyzed effects of *rhi* mutation on piRNA production from major piRNA clusters. Genomic loci that generate piRNAs in testis were recently *de novo* defined leading to identification of several novel piRNA clusters (*20*). Rhi was dispensable for expression of uni-strand piRNA clusters, *flam* and *20A* (Figure 4A), similar to results from ovary studies (*2*, *4*, *14*). Surprisingly, we found that Rhi was also dispensable for piRNA production from the Y-linked *Su(Ste)* locus, which is the most active piRNA cluster in testis (*20*). Unlike *flam* and *20A*, *Su(Ste)* is a dual-strand cluster that generates piRNAs from both genomic strands. We confirmed by RNA fluorescence *in situ* hybridization (FISH) that piRNA precursor transcription from *Su(Ste)* appeared intact in testes without Rhi (Figure 4C). In contrast to *Su(Ste)*, piRNA production from other major dual-strand clusters, including the Y-linked *petrel* cluster, collapses in *rhi* mutant testes (Figure 4A), indicating that expression of the majority of dual-strand piRNA clusters in testis relies on Rhi. Interestingly, dependence of dual-strand cluster expression on Rhi varies between sexes: *38C* is more affected by loss of *rhi* than *42AB* in testis, while the

opposite is found in ovary. Furthermore, piRNA production from both strands of complex satellites, which we recently found to behave as dual-strand piRNA clusters (*20*), also drastically declined in *rhi* mutant testes and ovaries (Figure 4B). These results show that Rhi is essential for piRNA production from a large fraction of piRNA clusters in male germline.

Since RDC complex forms distinct foci in the nuclei of germ cells (Figure 2A), we set out to test if Rhi binds the chromatin of dual-strand clusters whose expression depends on RDC complex, as reported in ovary (*2, 4, 14*). Given that expression of RDC complex is restricted to a small number of cells at the apical tip of testis, we used *bam* mutant testes, where Rhi-expressing spermatogonia are expanded (Figure 2E), to perform ChIP-seq of Rhi. All major dual-strand clusters, with the exception of *Su(Ste),* were enriched for Rhi binding (Figure 4D). In agreement with ChIP-seq, independent ChIP-qPCR showed no evidence of Rhi binding on *Su(Ste)* locus (n=4) (Figure 4G). Rhi was also absent on chromatin of uni-strand clusters, *flam* and *20A*. Importantly, the binding of Rhi on different loci seems to correlate with its effect on promoting piRNA cluster expression. Rhi does not bind, and is dispensable for piRNA production from, uni-strand clusters and *Su(Ste),* while it binds, and is required for expression of, other dual-strand clusters (Figure 4A and 4D). Also, dual-strand clusters that show the highest levels of overall Rhi binding, *38C* and *AT-chX*, demonstrate the strongest Rhi dependence for piRNA production. To characterize the relationship between Rhi binding and piRNA production on a genome-wide scale, we analyzed Rhi binding and piRNA production in 1Kb genomic windows spanning the entire genome (Figure 4E). For loci that depend on Rhi to produce piRNAs, we observed strong correlation between Rhi

binding and piRNA levels (Spearman's ρ = 0.97; Pearson's ρ = 0.99). On the other hand, loci that continue to produce piRNAs at the absence of Rhi usually have little, if any, Rhi binding. Collectively, our data indicate that Rhi physically binds the chromatin and ensures the expression of dual-strand piRNA clusters, with a notable exception of *Su(Ste)*.

Sexually dimorphic genome occupancy of RDC complex sculpts sex-specific piRNA program

piRNA profiles are distinct in male and female gonads, and expression of dual-strand piRNA clusters are sexually dimorphic (*20*). To explore if RDC complex might be involved in orchestrating sex-specific piRNA programs, we profiled Rhi binding on the genome in ovaries under identical ChIP-seq conditions as in testes. This analysis revealed differences in Rhi genome occupancy between sexes among top piRNA clusters (Figure 5A). For example, Rhi is more enriched on *38C* than *42AB* in testes, whereas the reciprocal is seen in ovaries, correlating with differential piRNA production from these two loci in two sexes. In addition, *80EF* and *40F7* clusters have high levels of Rhi binding in ovary but low in testis, mirrored by abundant piRNA production from these two loci in ovary but not in testis. Finally, an ovary-specific dual-strand piRNA cluster, *Sox102F*, is bound by Rhi in ovary, while there is no evidence of Rhi binding at *Sox102F* in testis where it is inactive (Figure 5B). Altogether, the observed link between Rhi binding and piRNA production between males and females suggests that the sex-specific Rhi binding on piRNA clusters is responsible for sculpting a sexually dimorphic piRNA program.

RDC complex enables dynamic piRNA production during spermatogenesis

ChIP-seq provides the genome-wide profile of Rhi binding, but it masks possible differences of Rhi localization among individual cells. Imaging of Rhi revealed distinct Rhi localization in nuclei of germ cells at different stages of spermatogenesis (Figure 6A). In the nuclei of GSCs and spermatogonia, Rhi forms many discrete foci, suggesting—in agreement with ChIP-seq results—that it binds multiple genomic loci. As male germ cells differentiate into spermatocytes and prepare for meiosis, however, Rhi concentrates as one single dot in nuclei of early spermatocytes, suggesting its specific localization at one single locus. Notably, while homologous chromosomes pair in about half of late spermatogonia and very early spermatocytes (*21*), we only observed one bright Rhi dot in virtually all nuclei at these stages, suggesting that this locus resides on one of the two sex chromosomes rather than an autosome. To explore where Rhi binds at this stage, we first examined testes from XO males that lack the Y chromosome and found that Rhi still localizes as one single dot in early spermatocytes, arguing against Y-linked loci such as *Su(Ste)* and *petrel* (Figure 6B). Simultaneous imaging of Rhi and RNA FISH of transcripts from the X-linked *AT-chX* locus revealed co-localization of *AT-chX* nascent transcripts and Rhi in one single dot from late spermatogonia to early spermatocytes, indicating that Rhi concentrates on *AT-chX* locus at this stage (Figure 6C). Indeed, single *AT-chX* RNA foci in individual nuclei became undetectable in *rhi* mutant testes, and expressing GFP-tagged Rhi transgene by *nanos-Gal4* in *rhi* mutant background restored *AT-chX* expression (Figure 6C). Importantly, even though *nanos-Gal4* drives stronger expression of GFP-Rhi in earlier stages (GSC and early spermatogonia), *AT-chX* transcripts remained highly expressed specifically during later

stages (late spermatogonia and early spermatocyte) (Figure 6C). This finding suggests that the spatio-temporally regulated gene expression of *AT-chX* piRNA cluster is rather robust to perturbations to the Rhi protein level, and the low expression of *AT-chX* piRNA cluster earlier in GSCs and early spermatogonia is not limited by the level of Rhi protein. In sum, these results show that Rhi binds multiple genomic loci in GSCs and spermatogonia but appears to concentrate on a single *AT-chX* locus later.

As Rhi is required for non-canonical transcription of dual-strand piRNA clusters, depletion of Rhi from clusters other than *AT-chX* should cease their expression. To test this, we set out to conduct RNA FISH of piRNA precursors from other Rhi-dependent dual-strand piRNA clusters (Figure 5A). Previously, FISH detection of piRNA precursor transcripts was performed in polyploid nurse cells in fly ovary, whose genome is endo-replicated up to 1032C with a much higher expression level of piRNA precursor transcripts (*2, 8, 22*). In addition to low expression levels in the diploid male germline, piRNA cluster transcripts are difficult to target, as they are highly repetitive and share extensive sequence homology with TEs. To tackle these challenges, we employed *in situ* HCR, which permits enzyme-free signal amplification and automatic background suppression (*23*), to target transcripts from *38C* piRNA cluster that has a relatively high expression level in testis (Figure 5A). *in situ* HCR detected nascent *38C* piRNA precursors that co-localized with Rhi in diploid male germ cells (Figure 6D). These signals from nascent *38C* piRNA cluster transcripts were absent in *rhi* mutant testes (Figure 6D). Moreover, both *38C* and *AT-chX* piRNA precursors can be seen in nuclei of spermatogonia. In contrast, only *AT-chX*, but

not *38C*, cluster continues to be expressed in 87.7% (n=50/57) of early spermatocyte nuclei (Figure 6E). Therefore, the expression of *38C* piRNA cluster is turned off when Rhi concentrates onto the *AT-chX* locus.

During early spermatogenesis, satellite DNAs located at peri-centromeric heterochromatin of different chromosomes cluster together to form distinct nuclear compartments called chromocenters (*24*, *25*). To explore the possibility that the repetitive *AT-chX* locus is also recruited to a chromocenter, we compared Rhi localization with that of D1, which binds the AATAT satellite on chromosome X, Y and 4 (*24*). We found that D1 does not co-localize with the single Rhi dot in early spermatocytes (Figure 6F). We also analyzed localization of Cid, a centromere-specific histone H3 variant that functions as an epigenetic mark for centromere identity (*26*). The single Rhi dot is present in nuclei that contain four or more separate Cid foci (Figure 6G), indicating that non-homologous chromosomes are separated and occupy distinct chromosome territories in these nuclei. These results show that the *AT-chX* locus bound by Rhi occupies a nuclear compartment distinct from chromocenters.

Taken together, our results suggest that RDC complex binds, and thus enables, the expression of many dual-strand piRNA clusters from GSCs to spermatogonia, including *38C*, but gradually concentrates onto a single locus, *AT-chX*, in early spermatocytes (Figure 6H). Dynamic association of RDC with piRNA clusters during early spermatogenesis executes a fluid piRNA program to allow different piRNA clusters to be engaged at different developmental stages.

DISCUSSION

Most proteins involved in piRNA pathway in *Drosophila* were initially identified in screens for female sterility or TE de-repression in ovaries (*9*, *10*, *27–30*). However, the first described case of piRNA repression, silencing of *Stellate* by *Su(Ste)* piRNAs during spermatogenesis, indicates that piRNA pathway operates in gonads of both sexes (*31*, *32*). Many proteins involved in piRNA pathway in ovaries are also required for male fertility and *Stellate* silencing in testes, supporting the conservation of piRNA pathway machinery between sexes (*31*, *33–39*). Notably, a few proteins stood out as exceptions: Rhi, Del and Cuff that form a complex to enable transcription of dual-strand piRNA clusters in ovary (*2*, *4*, *14*), for which no fertility defects or *Stellate* de-repression were observed in mutant males (*10*, *13*, *14*). This suggested that molecular mechanisms controlling piRNA cluster expression in testis might be different from ovary. Our results, however, demonstrate that RDC complex is assembled in testis (Figure 2) and required for TE silencing (Figure 3) in male germline, indicating that the molecular machinery regulating piRNA cluster expression is conserved between sexes.

piRNA pathway in *Drosophila* testis: beyond *Stellate* silencing

Genomic loci that encode ovarian piRNAs were systematically identified across the genome in several studies (*2*, *40*). In contrast, piRNA studies on testes were mostly focused on a single locus, *Su(Ste)*, that encodes piRNAs to silence *Stellate*, and no systematic search for piRNA clusters in testis has been performed to date. We recently *de novo* identified piRNA

clusters in testis (*20*), laying the foundation for broader understanding of piRNA biogenesis and function in male gonads. We found that RDC is essential for expression of all major dual-strand piRNA clusters in testis, with a remarkable exception of *Su(Ste)* (Figure 4). This explains the previous observation that *rhi* and *cuff* are dispensable for *Stellate* silencing (*14*, *15*). In contrast, many other piRNA pathway factors such as Aub, Ago3 and Zuc are involved in both *Stellate* and TE repression (*31*, *33*, *35*). Mutations in these genes cause dramatic disruption of spermatogenesis, often leading to complete male sterility. In comparison, *rhi, del* and *cuff* mutant males demonstrate milder fertility defects, suggesting that *Stellate* de-repression is the major cause of spermatogenesis failure in other piRNA pathway mutants. Indeed, previous studies demonstrated that *Stellate* de-repression induced by deleting the *Su(Ste)* locus alone, without global perturbations of piRNA pathway, disrupts spermatogenesis and causes male sterility (*41*). Thus, the finding that RDC is dispensable for *Stellate* repression provides a unique opportunity to understand impacts of silencing other piRNA targets in the male germline.

Our results indicate that piRNA-guided repression plays a crucial role in spermatogenesis beyond *Stellate* silencing, as *rhi, del* and *cuff* mutant males show rapid fertility decline, germline DNA damage and severe loss of germline content including GSCs (Figure 3). These phenotypes are likely caused by TE de-repression and the resultant genome instability, though we also identified complex satellites and a host protein-coding gene, SUMO protease *CG12717*/*pirate*, as targets of piRNA silencing in testes (*20*). In the future, it will be

important to disentangle the contributions of de-repressing different piRNA targets to spermatogenesis defects observed in testes of piRNA pathway mutants.

Even though the fertility of *rhi* mutant males is substantially compromised, they produce a small number of functional spermatozoa, at least when they are young (Figure 1). In contrast, females lacking *rhi, del* or *cuff* are completely sterile (*10, 13*). This distinction between the two sexes might result from differences in the TE threat faced by male and female germ cells. Different TE families are activated upon disruption of piRNA pathway in ovary and testis and, generally, there is a stronger TE threat in ovary (*20*). Differential TE de-repression in two sexes might be responsible for stronger defects in oogenesis at the absence of RDC. Alternatively, it might reflect differences in DNA damage response in two sexes. An egg is energetically more expensive to make than a spermatozoon. In line with this, DNA damage responses (activated by, e.g., TE transposition) often arrest oogenesis to avoid wasting resources on a defective egg (*42, 43*). Since oogenesis is usually shut down to attempt repair, incomplete oogenesis results in female sterility. In contrast, quality control mechanisms of spermatogenesis frequently kill unqualified germ cells (*44*), without pausing the developmental program of those surviving ones. Because spermatogenesis permits a large number of germ cells to develop in parallel, even though the unqualified ones are killed, a few surviving germ cells might be able to complete sperm development. Together, differential TE threats coupled with distinct response strategies to DNA damage might underlie sex-specific sterility when RDC is lost.

***Su(Ste)*: an RDC-independent dual-strand piRNA cluster free of RDC binding**

Su(Ste) locus on Y chromosome is the most prolific source of piRNAs in testes (*20*). piRNAs are generated from both genomic strands of *Su(Ste)* repeats, making it akin to other dual-strand piRNA clusters (*31*, *45*). However, our results showed that RDC is dispensable for expression of *Su(Ste)* piRNAs, while it is required for piRNA production from all other dual-strand clusters in testis (Figure 4A) and ovary (*2*, *4*). What might explain the ability of *Su(Ste)* to generate piRNAs in an RDC-independent fashion? RDC ensures transcription of piRNA precursors by suppressing their premature termination (*3*) and promoting non-canonical transcription initiation (*8*). Interestingly, the structure of *Su(Ste)* locus is different from other dual-strand piRNA clusters, as it is composed of many almost identical, relatively short units, all of which are flanked by two canonical, albeit convergent, promoters driving expression of both genomic strands (*31*). In fact, sense transcripts of *Su(Ste)* were found to be spliced and polyadenylated (*31*), consistent with the absence of RDC, which suppresses splicing and polyadenylation (*2–4*), at this locus. Thus, the presence of canonical promoters flanking individual short units of *Su(Ste)* repeats might enable their expression without engaging RDC complex.

Consistent with a role in promoting piRNA precursor transcription, Rhi is enriched on chromatin of dual-strand piRNA clusters. Such a correlation between direct Rhi binding and Rhi-dependent piRNA expression was previously reported in ovary (*2*). The function of RDC explains why piRNA production depends on its presence on chromatin; however, the molecular mechanism responsible for specific recruitment of RDC to piRNA clusters

remained poorly understood. It has been shown that piRNAs expressed during oogenesis are deposited into the oocyte and play an important role in jump-starting piRNA biogenesis in the progeny (*46*, *47*). Maternally supplied piRNAs were shown to induce deposition of Rhi on cognate genomic locus in the progeny (*7*). Thus, piRNAs expressed from the cluster and RDC binding to the cluster seem to form a positive feedback loop: RDC is required for piRNA production, and piRNAs in turn guide deposition of Rhi on cognate genomic loci. piRNA-dependent Rhi deposition might be mediated by the nuclear Piwi protein that directs the establishment of histone H3K9me3 mark (*48–51*), which provides a binding site for Rhi chromo-domain (*2*, *7*). Importantly, Piwi- and piRNA-dependent Rhi recruitment seems to occur in a narrow developmental window during early embryogenesis. This was demonstrated by the observation that depleting Piwi during early embryogenesis is sufficient to perturb Rhi localization on piRNA clusters, while depleting Piwi during larval or adult stages does not change Rhi localization (*52*). Our finding that Rhi is localized to all dual-strand piRNA clusters in testis except *Su(Ste)* (Figure 4D) is compatible with an idea that Rhi binding to genomic loci in the zygote is guided by maternal piRNAs. Indeed, in contrast to most other dual-strand clusters that are active in the germline of both sexes, Y-linked *Su(Ste)* locus generates piRNAs only in males. Therefore, unlike other piRNAs, *Su(Ste)* piRNAs are not deposited into the oocyte, resulting in the inability to recruit Rhi to *Su(Ste)* in the progeny. It is interesting to compare *Su(Ste)* with another dual-strand piRNA cluster on Y chromosome, *petrel*. Unlike *Su(Ste)*, Rhi is enriched on *petrel* chromatin and piRNA production from this cluster depends on Rhi (Figure 4A and 4D). However, *petrel* piRNAs derived from male-specific Y should be absent in ovaries and hence no *petrel* piRNAs can be deposited into the oocyte. At the first glance, these observations argue against a possibility

that maternal piRNAs guide Rhi deposition on *petrel* cluster. However, unlike *Su(Ste),* *petrel* is enriched of different TE sequences. As a result, TE-mapping piRNAs produced from other clusters in ovary might be able to target *petrel*. Indeed, piRNAs mapping to TEs present at the *petrel* locus can be found in unfertilized eggs (*46*). For example, *roo* piRNAs are the most abundant TE-mapping, maternally deposited piRNAs in the early embryo (*46*). Several *roo* fragments are present at *petrel,* and these sequences occupy ~6% of the total length of this cluster, making *roo* the third most represented TE at *petrel* (after *IDEFIX* and *ninja*) (*20*). piRNAs against other TEs located at *petrel* are also deposited in the embryo. Overall, while our results did not directly address the mechanism of Rhi recruitment to specific genomic loci, they show that studying Rhi occupancy on Y-linked piRNA clusters provides a novel angle to study this problem.

Dynamic organization of piRNA pathway during spermatogenesis

Our results showed that components of RDC complex are expressed exclusively in male germline during early stages of spermatogenesis, from GSCs to early spermatocytes (Figure 1; Figure 2). Interestingly, expression of Piwi and Ago3, two of the three PIWI proteins in *Drosophila*, is also restricted to the same developmental stages (*34*, *53*). Piwi is required for piRNA-guided transcriptional silencing in the nucleus (*48–51*), while Ago3 is involved in heterotypic ping-pong cycle in cytoplasm (*33*, *40*, *54*), indicating that these processes operate in the same cells that have RDC-dependent transcription of piRNA clusters. In contrast to RDC, Piwi and Ago3, expression of the third PIWI protein, Aub, continues through spermatocyte stage until meiosis (*34*). How can developing male germ cells be protected

when the piRNA pathway is greatly simplified? It is possible that the silencing network initiated by piRNA pathway factors earlier can self-sustain later. piRNAs produced with the help of RDC complex and Ago3-dependent heterotypic ping-pong could persist and continue to function through spermatocyte differentiation, as long as they load onto Aub.

Interestingly, the cessation of RDC, Piwi and Ago3 expression during spermatogenesis coincides with the mitosis-to-differentiation transition. This transition is accompanied by one of the most dramatic changes in gene expression programs, with the general transcriptional machinery replaced by the testis-specific ones like tTAF and tMAC (*55*, *56*). Thus, the following stage of male germline development can be seen as a less-protected window of spermatogenesis, providing an opportunity for TEs and other selfish genetic elements to thrive.

Restriction of the piRNA pathway to early stages of spermatogenesis contrasts with its activity during oogenesis. piRNA pathway factors appear to be expressed during all stages of oogenesis from GSCs to late-stage nurse cells. Recent studies reported several new factors involved in piRNA pathway in ovary, such as Moon, Boot, Nxf3, Panx, Arx and Nxf2 (*8*, *22*, *30*, *57–65*). These proteins are expressed at a low level in testis and their functions during spermatogenesis have not yet been reported. Our results suggest that, similar to RDC, these proteins might function in piRNA pathway in testis, and their low expression could be explained by restricted expression in early male germline. Indeed, we found that Moon co-expresses with, and acts genetically downstream of, RDC in testes (Figure 2; Figure S4).

Overall, our results suggest that piRNA pathway machinery is likely conserved between sexes. However, the developmental organization of piRNA pathway during gametogenesis is different: whereas the entire pathway is active throughout oogenesis, processes that require RDC, Piwi or Ago3 likely terminate when mitotic spermatogonia differentiate into spermatocytes and prepare for meiosis.

Dynamic expression of piRNA clusters during spermatogenesis

Our results revealed dynamic association of RDC complex with different genomic loci during spermatogenesis (Figure 6). In GSC and spermatogonia nuclei, RDC complex localizes to many foci, and ChIP-seq data indicate that Rhi associates with multiple dual-strand clusters. As germ cells differentiate into early spermatocytes, however, RDC gradually concentrates onto a single locus, *AT-chX,* on X chromosome. Since transcription of dual-strand piRNA clusters is dependent on RDC complex, the dynamic localization of RDC suggests that expression of piRNA clusters changes as germ cells progress from GSCs to early spermatocytes. Indeed, detection of nascent cluster transcripts revealed that *38C* is active early, but not later when most RDC concentrates onto *AT-chX*. Notably, though not dependent on Rhi, transcription of *Su(Ste)* piRNA cluster was shown to span a narrow window from late spermatogonia to early spermatocyte as well (*45*), likely reflecting the promoter activity that drives *Su(Ste)* transcription (*31*). Dynamic expression of piRNA clusters during spermatogenesis is also supported by the study that showed spermatogonia and spermatocytes have distinct piRNA populations (*66*). In contrast to dynamic localization of RDC complex and cluster expression in testes, previous work depicted a static view of

piRNA production in female gonads. There has been no evidence of dynamic localization of RDC complex to different clusters during oogenesis, and all clusters appear active throughout female germline development. It will be interesting to explore whether expression of piRNA clusters changed dynamically during oogenesis. Through studying a protein complex thought to be absent during spermatogenesis, we uncovered the sexually dimorphic and dynamic behaviors of a molecular machinery that drives dual-strand piRNA cluster expression during *Drosophila* gametogenesis.

REFERENCES

1. D. M. Ozata, I. Gainetdinov, A. Zoch, D. O'Carroll, P. D. Zamore, PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet* **20**, 89–108 (2019).

2. F. Mohn, G. Sienski, D. Handler, J. Brennecke, The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in Drosophila. *Cell* **157**, 1364–1379 (2014).

3. Y.-C. A. Chen, E. Stuwe, Y. Luo, M. Ninova, A. Le Thomas, E. Rozhavskaya, S. Li, S. Vempati, J. D. Laver, D. J. Patel, C. A. Smibert, H. D. Lipshitz, K. Fejes Toth, A. A. Aravin, Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. *Molecular Cell* **63**, 97–109 (2016).

4. Z. Zhang, J. Wang, N. Schultz, F. Zhang, S. S. Parhad, S. Tu, T. Vreven, P. D. Zamore, Z. Weng, W. E. Theurkauf, The HP1 Homolog Rhino Anchors a Nuclear Complex that Suppresses piRNA Precursor Splicing. *Cell* **157**, 1353–1363 (2014).

5. P. Rangan, C. D. Malone, C. Navarro, S. P. Newbold, P. S. Hayes, R. Sachidanandam, G. J. Hannon, R. Lehmann, piRNA Production Requires Heterochromatin Formation in Drosophila. *Current Biology* **21**, 1373–1379 (2011).

6. T. H. Chang, E. Mattei, I. Gainetdinov, C. Colpan, Z. Weng, P. D. Zamore, Maelstrom Represses Canonical Polymerase II Transcription within Bi-directional piRNA Clusters in Drosophila melanogaster. *Molecular Cell* **73**, 291–303.e6 (2019).

7. A. Le Thomas, E. Stuwe, S. Li, J. Du, G. Marinov, N. Rozhkov, Y.-C. A. Chen, Y. Luo, R. Sachidanandam, K. F. Toth, D. Patel, A. A. Aravin, Transgenerationally inherited

piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev.* **28**, 1667–1680 (2014).

8. P. R. Andersen, L. Tirian, M. Vunjak, J. Brennecke, A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* **549**, 54–59 (2017).

9. C. A. Berg, A. C. Spradling, Studies on the rate and site-specificity of P element transposition. *Genetics* **127**, 515–524 (1991).

10. T. Schüpbach, E. Wieschaus, Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129**, 1119–1136 (1991).

11. J. B. Brown, N. Boley, R. Eisman, G. E. May, M. H. Stoiber, M. O. Duff, B. W. Booth, J. Wen, S. Park, A. M. Suzuki, K. H. Wan, C. Yu, D. Zhang, J. W. Carlson, L. Cherbas, B. D. Eads, D. Miller, K. Mockaitis, J. Roberts, C. A. Davis, E. Frise, A. S. Hammonds, S. Olson, S. Shenker, D. Sturgill, A. A. Samsonova, R. Weiszmann, G. Robinson, J. Hernandez, J. Andrews, P. J. Bickel, P. Carninci, P. Cherbas, T. R. Gingeras, R. A. Hoskins, T. C. Kaufman, E. C. Lai, B. Oliver, N. Perrimon, B. R. Graveley, S. E. Celniker, Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**, 393–399 (2014).

12. D. Vermaak, S. Henikoff, H. S. Malik, Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. *PLoS Genet.* **1**, 96–108 (2005).

13. A. M. Volpe, H. Horowitz, C. M. Grafer, S. M. Jackson, C. A. Berg, Drosophila rhino encodes a female-specific chromo-domain protein that affects chromosome structure and egg polarity. *Genetics* **159**, 1117–1134 (2001).

14. C. Klattenhoff, H. Xi, C. Li, S. Lee, J. Xu, J. S. Khurana, F. Zhang, N. Schultz, B. S. Koppetsch, A. Nowosielska, H. Seitz, P. D. Zamore, Z. Weng, W. E. Theurkauf, The Drosophila HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters. *Cell* **138**, 1137–1149 (2009).

15. M. V. Kibanov, K. S. Egorova, S. S. Ryazansky, O. A. Sokolova, A. A. Kotov, O. M. Olenkina, A. D. Stolyarenko, V. A. Gvozdev, L. V. Olenina, A novel organelle, the piNG-body, in the nuage of *Drosophila* male germ cells is associated with piRNA-mediated gene silencing. *MBoC* **22**, 3410–3419 (2011).

16. S. Sun, C.-T. Ting, C.-I. Wu, The normal function of a speciation gene, Odysseus, and its hybrid sterility effect. *Science* **305**, 81–83 (2004).

17. D. Vermaak, H. S. Malik, Multiple roles for heterochromatin protein 1 genes in Drosophila. *Annu. Rev. Genet.* **43**, 467–492 (2009).

18. A. I. Kalmykova, M. S. Klenov, V. A. Gvozdev, Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. *Nucleic Acids Res.* **33**, 2052–2059 (2005).

19. E. Pasyukova, S. Nuzhdin, W. Li, A. J. Flavell, Germ line transposition of the copia retrotransposon in Drosophila melanogaster is restricted to males by tissue-specific control of copia RNA levels. *Mol. Gen. Genet.* **255**, 115–124 (1997).

20. P. Chen, A. A. Kotov, B. K. Bazylev, L. V. Olenina, A. A. Aravin.. piRNA-mediated gene regulation and adaptation to sex-specific transposon expression in *D. melanogaster* male germline. BioRxiv [Preprint]. 2020 Aug. doi:10.1101/2020.08.25.266585

21. J. Vazquez, A. S. Belmont, J. W. Sedat, The dynamics of homologous chromosome pairing during male Drosophila meiosis. *Curr Biol* **12**, 1473–1483 (2002).

22. M. F. ElMaghraby, P. R. Andersen, F. Pühringer, U. Hohmann, K. Meixner, T. Lendl, L. Tirian, J. Brennecke, A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production. *Cell* **178**, 964–979.e20 (2019).

23. H. M. T. Choi, M. Schwarzkopf, M. E. Fornace, A. Acharya, G. Artavanis, J. Stegmaier, A. Cunha, N. A. Pierce, Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145** (2018).

24. M. Jagannathan, R. Cummings, Y. M. Yamashita, A conserved function for pericentromeric satellite DNA. 19 (2018).

25. M. Jagannathan, R. Cummings, Y. M. Yamashita, The modular mechanism of chromocenter formation in Drosophila. *eLife* **8**, e43938 (2019).

26. S. Henikoff, K. Ahmad, J. S. Platero, B. van Steensel, Heterochromatic deposition of centromeric histone H3-like proteins. *Proc Natl Acad Sci U S A* **97**, 716–721 (2000).

27. B. Czech, J. B. Preall, J. McGinn, G. J. Hannon, A Transcriptome-wide RNAi Screen in the Drosophila Ovary Reveals Factors of the Germline piRNA Pathway. *Molecular Cell* **50**, 749–761 (2013).

28. D. Handler, K. Meixner, M. Pizka, K. Lauss, C. Schmied, F. S. Gruber, J. Brennecke, The Genetic Makeup of the Drosophila piRNA Pathway. *Molecular Cell* **50**, 762–777 (2013).

29. G. H. Karpen, A. C. Spradling, Analysis of subtelomeric heterochromatin in the Drosophila minichromosome Dp1187 by single P element insertional mutagenesis. *Genetics* **132**, 737–753 (1992).

30. F. Muerdter, P. M. Guzzardo, J. Gillis, Y. Luo, Y. Yu, C. Chen, R. Fekete, G. J. Hannon, A Genome-wide RNAi Screen Draws a Genetic Framework for Transposon Control and Primary piRNA Biogenesis in Drosophila. *Molecular Cell* **50**, 736–748 (2013).

31. A. A. Aravin, N. M. Naumova, A. V. Tulin, V. V. Vagin, Y. M. Rozovsky, V. A. Gvozdev, Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Current Biology* **11**, 1017–1027 (2001).

32. V. V. Vagin, A. Sigova, C. Li, H. Seitz, V. Gvozdev, P. D. Zamore, A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).

33. C. Li, V. V. Vagin, S. Lee, J. Xu, S. Ma, H. Xi, H. Seitz, M. D. Horwich, M. Syrzycka, B. M. Honda, E. L. W. Kittler, M. L. Zapp, C. Klattenhoff, N. Schulz, W. E. Theurkauf, Z. Weng, P. D. Zamore, Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. *Cell* **137**, 509–521 (2009).

34. A. Nagao, T. Mituyama, H. Huang, D. Chen, M. C. Siomi, H. Siomi, Biogenesis pathways of piRNAs loaded onto AGO3 in the Drosophila testis. *RNA* **16**, 2503–2515 (2010).

35. A. Pane, K. Wehr, T. Schüpbach, zucchini and squash Encode Two Putative Nucleases Required for rasiRNA Production in the Drosophila Germline. *Developmental Cell* **12**, 851–862 (2007).

36. A. Schmidt, G. Palumbo, M. P. Bozzetti, P. Tritto, S. Pimpinelli, U. Schäfer, Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of Drosophila melanogaster. *Genetics* **151**, 749–760 (1999).

37. W. Stapleton, S. Das, B. D. McKee, A role of the Drosophila homeless gene in repression of Stellate in male meiosis. *Chromosoma* **110**, 228–240 (2001).

38. Y. Tomari, T. Du, B. Haley, D. S. Schwarz, R. Bennett, H. A. Cook, B. S. Koppetsch, W. E. Theurkauf, P. D. Zamore, RISC assembly defects in the Drosophila RNAi mutant armitage. *Cell* **116**, 831–841 (2004).

39. Z. Zhang, J. Xu, B. S. Koppetsch, J. Wang, C. Tipping, S. Ma, Z. Weng, W. E. Theurkauf, P. D. Zamore, Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. *Mol. Cell* **44**, 572–584 (2011).

40. J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G. J. Hannon, Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128**, 1089–1103 (2007).

41. R. W. Hardy, D. L. Lindsley, K. J. Livak, B. Lewis, A. L. Siversten, G. L. Joslyn, J. Edwards, S. Bonaccorsi, Cytogenetic analysis of a segment of the Y chromosome of Drosophila melanogaster. *Genetics* **107**, 591–610 (1984).

42. Y. Chen, A. Pane, T. Schüpbach, Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in Drosophila. *Curr. Biol.* **17**, 637–642 (2007).

43. C. Klattenhoff, D. P. Bratu, N. McGinnis-Schultz, B. S. Koppetsch, H. A. Cook, W. E. Theurkauf, Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev. Cell* **12**, 45–55 (2007).

44. K. Yacobi-Sharon, Y. Namdar, E. Arama, Alternative germ cell death pathway in Drosophila involves HtrA2/Omi, lysosomes, and a caspase-9 counterpart. *Dev. Cell* **25**, 29–42 (2013).

45. A. A. Aravin, M. S. Klenov, V. V. Vagin, F. Bantignies, G. Cavalli, V. A. Gvozdev, Dissection of a Natural RNA Silencing Process in the Drosophila melanogaster Germ Line. *MCB* **24**, 6742–6750 (2004).

46. J. Brennecke, C. D. Malone, A. A. Aravin, R. Sachidanandam, A. Stark, G. J. Hannon, An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**, 1387–1392 (2008).

47. A. de Vanssay, A.-L. Bougé, A. Boivin, C. Hermant, L. Teysset, V. Delmarre, C. Antoniewski, S. Ronsseray, Paramutation in Drosophila linked to emergence of a piRNA-producing locus. *Nature* **490**, 112–115 (2012).

48. A. Le Thomas, A. K. Rogers, A. Webster, G. K. Marinov, S. E. Liao, E. M. Perkins, J. K. Hur, A. A. Aravin, K. F. Tóth, Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* **27**, 390–399 (2013).

49. N. V. Rozhkov, M. Hammell, G. J. Hannon, Multiple roles for Piwi in silencing Drosophila transposons. *Genes Dev.* **27**, 400–412 (2013).

50. G. Sienski, D. Dönertas, J. Brennecke, Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**, 964–980 (2012).

51. S. H. Wang, S. C. R. Elgin, Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 21164–21169 (2011).

52. A. Akkouche, B. Mugat, B. Barckmann, C. Varela-Chavez, B. Li, R. Raffel, A. Pélisson, S. Chambeyron, Piwi Is Required during Drosophila Embryogenesis to License Dual-Strand piRNA Clusters for Transposon Repression in Adult Ovaries. *Mol. Cell* **66**, 411–419.e4 (2017).

53. D. N. Cox, A. Chao, H. Lin, piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127**, 503–514 (2000).

54. C. D. Malone, J. Brennecke, M. Dus, A. Stark, W. R. McCombie, R. Sachidanandam, G. J. Hannon, Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **137**, 522–535 (2009).

55. E. L. Beall, P. W. Lewis, M. Bell, M. Rocha, D. L. Jones, M. R. Botchan, Discovery of tMAC: a Drosophila testis-specific meiotic arrest complex paralogous to Myb-Muv B. *Genes Dev.* **21**, 904–919 (2007).

56. M. Hiller, X. Chen, M. J. Pringle, M. Suchorolski, Y. Sancak, S. Viswanathan, B. Bolival, T.-Y. Lin, S. Marino, M. T. Fuller, Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development* **131**, 5297–5308 (2004).

57. J. Batki, J. Schnabl, J. Wang, D. Handler, V. I. Andreev, C. E. Stieger, M. Novatchkova, L. Lampersberger, K. Kauneckaite, W. Xie, K. Mechtler, D. J. Patel, J. Brennecke, The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. *Nat. Struct. Mol. Biol.* **26**, 720–731 (2019).

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

59. M. H. Fabry, F. Ciabrelli, M. Munafò, E. L. Eastwood, E. Kneuss, I. Falciatori, F. A. Falconio, G. J. Hannon, B. Czech, piRNA-guided co-transcriptional silencing coopts nuclear export factors. *Elife* **8** (2019).

60. E. Kneuss, M. Munafò, E. L. Eastwood, U.-S. Deumer, J. B. Preall, G. J. Hannon, B. Czech, Specialization of the Drosophila nuclear export family protein Nxf3 for piRNA precursor export. *Genes Dev.* **33**, 1208–1220 (2019).

61. K. Murano, Y. W. Iwasaki, H. Ishizu, A. Mashiko, A. Shibuya, S. Kondo, S. Adachi, S. Suzuki, K. Saito, T. Natsume, M. C. Siomi, H. Siomi, Nuclear RNA export factor variant initiates piRNA-guided co-transcriptional silencing. *EMBO J.* **38**, e102870 (2019).

62. H. Ohtani, Y. W. Iwasaki, A. Shibuya, H. Siomi, M. C. Siomi, K. Saito, DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the Drosophila ovary. *Genes Dev.* **27**, 1656–1661 (2013).

63. G. Sienski, J. Batki, K.-A. Senti, D. Dönertas, L. Tirian, K. Meixner, J. Brennecke, Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. *Genes Dev.* **29**, 2258–2271 (2015).

64. Y. Yu, J. Gu, Y. Jin, Y. Luo, J. B. Preall, J. Ma, B. Czech, G. J. Hannon, Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science* **350**, 339–342 (2015).

65. K. Zhao, S. Cheng, N. Miao, P. Xu, X. Lu, Y. Zhang, M. Wang, X. Ouyang, X. Yuan, W. Liu, X. Lu, P. Zhou, J. Gu, Y. Zhang, D. Qiu, Z. Jin, C. Su, C. Peng, J.-H. Wang, M.-Q. Dong, Y. Wan, J. Ma, H. Cheng, Y. Huang, Y. Yu, A Pandas complex adapted for

piRNA-guided transcriptional silencing and heterochromatin formation. *Nat. Cell Biol.* **21**, 1261–1272 (2019).

66. E. Quénerch'du, A. Anand, T. Kai, The piRNA pathway is developmentally regulated during spermatogenesis in Drosophila. *RNA* **22**, 1044–1054 (2016).

67. A. A. Kotov, V. E. Adashev, B. K. Godneeva, M. Ninova, A. S. Shatskikh, S. S. Bazylev, A. A. Aravin, L. V. Olenina, piRNA silencing contributes to interspecies hybrid sterility and reproductive isolation in Drosophila melanogaster. *Nucleic Acids Research* **47**, 4255–4271 (2019).

68. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

69. T. E. Oliphant, *Guide to NumPy* (Continuum Press, Austin, Tex., 2015).

70. W. McKinney, "Data Structures for Statistical Computing in Python" (Austin, Texas, 2010; https://conference.scipy.org/proceedings/scipy2010/mckinney.html), pp. 56–61.

71. J. VanderPlas, B. Granger, J. Heer, D. Moritz, K. Wongsuphasawat, A. Satyanarayan, E. Lees, I. Timofeev, B. Welsh, S. Sievert, Altair: Interactive Statistical Visualizations for Python. *JOSS* **3**, 1057 (2018).

72. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

73. J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, J. P. Mesirov, Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).

74. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics* **14**, 178–192 (2013).

Figure 1. Chen et al.

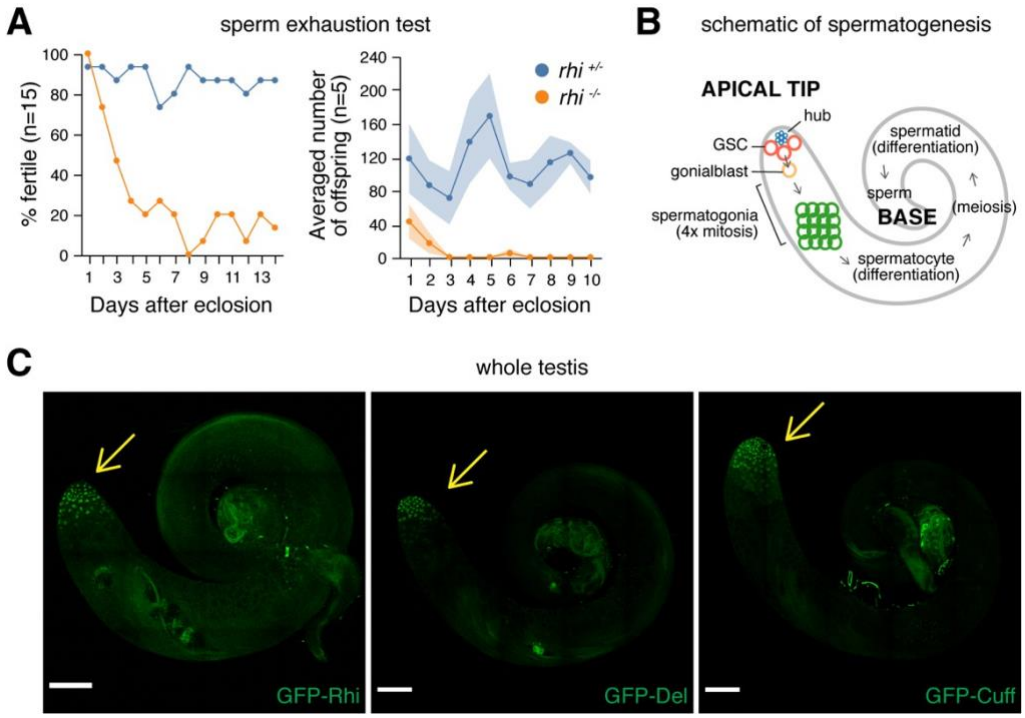


Figure 1. Rhi is required for normal male fertility, and components of RDC complex are expressed at the apical tip of testis.

(A) Compromised fertility *of rhi* mutant males. Sperm exhaustion test of $rhi^{2/KG}$ mutant (orange) and heterozygous sibling control (blue) males. Left: percentages of fertile males 1–14 days post-eclosion (n=15). Right: averaged numbers of offspring per male 1–10 days after eclosion (n=5). Shaded areas display standard error. Two charts report results from two independent sperm exhaustion tests.

(B) A schematic of spermatogenesis, showing major developmental stages of male germline as well as the somatic hub that serves as GSC niche.

(C) Expression of GFP-tagged Rhi (left), Del (middle) and Cuff (right) transgenes driven by their respective regulatory regions. Expression of all three proteins can only be seen at the apical tip of testis (pointed to by the yellow arrow). Scale bar: 100μm.

Figure 2. Chen et al.

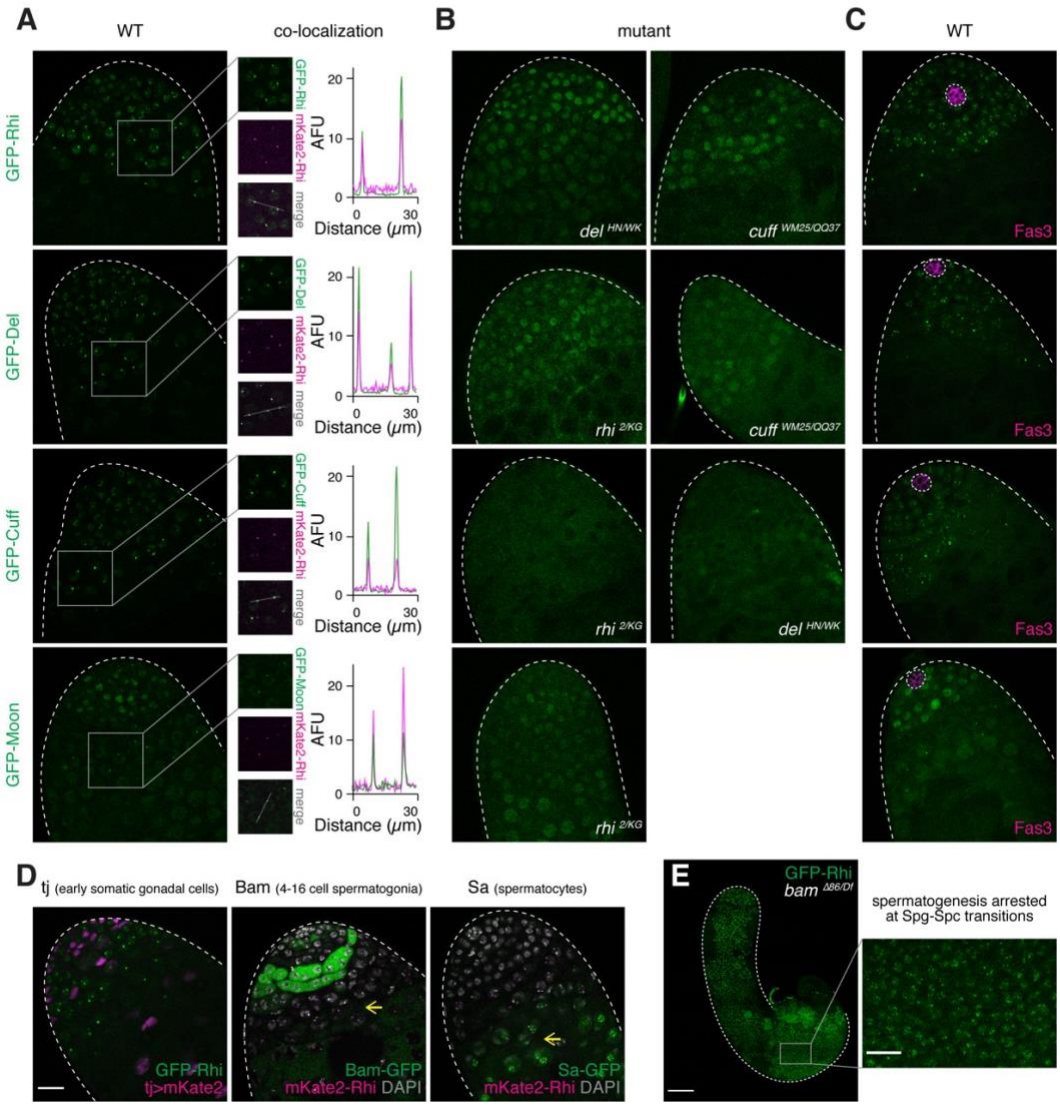


Figure 2. RDC complex is assembled in early male germ cells.

(A) Rhi, Del, Cuff and Moon co-localize in nuclear foci. Confocal images showing apical tips of testes expressing GFP-tagged Rhi, Del, Cuff and Moon (top to bottom) transgenes

driven by their native regulatory regions. Co-localization with mKate2-tagged Rhi in nuclear foci are shown on the right. Signal intensities along the marked line are plotted for each of four co-localization analysis. AFU, arbitrary fluorescence units.

(B) Inter-dependence of Rhi, Cuff, and Del localization in nuclear foci, as well as the dependence of Moon localization on Rhi. Confocal images showing apical tips of testes expressing GFP-Rhi in $del^{HN/WK}$ and $cuff^{WM25/QQ37}$, GFP-Del in $rhi^{2/KG}$ and $cuff^{WM25/QQ37}$, GFP-Cuff in $rhi^{2/KG}$ and $del^{HN/WK}$, and GFP-Moon in $rhi^{2/KG}$ mutant backgrounds. Nuclear foci of each protein dispersed or disappeared in respective mutants.

(C) Rhi, Del, Cuff and Moon are expressed in GSCs. Immuno-fluorescence of testes expressing GFP-tagged Rhi, Del, Cuff and Moon, stained for somatic hub marker Fas3. Note that GSCs directly adjacent to Fas3-positive hub express all four proteins.

(D) Rhi is not expressed in somatic gonadal cells, and its germline expression continues beyond spermatogonia till early spermatocytes. Confocal images showing a fluorescently tagged Rhi transgene with somatic gonadal cells marked by *tj-Gal4>UASp-mKate2* (left), 4–16 cell spermatogonia marked by Bam-GFP (middle) and spermatocytes marked by Sa-GFP (right). Expression of Rhi in early spermatocytes is pointed to by yellow arrows.

(E) Rhi expression upon arrest of spermatogenesis in *bam* mutants. Confocal image showing $bam^{\Delta 86/Df}$ mutant testis expressing GFP-Rhi, where spermatogenesis is arrested at the spermatogonia-to-spermatocyte transition stage. Note that spermatogonia are expanded and virtually all germ cells express Rhi. An enlarged view of the basal part of mutant testis is shown at the bottom, which shares scale bar with (D).

All images share scale bars with (D), except for (E). Scale bars: 20µm (D) and 100µm (E).

Figure 3. Chen et al.

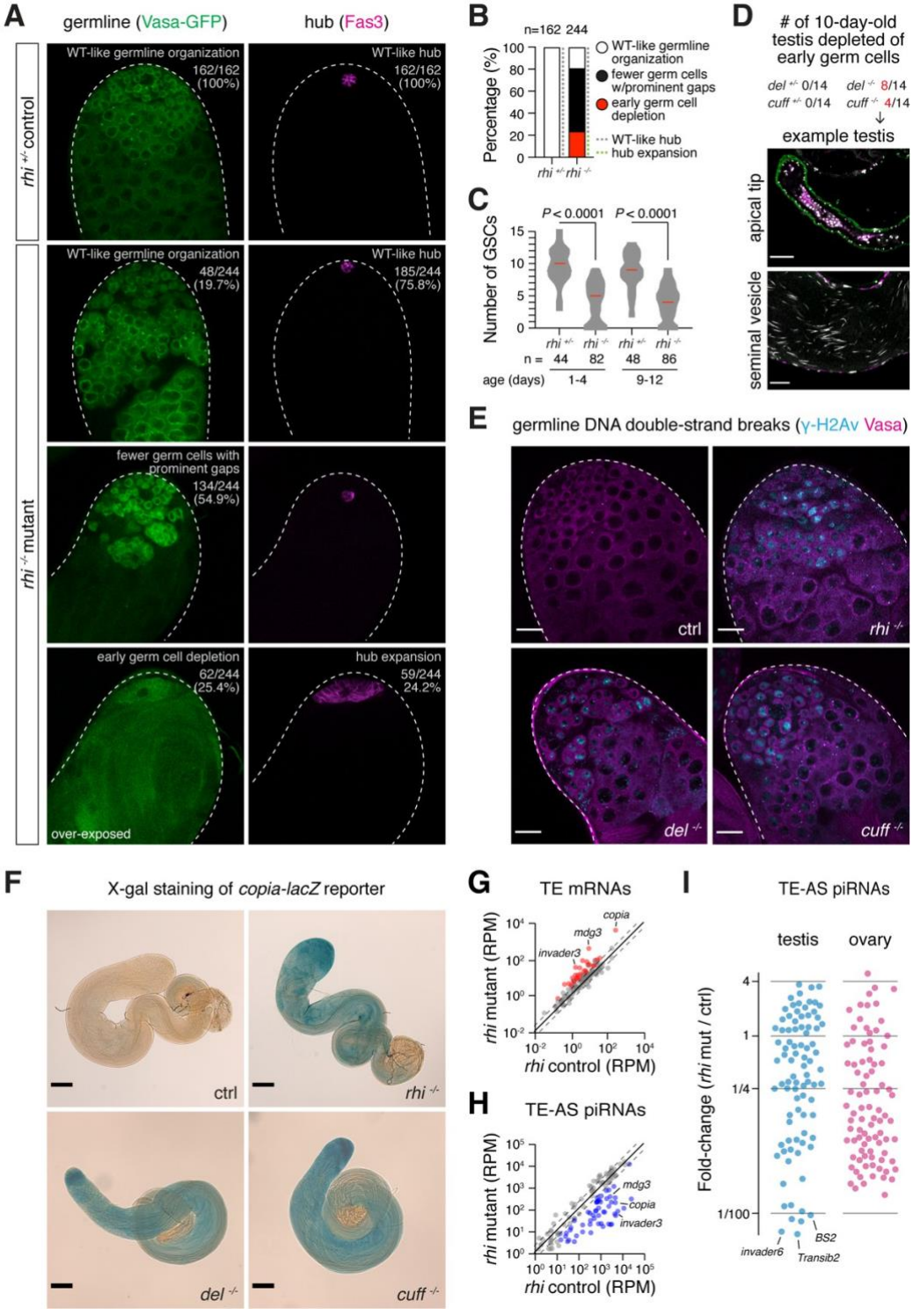


Figure 3. Loss of Rhi causes germ cell death, DNA damage and TE de-repression.

(A) Loss of germ cells in testes of *rhi* mutants. Left: expression of germ cell marker, Vasa-GFP driven by *vasa* promoter, in testes of $rhi^{2/KG}$ mutants and heterozygous control. Right: somatic hub cells that form a niche for GSCs are marked with Fas3. Classification of germ cell phenotype (defined by number and organization of germ cells that express Vasa) and hub size phenotype (defined by cells that express Fas3) are labeled at the top right with corresponding statistics. All images on the same scale as (D).

(B) Quantification of germ cell survival and hub size in *rhi* heterozygous control (left) and mutant (right) testes shown in (A). n, number of testes examined.

(C) Loss of GSCs in testes of *rhi* mutants. Violin plot showing GSC numbers in $rhi^{2/KG}$ mutant and age-matched heterozygous sibling control testes. Median of GSC number is marked red. GSC number is acquired by counting the total number of Vasa-positive cells directly adjacent to Fas3-positive hub in 3D. $P<0.0001$ based on Mann–Whitney–Wilcoxon test. n, number of testes counted for each genotype and age group.

(D) Testes of *del* and *cuff* mutant males showed frequent loss of early germline but produce mature sperm indicating no block of spermatogenesis. Number of 10-day-old virgin males depleted of early germline is listed for each genotype at the top. The apical tip of testis and the seminal vesicle (SV) of such testes are shown below as an example, stained for Vasa, Fas3 and DAPI. Scale bar: 40μm and 20μm for testis and SV, respectively.

(E) Accumulation of DNA DSBs in male germline of *rhi*$^{2/KG}$, *del*$^{HN/WK}$ and *cuff*$^{WM25/QQ37}$ mutants. Immuno-fluorescence of heterozygous control and mutant testes, stained for γ-H2Av, a marker for DNA DSBs, and Vasa, a germline marker. Scale bars: 20µm.

(F) De-repression of *copia* reporter in testes of *rhi* mutants. Brightfield images showing heterozygous control and *rhi*$^{2/KG}$, *del*$^{HN/WK}$ and *cuff*$^{WM25/QQ37}$ mutant testes expressing *copia-lacZ*, after X-gal staining. *copia* LTR containing its promoter is fused upstream to *lacZ* gene. Note that part of *copia* LTR is transcribed as well. Scale bar: 100µm.

(G) De-repression of TEs in testes of *rhi* mutants measured by polyA+ RNA-seq. Scatter plot showing expression of TE mRNAs in *rhi*$^{2/KG}$ mutant versus heterozygous control testes. TEs that show ≥2-fold de-repression (FDR < 0.05) and ≥1 RPM averaged levels are marked red. The mean of two biological replicates is shown.

(H) Loss of TE-targeting piRNAs in testes of *rhi* mutants. Scatter plot showing expression of TE-antisense piRNAs in *rhi*$^{2/KG}$ mutant versus heterozygous control testes. TE-antisense piRNAs that show ≥2-fold reduction (FDR < 0.05) and ≥10 RPM averaged levels are marked blue. Shown are averages of two biological replicates.

(I) Loss of TE-targeting piRNAs in *rhi* males and females. Scatter plot showing fold change of TE-antisense piRNAs upon loss of *rhi* in testis (left, blue) and ovary (right, pink). Note that piRNAs targeting several TE families demonstrate over 100-fold reduction in *rhi* testis, a magnitude not observed in ovary. Averages of two biological replicates are shown.

Figure 4. Chen et al.

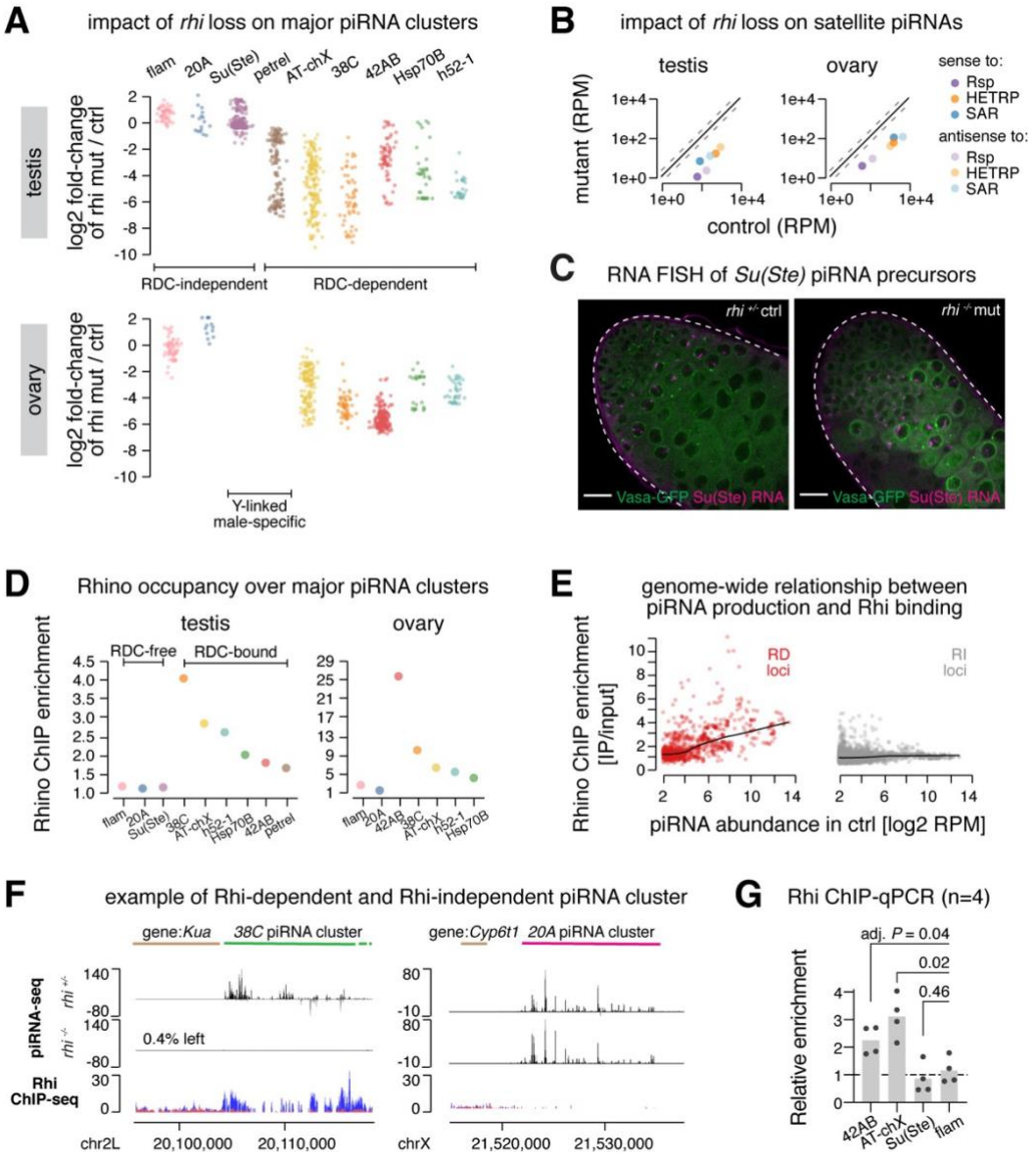


Figure 4. RDC complex is required for piRNA production from dual-strand piRNA clusters.

(A) Impacts of *rhi* loss on piRNA production from major piRNA clusters in testis and ovary. Scatter plot showing fold change of piRNA production from 1Kb genomic windows spanning major piRNA clusters in testis (top) and ovary (bottom), upon loss of *rhi*. Two Y-linked, male-specific clusters are not present in female genome. Each cluster is given a unique color. Averages of two biological replicates are shown.

(B) Impact of *rhi* loss on piRNA production from satellite repeats in testis and ovary. Scatter plot showing levels of complex satellite-mapping piRNAs in $rhi^{2/KG}$ mutant and control testis (left) and ovary (right). Each complex satellite is assigned a color, and piRNAs sense to satellite consensus are marked with higher opacity than antisense ones. Mean of two biological replicates is shown.

(C) Unperturbed expression of *Su(Ste)* piRNA precursors in testes of *rhi* mutants. RNA fluorescence *in situ* hybridization of *Su(Ste)* piRNA precursors in *rhi* control (left) and $rhi^{2/KG}$ mutant (right) testes that express Vasa-GFP. Scale bar: 20μm.

(D) Rhi occupancy over major piRNA clusters in testes and ovary. Scatter plot showing Rhi ChIP-seq enrichment over major piRNA clusters in testis (left) and ovary (right). Shown are averages of two biological replicates. Two Y-linked, male-specific clusters are not present in female genome. Each cluster is colored the same way as in (A).

(E) Scatter plot showing the relationship between Rhi ChIP enrichment and piRNA production over 1Kb genomic windows. Loci that show ≥4-fold decline in piRNA production at the absence of *rhi* are defined as Rhi-dependent (“RD”, red), otherwise Rhi-independent (“RI”, gray). Each dot is the average of two biological replicates. Black lines show local regression.

(F) Examples of Rhi-dependent (*38C*, left) and Rhi-independent (*20A*, right) piRNA clusters. piRNA-seq in *rhi* mutant and control testes are shown at the top, and Rhi ChIP-seq is shown at the bottom (blue: IP, red: input). The profile of a representative replicate is shown.

(G) Rhi does not bind *Su(Ste)* cluster in testis. Bar graphs showing Rhi ChIP-qPCR (n=4) over four piRNA clusters in testis. Adjusted *P*-values are from multiple t-tests corrected for multiple comparisons by the Holm-Sidak method. Uni-strand piRNA cluster *flam* not bound by Rhi serves as a negative control.

Figure 5. Chen et al.

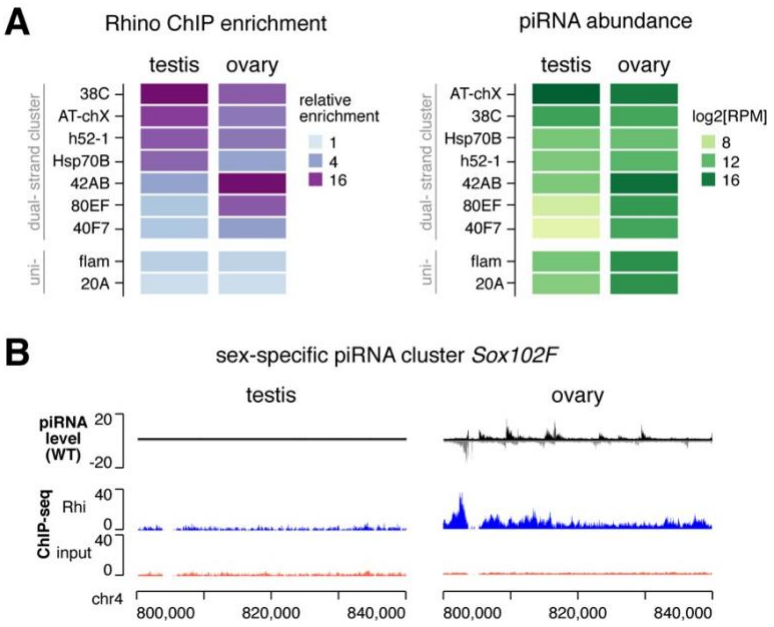


Figure 5. Sexually dimorphic RDC genome occupancy sculpts sex-specific piRNA program.

(A) Heatmaps showing relative enrichment of Rhi binding over major piRNA clusters determined by ChIP-seq (left) and piRNA production from major piRNA clusters (right) in two sexes. The mean of two biological replicates is shown. Rhi-dependent dual-strand clusters are shown at the top, and Rhi-independent uni-strand clusters at the bottom.

(B) *Sox102F* piRNA cluster produces piRNAs exclusively in ovary. Rhi is enriched on *Sox102F* in ovary, but not in testis. Shown is the profile from a representative replicate.

Figure 6. Chen et al.

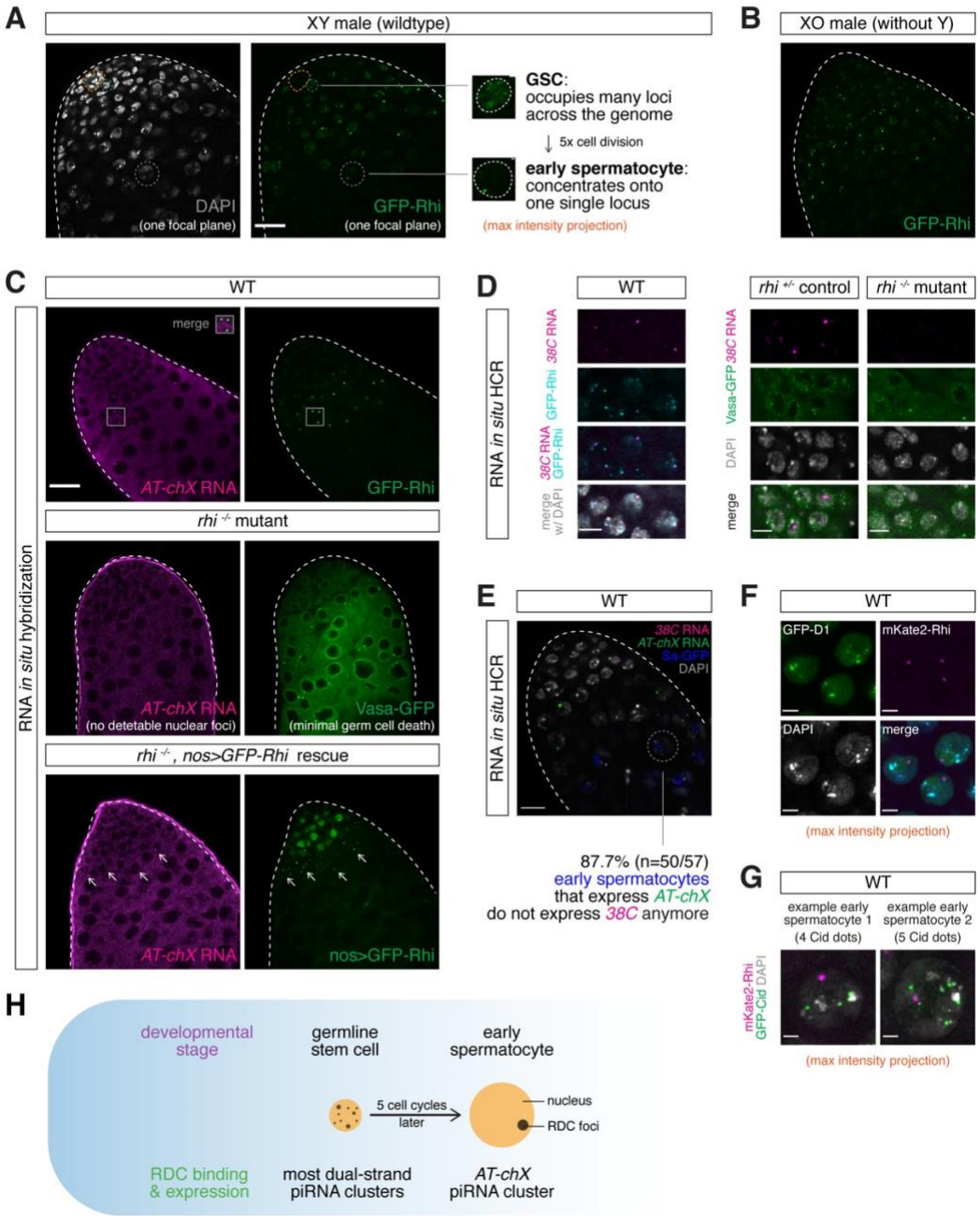


Figure 6. RDC complex enables dynamic piRNA production during early spermatogenesis.

(A) Rhi localizes to multiple nuclear foci in GSCs and early spermatogonia but concentrates in a single dot in early spermatocytes. Confocal images of the apical tip of testis expressing GFP-Rhi transgene. A single focal plane is shown for DAPI and GFP-Rhi, while maximum-intensity projections covering entire nuclei are shown for a GSC and an early spermatocyte on the right. Orange circle outlines the somatic hub for identification of GSCs next to it. Early spermatocyte is identified by formation of chromosome territory and increased nuclear size by DAPI.

(B) Rhi localization to single nuclear dot is not affected in early spermatocytes that lack Y chromosome. Confocal image of the apical tip of testis from XO male (that lacks Y chromosome) expressing GFP-Rhi transgene.

(C) Rhi localizes exclusively to *AT-chX* piRNA cluster in early spermatocytes. Top: RNA FISH of *AT-chX* piRNA precursors in wildtype testis expressing GFP-Rhi. Middle: RNA FISH of *AT-chX* piRNA precursors in $rhi^{2/KG}$ mutant testis. Note that the absence of *AT-chX* RNA foci is not due to the loss of germ cell as Vasa expression shows minimal germ cell death in this particular testis. Bottom: expression of Rhi transgene driven by *nos-Gal4* rescues expression of *AT-chX* cluster in testis of $rhi^{2/KG}$ mutant.

(D) Rhi binds *38C* piRNA cluster in spermatogonia. RNA *in situ* HCR of *38C* piRNA precursors in testis expressing GFP-Rhi (left) or Vasa-GFP in heterozygous control and *rhi* mutant background (right) . Shown are spermatogonia as indicated by DAPI staining. Note that, in contrast to exclusive co-localization with *AT-chX*

cluster in spermatocytes shown in (C), only a subset of Rhi foci co-localize with *38C* RNA foci, indicating that Rhi binds other piRNA clusters besides *38C* in spermatogonia. *38C* RNA signal is present in control but absent in *rhi* mutant testes.

(E) *AT-chX* and *38C* co-express in spermatogonia, but only *AT-chX* is expressed in most early spermatocytes. Dual *in situ* HCR of *38C* and *AT-chX* piRNA precursors in testis expressing Sa-GFP (marker for spermatocytes). Circled is an example of early spermatocytes that express *AT-chX* but not *38C*. Quantification is shown at the bottom. Note that co-expression of *38C* and *AT-chX* can be seen in Sa-negative spermatogonia.

(F) D1 and Rhi do not co-localize in early spermatocytes. Confocal image of early spermatocytes expressing GFP-D1 (a protein trap line) and mKate2-Rhi driven by the *rhi* promoter.

(G) Rhi is localized in a single dot in spermatocyte nuclei, where individual chromosomes occupy distinct chromosomal territories. Confocal image of early spermatocytes expressing mKate2-Rhi and the centromere marker GFP-Cid. The presence of four or more Cid foci indicates that centromeres of individual non-homologous chromosomes are not clustered together.

(H) Proposed model of how Rhi switches genomic binding sites during gametogenesis from GSC to early spermatocyte stage, in order to allow dynamic employment of different piRNA clusters.

Scale bars: 20μm (A-C, E), 5μm (D), 4μm (F) and 2μm (G).

MATERIALS AND METHODS

Fly stocks

The following stocks were used: *bam*$^{\Delta 86}$ (BDSC5427), *bam*^{Df} (BDSC27403), C(1)RM (BDSC9460), *nos-Gal4* (BDSC4937), *UASp-shRhi* (BDSC35171), *iso-1* (BDSC2057), *GFP-Cid* (BDSC25047), *GFP-D1* (BDSC50850) were obtained from Bloomington Drosophila Stock Center; *GFP-Rhi* (VDRC313340), *GFP-Del* (VDRC313271), *GFP-Cuff* (VDRC313269), *moon*$^{\Delta 1}$ (VDRC313735), *moon*$^{\Delta 28}$ (VDRC313738), *Bam-GFP* (VDRC318001), *w*^{1118} (VDRC60000) were obtained from Vienna Drosophila Resource Center; *tj-Gal4* (DGRC104055) was obtained from Kyoto Stock Center; *rhi*^{2}, *rhi*^{KG} and *UASp-GFP-Rhi* were gifts of William Theurkauf; *del*^{WK36}, *del*^{HN56}, *cuff*^{QQ37}and *cuff*^{WM25} were gifts of Trudi Schüpbach; *GFP-Vasa* (gift of Paul Lasko), *copiaLTR-lacZ* (gift of Elena Pasyukova), *GFP-Moon* (gift of Peter Andersen), *Sa-GFP* (gift of Xin Chen). *UASp-mKate2* was described before (*3*). GFP-tagged Rhi, Del, Cuff and Moon are previously described transgenes constructed by inserting N-terminal GFP into genomic BACs via recombineering (*2, 8*). XO males were generated by crossing *GFP-Rhi* males to C(1)RM females. To perform GFP-Rhi ChIP *bam*$^{\Delta 86}$ and *GFP-Rhi* were recombined.

Generation of transgenic flies

To make mKate2-tagged Rhi driven by endogenous *rhi* promoter, the ~2Kb region upstream of *rhi* gene that includes the putative endogenous *rhi* promoter [4] was cloned from genomic DNA of *Drosophila melanogaster* by PCR (forward primer: AGG CCT ATG TAC CAA

GTT GTT AAC TCT ATC G, reverse primer: GGT ACC AGA CAT AAC TTA TCC GCT CAC AGG). PCR product was digested by Stu1 and Kpn1, then ligated into Stu1- and Kpn1-digested vector that contains mKate2-Rhino, mini white gene and the ΦC31 attB site. The construct was inserted into genomic site 76A2 (y1 w1118; PBac{y+-attP-9A}VK00013) on chromosome 3 using ΦC31-mediated recombination.

Sperm exhaustion test

The test was modified from Sun et al. (2004) and done at 25°C. Individual 1-day old virgin males (either $rhi^{2/KG}$ or heterozygous siblings, n=15) were allowed to mate with three 4-day old wildtype virgin females (*iso-1*) for 24hrs. Each male was then moved to mate with another three 4-day old wildtype virgin females (*iso-1*) for another 24hrs, and this was repeated every 24hrs for a total of 14 days. Since each male encountered multiple young virgin females every day, their sperm were exhausted, and the number of progeny produced by females can be used to represent the daily male fertility. Inseminated females were flipped every other day and kept for 20 days (without contact with other males) to achieve maximal egg laying. A binary result of whether there is offspring or not was used to approximate whether a male is fertile or not on a given day, and we plotted the percentage of fertile male out of fifteen tested males each day for $rhi^{2/KG}$ and heterozygous siblings. To probe the male fertility more quantitatively, we repeated the test with five males for ten days. Instead of recording a binary result, we counted the number of progeny. The number of adult offspring was counted 15 days after female fly removal to allow most laid eggs to develop into

adulthood. The averaged total number of offspring for each male each day was plotted for each genotype. This was repeated for $del^{HN/WK}$, $cuff^{WM25/QQ37}$ and respective heterozygous controls (n=5 per genotype), where each 1-day-old virgin male was mated with two 4-day-old w^{1118} virgin females every day for a total of eleven days. After male removal, inseminated females were flipped every 3 days for a total of 15 days for counting.

Female fertility test

Each 1-day-old virgin female with genotype of interest was allowed to mate with two 5-day-old w^{1118} males for four days. Next, flies were discarded but vials were kept for another 14 days before counting, so eggs laid had 14–18 days to develop to adulthood. The number of adult flies from each vial was counted to approximate the female fertility. For each of the four groups shown in Figure S2, all three genotypes (control, mutant and rescue) were siblings from the same cross with similar genetic backgrounds. Results were obtained from three biological replicates (n=3).

Immunofluorescence staining

Testes were dissected in PBS, fixed in 4% formaldehyde for 20mins and washed by PBSTw (PBS with 0.1% Tween-20) for three times. Permeabilization of testes was done by incubation with PBST (PBS with 0.5% Triton-X) for 30mins. Testes were then blocked by 5% BSA in PBSTw for at least an hour, before incubation with primary antibody in 5% BSA in PBSTw at 4°C overnight. Testes were washed 3 times with PBSTw and incubated with secondary

antibody in 5% BSA in PBSTw at room temperature for 2hrs, followed by another 3 washes with PBSTw. Before mounting in VECTA-SHIELD, testes were stained by DAPI (1:5000) for 10mins and rinsed once with PBS. The following primary antibodies were used: mouse anti-Fas3 (7G10, 1:200), mouse anti-γH2Av (UNC93-5.2.1, 1:400) and rat anti-Vasa (concentrated, 1:100) were obtained from Developmental Studies Hybridoma Bank.

RNA fluorescence *in situ* hybridization (RNA FISH)

RNA FISH was done as described previously (*67*). Fixed testes were prepared as above for immunofluorescence staining. Permeabilization was done by incubation in PBST with 0.5% sodium deoxycholate for an hour, followed by three washes of PBSTw. Testes were transferred to first 25% and then 50% formamide, both for 10mins. Next, testes were prehybridized in hybridization buffer (50% formamide, 0.5mg/ml yeast tRNA, 0.2mg/ml) for 1hr at 42°C, before incubation with 0.1–1µg DIG-labeled RNA probe in 50µl hybridization buffer overnight at 42°C with shaking. Testes were rinsed twice with 50% formamide for 20mins at 42°C and then transferred to wash in PBSTw for four times. Subsequent blocking, staining by sheep anti-DIG antibody (PA1-85378, 1:200, Life Technologies) and mounting were the same as described above for immunofluorescence staining. DIG-labeled RNA probes were transcribed by T7 according to manufacturer's instructions. DNA template was made from genomic PCR using primers listed below, with T7 promoter sequence added 5' to the reverse primers.

Su(Ste) (*45*)

F: 5’-CAGGTGATTACCACTATTAACGAAAAGTATGC

R: 5’-ATCCTCGGCCAGCTAGTCCT

AT-chX (*67*)

F: 5’-AGCGATCCCACTGCTAAAGA

R: 5’-ATAAAAGGTGACCG-GCAACG

RNA *in situ* hybridization chain reaction (HCR)

A kit containing a DNA probe set, a DNA probe amplifier and hybridization, amplification and wash buffers were purchased from Molecular Instruments (molecularinstruments.org) for *AT-chX* and *38C* transcripts. To minimize off-targets, we designed probes targeting unique regions at *AT-chX* and *38C.* For *38C*, we specifically targeted junction sites of two different TEs, the simultaneous presence of which is required to generate amplified HCR signals. The *AT-chX* (unique identifier: 3893/E038) and *38C* (unique identifier: 4026/E138-E140) probe sets initiated B1 (Alexa647) and B3 (Alexa 546) amplifiers, respectively. *In situ* HCR v3.0 (*23*) was performed according to manufacturer’s recommendations for generic samples in solution.

X-gal staining

Testes were dissected in PBS, fixed in 0.5% glutaraldehyde containing 1mM $MgCl_2$ for 5mins and washed twice in PBS. Testes were incubated with 0.02% X-gal in X-gal buffer

(1mM $MgCl_2$, 150mM NaCl, 10mM Na_2HPO_4, 10mM NaH_2PO_4, 3.5mM $K_4Fe(CN)_6$ and 3.5mM $K_3Fe(CN)_6$) at 37°C in dark for time of interest. Staining of *copiaLTR-lacZ* in $rhi^{2/KG}$, $del^{HN/WK}$ and $cuff^{WM25/QQ37}$ typically took 1.5–2.5hrs to develop. Reaction was then stopped by two washes of PBS and mounted as above for RNA FISH.

Image acquisition and analysis

Images were acquired using confocal microscope Zeiss LSM 800 with 63x oil immersion objective (NA=1.4) and processed using the software Fiji (*68*). X-gal stained testes were imaged with 10x objective (NA=0.3). Maximum-intensity z-projection was done in Fiji, and line intensity profiles were obtained in Fiji. All images shown were from single focal planes, unless otherwise stated. Dotted outlines were drawn for illustration purposes. To quantify the number of GSCs, we stained the hub by Fas3. Z-stacks were acquired with 0.5µm intervals to cover depths well above and below the entire hub. Vasa-positive germ cells directly adjacent to the hub in 3D were deemed as GSCs and manually counted for each testis. Even though a molecular marker for GSCs was not used, any bias in GSC counting should be shared by both $rhi^{2/KG}$ and heterozygous sibling controls.

RNA-seq and analysis

RNA was extracted from dissected testes of 0–3 days old $rhi^{2/KG}$ and heterozygous sibling controls using TRIzol (Invitrogen). About 1µg RNA for each sample was subject to polyA+ selection using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB E7490) and then

strand-specific library prep using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB E7760) according to manufacturer's instructions. For each genotype, two biological replicates were sequenced on Illumina HiSeq 2500 yielding 25–33 million 50bp single-end reads. Reads mapped to *D. mel* rRNA were discarded by bowtie 1.2.2 allowing 3 mismatches (<2% across all polyA-selected samples). For TE analysis, rRNA-depleted reads were mapped to TE consensus from RepBase17.08 using bowtie 1.2.2 with -v 3 -k 1. Mapped reads were normalized to the total number of reads that can be mapped to dm6 genome. Consistency between biological replicates was confirmed by >0.98 correlation coefficient, so the mean of them was used for all analyses. For simplicity, reads mapped to LTR and internal sequence were merged for each LTR TE given their well-correlated behaviors. Note that polyA-selection was done for TE quantification in order to exclude piRNA precursor transcripts from dual-strand clusters and non-canonical transcripts from individual TEs, which are not polyadenylated but share sequence homology with TEs.

piRNA-seq and analysis

RNA extraction was done as above for RNA-seq. 19–30nt small RNAs were purified by PAGE (15% polyacrylamide gel) from ~1µg total RNA. Purified small RNA was subject to library prep using NEBNext Multiplex Small RNA Sample Prep Set for Illumina (NEB E7330) according to manufacturer's instructions. Adaptor-ligated, reverse-transcribed, PCR-amplified samples were purified again by PAGE (6% polyacrylamide gel). Two biological replicates per genotype were sequenced on Illumina HiSeq 2500 yielding 15–20 million 50bp single-end reads. Adaptors were trimmed with cutadapt 2.5 and size-selected for 23–29nt

sequences for piRNA analysis. 23–29nt reads that mapped to rRNA were discarded by bowtie 1.2.2 tolerating 3 mismatches (<30% in control samples). For TE analysis, 23–29nt small RNA reads were mapped and normalized as done for polyA+ RNA described above, with correlation coefficients between replicates all >0.94. Averages of two well-correlated replicates were used for all analyses. Complex satellite-mapping small RNAs were analyzed similarly (with ovary data downloaded from GSE126578). For piRNA cluster analysis, we used piRNA clusters defined in (*20*). 1Kb genomic windows in individual piRNA clusters were generated with bedtools v2.28.0, and the ones including highly expressed miRNA, snRNA, snoRNA, hpRNA or 7SL SRP RNA were excluded. Coverage over individual piRNA clusters were computed using the pipeline tolerating local repeats described in (*20*). A pseudo-count of 1 was added before calculating log2 fold change of *rhi* mutant over control.

ChIP-qPCR, ChIP-seq and analysis

ChIP protocol was modified based on Le Thomas et al. (2014). For each biological replicate, 200 pairs of 0–2 days old testes or 100 pairs of 4–5 days old ovaries (yeast-fed for 3 days) were fixed in 1% formaldehyde for 10mins, quenched by 25mM glycine for 5mins and washed 3 times with PBS. Fixed testes were snap-frozen in liquid nitrogen and stored in -80°C before ChIP. Frozen testes were first resuspended in PBS and then washed in Farnham Buffer (5mM HEPES pH8.0, 85mM KCl, 0.5% NP-40, protease inhibitor, 10mM NaF, 0.2mM Na_3VO_4) twice. Testes were then homogenized in RIPA buffer (20mM Tris pH7.4, 150mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, protease inhibitor, 10mM

NaF, 0.2mM Na3VO4) using a glass douncer and a tight pestle. Sonication was done in Bioruptor (Diagenode) on high power for 25 cycles (30s on and 30s off). Sonicated tissues were centrifugated to obtain the supernatant. The supernatant was pre-cleared with Dynabeads Protein G beads (Invitrogen) for 2hrs at 4°C. 5% of the pre-cleared sample was set aside as the input, while the rest was incubated with 5µl anti-GFP antibody (A-11122, Invitrogen) overnight at 4°C. The immune-precipitated (IP) sample was incubated with Dynabeads Protein G beads for 5hrs at 4°C to allow beads binding. After that, beads were washed 5 times in LiCl Wash Buffer (10mM Tris pH7.4, 500mM LiCl, 1% NP-40, 1% sodium deoxycholate), while the input sample was incubated with 1µl 10mg/ml RNase A at 37°C for 1hr. Both IP and Input samples were incubated with 100µg proteinase K in PK Buffer (200mM Tris pH7.4, 25mM EDTA, 300mM NaCl, 2% SDS) first at 55°C for 3hrs and then at 65°C overnight. DNA was purified by phenol-chloroform extraction and the concentration was measured by Qubit. Four biological replicates of testis (*bam*$^{\Delta 86/Df}$, *GFP-Rhi*) ChIP were done and used in qPCR, with two randomly selected replicates sequenced. Two biological replicates of ovary ChIP (*GFP-Rhi*) were done and sequenced. ChIP-qPCR was normalized first to input and then to a negative control region (free of Rhi binding in ovary according to Mohn et al. 2014) to obtain Rhi enrichment (primers listed below). ChIP DNA was subject to library prep using NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (NEB E6240) according to manufacturer's instructions. Two biological replicates per sex were sequenced on Illumina HiSeq 2500 yielding 13–22 million 50bp single-end reads. Reads were mapped to the genome as described in (*20*) permitting local repeats. Coverage over piRNA clusters were computed and the enrichment of IP over input was

calculated. Two biological replicates were consistent with a correlation coefficient >0.96, so the average enrichment was plotted.

42AB (*14*)

F: 5’-GTG GAG TTT GGT GCA GAA GC

R: 5’-AGC CGT GCT TTA TGC TTT AC

flam (*14*)

F: 5’-TGA GGA ATG AAT CGC TTT GAA

R: 5’-TGG TGA AAT ACC AAA GTC TTG GGT CAA

Su(Ste) (*31*)

F: 5’-CTTGGACCGAACACTTTGAACCAAGTATT

R: 5’-GGCATGATTCACGCCCGATACAT

AT-chX (*67*)

F: 5’-AGCGATCCCACTGCTAAAGA

R: 5’-GTCGAAGACGTCCAGAGGAG

negative control for Rhi binding (this study)

F: 5’-AAGAGCAGAGGGGCCAAATC

R: 5’-TCCAAGTCGGCTTCCCTTTC

Genome-wide relationship between piRNA production and Rhi binding

This analysis was adapted from Mohn et al. (2014) with modifications. piRNA production and Rhi enrichment were computed for individual 1Kb windows tiling the entire dm6 genome. The average of two well-correlative biological replicates was used for this analysis. Only 1Kb windows having both ≥4 RPM piRNAs in controls and ≥30 RPM reads in IP samples were plotted. Rhi-dependent loci (RD loci) were defined as 1Kb windows showing ≥4-fold drop in piRNA production in *rhi* mutant testes, and the rest were treated as Rhi-independent loci (RI loci). Local regression was implemented with LOESS technique in Python.

Data visualization and statistical analysis

Most data visualization and statistical analysis were done in Python 3 via JupyterLab using the following software packages: numpy (*69*), pandas (*70*) and altair (*71*). Germ cell death, GSC loss and ChIP-qPCR were plotted in GraphPad Prism. Mann–Whitney–Wilcoxon test was done to compute p values for GSC loss. Multiple t-tests corrected for multiple comparisons by the Holm-Sidak method were done for Rhi ChIP-qPCR, using the uni-strand cluster *flam* known to be free of Rhi binding as a negative control. The UCSC Genome Brower (*72*) and IGV (*73*, *74*) were used to conduct explorative analysis of sequencing data.

Data availability

Sequencing data can be accessed via NCBI SRA with the accession number PRJNA646006.

Figure S1. Chen et al.

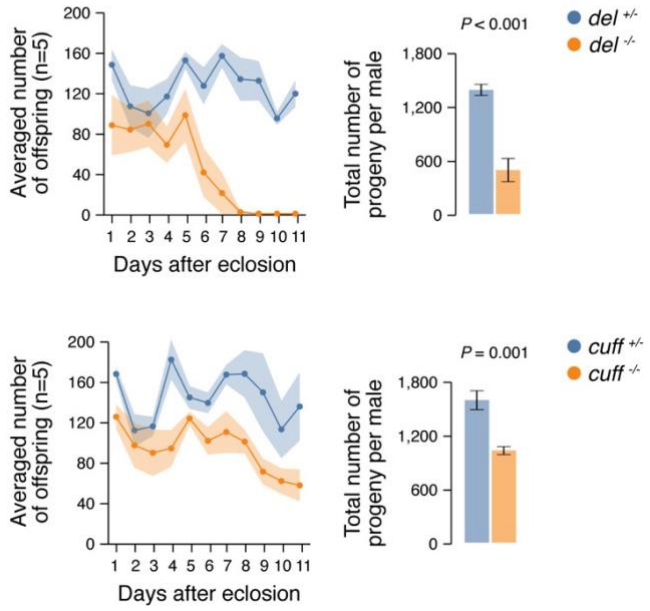


Figure S1. Del and Cuff are required for normal male fertility.

Compromised fertility *of del* (top) and *cuff* (bottom) mutant males. Sperm exhaustion test of $del^{HN/WK}$ or $cuff^{WM25/QQ37}$ mutant (orange) and respective heterozygous sibling control (blue) males. Left: averaged numbers of offspring per male 1–11 days after eclosion (n=5). Right: total number of progeny per male after mating for 11 days. Shaded areas and error bars display standard error. *P*-value from unpaired t-test.

Figure S2. Chen et al.

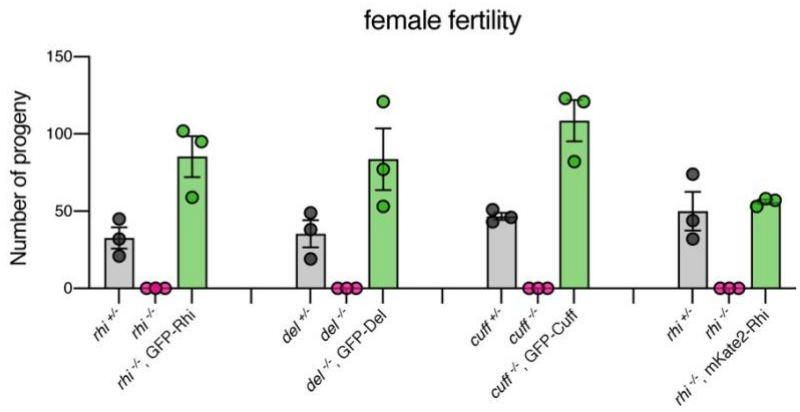


Figure S2. Rescue of female sterility by transgenes used in this study.

GFP-tagged Rhi, Del and Cuff transgenes as well as mKate2-tagged Rhi transgene driven by a putative *rhi* promoter fully rescue the female sterility of respective mutations.

Figure S3. Chen et al.

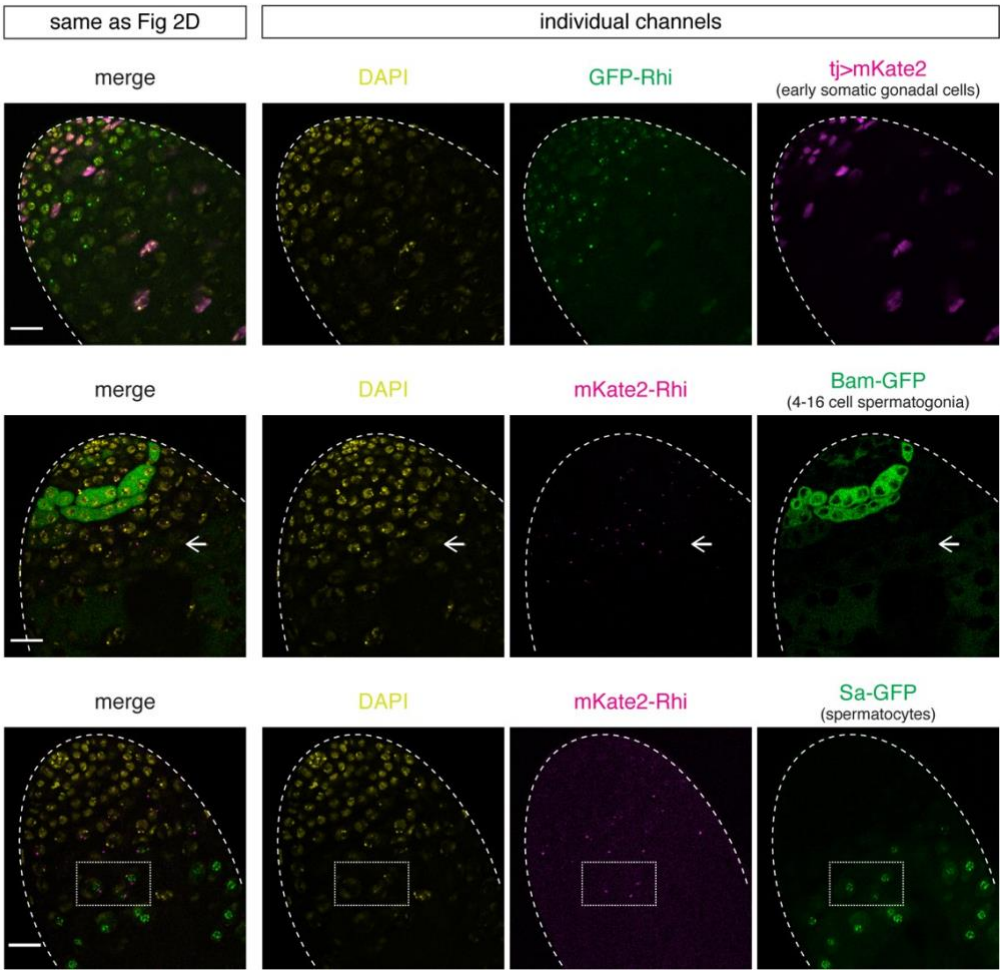


Figure S3. Characterization of the expression of Rhi in testis.

Individual channels of images shown in Figure 2D.

Figure S4. Chen et al.

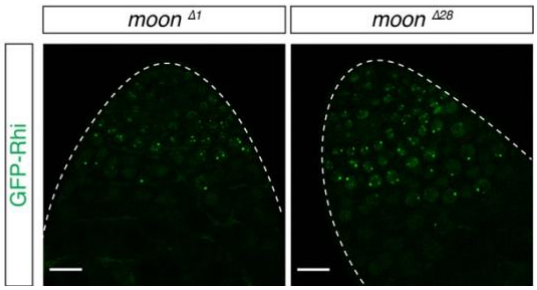


Figure S4. Moon acts downstream of Rhi in testis.

Confocal images of apical tips of testes expressing GFP-Rhi, in *moon*$^{\Delta 1}$ and *moon*$^{\Delta 28}$ mutant backgrounds. Note that *moon* is X-linked, so XY males only have one copy of *moon* and trans-heterozygous mutant cannot be generated.

Figure S5. Chen et al.

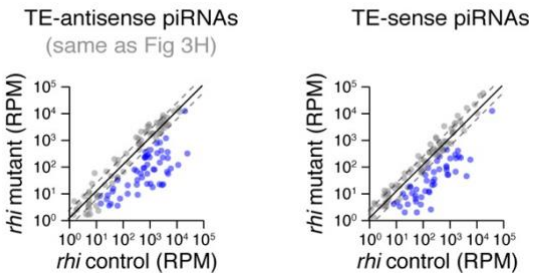


Figure S5. Production of TE-antisense and TE-sense piRNAs collapses without *rhi*.

Loss of TE-mapping piRNAs in testes of *rhi* mutants. Scatter plot showing expression of TE-antisense (left) and TE-sense (right) piRNAs in $rhi^{2/KG}$ mutant versus heterozygous control testes. piRNAs that show ≥2-fold reduction (FDR < 0.05) and ≥10 RPM average expression levels are marked in blue. Shown are averages of two biological replicates.

Chapter 5

GENETIC CONTROL OF A SEX-SPECIFIC PIRNA PROGRAM

P. Chen, A. A. Aravin (2023).

Curr Biol **33**, 1825–1835.e3. https://doi.org/10.1016/j.cub.2023.03.057

Abstract

Sexually dimorphic traits in morphologies are widely studied(*1–4*), but those in essential molecular pathways remain largely unexplored. Previous work showed substantial sex differences in *Drosophila* gonadal piRNAs(*5*), which guide PIWI proteins to silence selfish genetic elements thereby safeguarding fertility(*6–8*). However, the genetic control mechanisms of piRNA sexual dimorphism remain unknown. Here, we showed that most sex differences in the piRNA program originate from the germline rather than gonadal somatic cells. Building on this, we dissected the contribution of sex chromosome and cellular sexual identity towards the sex-specific germline piRNA program. We found that the presence of the Y chromosome is sufficient to recapitulate some aspects of the male piRNA program in a female cellular environment. Meanwhile, sexual identity controls the sexually divergent piRNA production from X-linked and autosomal loci, revealing a crucial input from sex determination into piRNA biogenesis. Sexual identity regulates piRNA biogenesis through Sxl and this effect is mediated in part through chromatin proteins Phf7 and Kipferl. Together, our work delineated the genetic control of a sex-specific piRNA program, where sex chromosome and sexual identity collectively sculpt an essential molecular trait.

Main Text

Results

Sexual dimorphism, where a trait is modified by the biological sex to manifest in distinct ways between males and females, is pervasive in nature. While sexually dimorphic traits in morphologies have been widely studied(*1–4*), those in essential molecular pathways remain largely unexplored. In *Drosophila melanogaster* gonads, the piRNA program executes a critical function by guiding the PIWI-clade Argonaute proteins to silence selfish genetic elements such as transposons(*6–8*), thereby safeguarding fertility. To pass on the transgenerational memory of proper piRNA targets, mothers deposit piRNAs to the embryo, instructing the zygotic genome to mount a homologous piRNA program in the next generation that reflects the maternal response to genomic parasites(*9–12*). However, males implement a piRNA program distinct from their female siblings(*5*), the underlying mechanism of which is elusive. We previously found evidence for both differential transcription of piRNA loci in the nucleus(*13*) and differential processing of piRNA precursor transcripts in the cytoplasm(*5*) between the two sexes, but the upstream control of these sexually dimorphic molecular events is unknown. In this work, we sought to decipher the genetic control of piRNA sexual dimorphism, in order to gain insights into the mechanisms by which sexual dimorphism in essential molecular traits is sculpted.

Prior work compared the male and female piRNA profiles from two different *D. melanogaster* lab strains(*5*), where distinct genetic backgrounds confounded the characterization of piRNA sexual dimorphism. In addition, the sex of *D. melanogaster* is determined independently of the presence of the Y chromosome (both XY and XO flies are

phenotypic males, while XX and XXY flies are phenotypic females)(*14–16*), so the morphology-based identification of males and females does not directly translate to an interpretation of Y chromosome status. Given that several piRNA-producing loci reside on the Y(*5*), the inability to infer Y chromosome content from the phenotypic sex complicates the characterization of piRNA sexual dimorphism. To circumvent these issues, we introduced a Y chromosome marked by y^+ and w^+ genes (hereafter y^+w^+Y) into an inbred *yw* stock and backcrossed it to *yw* for multiple consecutive generations (see STAR Methods). This line allowed us to unequivocally identify XY males (red-eyed flies with black body color and male genitalia) and XX females (white-eyed flies with yellow body color and female genitalia) (Figure 1A), from which we profiled the gonadal piRNAs in each of the two sexes. Analysis of the piRNA libraries showed substantial intersexual differences in the abundance of piRNAs targeting different transposon families (Figure 1B) and expression levels of individual major piRNA loci in the genome (Figure 1C), largely in agreement with our previous study(*5*). Having excluded the possible confounding effects of genetic backgrounds and Y chromosome status, we confirmed that the piRNA program in *D. melanogaster* gonads is sexually dimorphic.

Germline is the major cell type origin of piRNA sexual dimorphism.

In *D. melanogaster* gonads, the piRNA program operates in both the germline and gonadal somatic cells, but piRNA biogenesis and targets differ between the two cell types(*10*, *17*). Thus, the male-female differences seen in gonad-wide piRNA quantification could reflect sexual dimorphism in either germline or gonadal somatic cells, or both cell types, which could be further skewed by distinct germline-soma ratios in testis and ovary. To distinguish

these possibilities, we isolated the somatic piRNAs in the gonad, by immunoprecipitating Piwi upon germline-specific *piwi* knock-down (see STAR Methods) and then sequencing the small RNAs associated with Piwi in gonadal somatic cells (Figure 1D). This allowed us to profile the gonadal somatic piRNA program in each of the two sexes.

Experimentally isolated gonadal somatic piRNAs from testes and ovaries (Figure 1D) allowed us to directly compare the piRNA program in gonadal soma between sexes. We found that the *flamenco* piRNA locus shows a similar piRNA coverage profile and produces piRNAs that take up comparable fractions of total piRNAs in testicular and ovarian soma (Figure 1E). Most of the highly expressed piRNAs in gonadal soma are antisense to transposons in both males and females, and piRNAs targeting different transposon families display a strong positive correlation between the two sexes (Pearson's r = 0.91, $p < 0.0001$; Spearman's r = 0.89, $p < 0.0001$; Figure 1F). When normalized to *flamenco*, a few transposons are targeted more in either males (e.g., *idefix*) or females (e.g., *gtwin* and *mdg1*), but these biases are relatively mild (Figure 1F). To examine piRNA production across the genome, we defined piRNA-producing loci in gonadal soma and measured their expression levels (see STAR Methods). Akin to piRNA quantification based on their transposon targets, quantifying piRNAs based on their genomic origins also revealed a strong positive correlation between the two sexes (Pearson's r = 0.89, $p < 0.0001$; Spearman's r = 0.82, $p < 0.0001$; Figure 1G). We did, however, note an exception: a novel piRNA locus we identified in the gonadal soma, *77B* (Figure 1H), produces more piRNAs in males than females (Figure 1G). This locus resembles *flamenco*, as it makes piRNAs from one genomic strand downstream of a prominent RNA pol II peak that is indicative of a promoter, producing

antisense piRNAs against transposons active in the gonadal soma (e.g., *idefix*; Figure 1H). Nevertheless, the genome-wide view of the piRNA production in gonadal soma highly correlates between sexes. The 3' UTR of some genes (e.g., *tj*) is known to produce piRNAs in ovarian soma(*18*), and the same holds in the male counterpart (Figure 1G). Overall, the piRNA program operating in gonadal soma shows very few sex differences.

Taking advantage of the fact that the *flamenco* piRNA locus is active exclusively in the gonadal soma but not in the germline of both testis (Figure S1) and ovary(*7, 19*), we inferred that germline piRNAs make up about 97% and 79% of the gonadal piRNAs in testis and ovary, respectively (see STAR Methods). Because germline piRNAs dominate the whole gonad piRNA pool to comparable extents in both sexes (97% / 79% = 1.2-fold difference), gonad-wide piRNA quantification is a close approximation of the germline piRNA program when studying male-female differences. Consistent with this, almost all sex-biased piRNAs have been previously annotated as germline-specific(*5, 10, 17, 20*). These results suggest that the piRNA sexual dimorphism we observed in the whole gonads originates from the germline rather than gonadal soma. For the rest of this work, we used the gonad-wide piRNA sexual dimorphism to approximate germline piRNA sexual dimorphism.

Y chromosome is necessary and sufficient to recapitulate aspects of male piRNA program. Having found that germline is the major cell type origin of piRNA sexual dimorphism, we aimed to dissect its underlying genetic control mechanisms (Figure 2A). Distinct sex chromosome contents between sexes, specifically, the presence of Y chromosome in males, could in theory explain some sex differences in piRNAs. On the other hand, distinct sexual

identities could lead to differential piRNA production even from identical piRNA loci located outside the Y. Importantly, the sex determination in *D. melanogaster* does not involve the Y(*14*), which provides us with a unique opportunity to manipulate the Y chromosome without perturbing sexual identities.

To pinpoint the contribution of the Y chromosome to sex differences in the piRNA program, we first generated XY and XO male sibling flies that only differ in the Y chromosome content but are otherwise genetically identical. This is done by using spontaneous sex chromosome nondisjunction that occurs at about 10% frequency in *X^XY* females carrying the compound X chromosome, *C(1)A* (Figure 2C). When compared to their XY brothers, XO males lose piRNAs targeting several transposon families (Figure 2D), suggesting that the Y chromosome is likely a source of transposon-targeting piRNAs in the male. For example, the absence of the Y chromosome causes decreases in piRNAs against *nomad* and *invader2* (Figure 2D)—two transposons that are normally targeted by more piRNAs in males than females (Figure 1B), suggesting that these sex differences could be explained by males having the Y chromosome. We also note that, piRNAs targeting *I-element* appear upregulated in males lacking the Y chromosome (Figure 2D), which warrants future investigation. Removing the Y also led to a specific loss of piRNAs from two Y-linked loci—*Su(Ste)* and *petrel* (Figure 2B)—while leaving the piRNA production from other loci on X and autosomes unperturbed (Figures 2B and 2E). Therefore, Y chromosome contributes to the male piRNA program via production of piRNAs from two loci on the Y, *Su(Ste)* and *petrel*, as well as piRNAs targeting a select group of transposons. While it is possible that Y could upregulate the expression of transposon-targeting piRNAs from other chromosomes,

its specific effect on a few transposons led us to favor the possibility that Y chromosome encodes these transposon-targeting piRNAs.

Complementing the male experiment, we generated XX and XXY female sibling flies that only differ in their Y chromosome contents but are otherwise genetically identical. This is achieved by first obtaining an exceptional XXY female from primary sex chromosome nondisjunction that occurs naturally about 1 in 2,000 wildtype flies(*14*), and then crossing this XXY female with XY males to sire XX and XXY females through secondary sex chromosome nondisjunction (Figure 2F). The extra Y chromosome barely altered the overall transposon-targeting piRNA program in females (Figure 2G). Nonetheless, the presence of the Y chromosome in females triggers piRNA biogenesis from *Su(Ste)* and *petrel* (Figure 2H), two loci that reside on the Y chromosome. Even though these two Y-linked piRNA loci appear to be less active in females than males, they have expression levels comparable to other top piRNA loci, including *42AB*, *38C*, and *80F* (Figure 2H), suggesting that the Y is an active and productive piRNA source in a female cellular environment.

We generated genetically identical male and female siblings that only differ in their Y chromosome contents, however, these crosses necessitated the employment of mothers carrying a Y chromosome (Figures 2C and 2F). Given that maternally deposited piRNAs instruct piRNA biogenesis in the progeny(*9*, *11*, *12*), Y-bearing mothers might create a permissive environment to produce Y piRNAs in the offspring by depositing Y-derived piRNAs to the embryo. Consequently, it is unclear whether the effects of Y chromosome on male and female piRNA production we observed (Figures 2D, 2E, 2G, and 2H) depends on

mothers carrying a Y chromosome. To empirically test the role of Y chromosome in piRNA production without mothers bearing a Y, we devised a strategy to generate half siblings of both sexes that share similar, albeit not identical, genetic backgrounds with and without Y chromosome from XX mothers (Figures 2I and 2L). We observed similar effects of the Y chromosome on piRNA production when mothers do not have a Y: Y chromosome seems to be an important source of transposon-targeting piRNAs in males but not in females (Figures 2J and 2M), and the two Y-linked piRNA loci, *Su(Ste)* and *petrel*, produce piRNAs in both sexes, irrespective of the cellular sexual identity (Figures 2K and 2N). We noticed that the Y exerts a slightly greater effect on the transposon-targeting piRNA program in this latter cross scheme (Figures 2I and 2L) compared to the former one involving Y-bearing mothers (Figures 2C and 2F), which likely results from having different fathers and thus distinct paternally inherited haploid genome. Nevertheless, the results between mothers with and without a Y chromosome are qualitatively very similar. We conclude that the presence of the Y chromosome alters the piRNA profiles independently of its presence in the mothers.

Y chromosome in *D. melanogaster* is known to exhibit imprinting effects(*21–23*), that is, Y can behave differently when inherited from the mother or the father. To test if Y-linked piRNA loci show parent-of-origin effects, we designed crosses that allow females to inherit a Y chromosome from either parent. In the case of paternally inheriting the Y, we also designed crosses either with or without mothers bearing a Y (Figure 3A top). In all cases, we detected nascent transcripts from *Su(Ste)* and *petrel* piRNA loci located on the Y chromosome (Figure 3A middle), indicating that piRNA loci on the Y are transcriptionally active in the female germline when inherited from either parent. We also observed similar

behaviors of the Y-linked piRNA loci in the male counterpart (Figure 3B)—when inherited from either parent, with or without mothers carrying a Y, Y chromosome activates both *Su(Ste)* and *petrel* loci in the male germline. Thus, there are no obvious imprinting effects of the two Y-linked piRNA loci, and the mere presence of the Y can translate to an effect on the germline piRNA program in both sexes. Interestingly, in our cross scheme of passing the Y from mothers to daughters, *Su(Ste)* piRNA precursor transcription is also activated in the follicle cells (a gonadal somatic cell type), an unexpected finding that calls for future studies.

Whereas *Su(Ste)* piRNAs silence the *Stellate* genes that are only active in the male germline, *petrel* piRNAs silence the *pirate* gene that is ubiquitously expressed in all tissues including female germline, which allows us to explore if activating *petrel* piRNA biogenesis in the female germline leads to *pirate* silencing. In wildtype XX female germline, the *pirate* gene is active, and its transcripts can be readily detected by RNA *in situ* HCR (Figure 3A bottom). However, introducing a Y into the female germline from either parent led to a marked silencing of the *pirate* gene (Figure 3A bottom), suggesting that making *petrel* piRNAs in female germline has a direct functional outcome. Meanwhile, the presence of Y chromosome in mothers was neither necessary nor sufficient for the silencing of *pirate* in the female progeny. Thus, *pirate* silencing in the female germline requires the presence of the Y chromosome, regardless of the parental origin of the Y. Similarly, in the male germline, having a Y-bearing mother is neither sufficient nor required for the male germline to tame *Stellate* and *pirate*, and the presence of Y chromosome triggers silencing of both *Stellate* and *pirate* genes regardless of the Y's inheritance path (Figure 3B). Taken together, the

differences in piRNA profiles caused by the presence of Y chromosome directly translates to differential silencing of several targets, without obvious parent-of-origin effects.

Cellular sexual identity provides a key input into piRNA biogenesis.

Though necessary, the presence of the Y chromosome in males is not sufficient to explain sex differences in the piRNA program, as piRNA loci outside the Y are also differentially expressed in two sexes (Figure 1C). What underlies piRNA sexual dimorphism outside the Y? In *D. melanogaster*, a cascade of molecular switches takes place after counting the number of X, culminating in either male or female cellular sexual identity(*14–16*, *24–29*) (Figure 4A). To examine the contribution of sexual identities to germline piRNA sexual dimorphism without confounding impacts of the Y chromosome, we sought to masculinize XX female germline and compare it to XO male germline that lacks a Y (Figure 4B). Unlike the *Drosophila* soma, where sex determination occurs cell-autonomously, *Drosophila* germline receives an additional input from the soma on top of its own chromosomal content to determine the germline sex(*25*, *26*). When the germline sex does not match the somatic sex, germline either dies or becomes tumorous(*24–26*, *30–33*), so a productive germline sex reversal requires perturbing both the germline and somatic sex. Given that there is very little sexual dimorphism in the gonadal somatic piRNAs (Figures 1E, 1F, 1G, and 1H) and that gonad-wide sex-biased piRNAs are predominantly expressed in the germline(*5*, *10*, *17*, *20*), reversing the somatic sex should not significantly confound our study of germline piRNA sexual dimorphism. Hence, our germline sex reversal was done in sex-reversed soma, which allowed us to interrogate the effect of sexual identities on germline piRNA sexual dimorphism. With the caveat that somatic perturbations could still influence gonad-wide

piRNA profiles to some extent, we focused our analysis on piRNAs that are active in the germline.

To explore if and how sexual identities impact germline piRNA profiles, we masculinized XX female germline by germline-specific knock-down of *Sex lethal* (*Sxl*), the major factor that governs the female identity in germline(*24–27*), in the *transformer* (*tra*) mutant background that has a masculinized soma(*31*) (Figure 4C). Strikingly, masculinizing the XX female germline converted its transposon-targeting piRNA program to a state that closely resembles the XO male germline (Figure 4D left and S2). When quantified, the median extent of masculinization for piRNAs targeting different transposon families is 99% (see STAR Methods; Figure 4D right). Similarly, for major piRNA loci outside the Y, their expression levels and piRNA coverage profiles also switched from an XX female state to an XO male state upon masculinization of the XX germline (Figures 4E and 4F). For example, abundant piRNAs are made from both proximal and distal ends of the *42AB* piRNA locus in XX females, but masculinized XX germline only make some piRNAs from the distal end of *42AB* and barely any from the proximal side, reminiscent of the XO male germline (Figure 4F left). In sum, reversing the germline sexual identity is sufficient to switch the germline piRNA program from one sex to the other, suggesting that the cellular sexual identity provides a key input into piRNA biogenesis.

How does the germline interpret its sexual identity to elicit a sex-specific piRNA program? We showed that this sexual-identity effect on piRNA biogenesis is governed by the major switch protein Sxl (Figures 4D, 4E, and 4F), which is active in the female, but not male,

germline. Next, we looked into how Sxl orchestrates a female-specific piRNA program in the germline. Sxl is known to regulate two target genes that exhibit sex-specific expression patterns in the germline(*34*): Tdrd5l(*29*), a cytoplasmic protein that forms granules distinct from the piRNA processing sites, and Phf7(*35*), a chromatin reader protein that binds H3K4me2/3. Both Tdrd5l and Phf7 promote a male identity, and Sxl represses these two factors to bolster a female identity. Since Tdrd5l and Phf7 act genetically redundantly to support a male identity(*29*), we focused on Phf7 for this study and asked whether and, if so, to what extent Phf7 mediates the sexual-identity effect on piRNA biogenesis. Expressing Phf7 in the female germline accompanied by somatic masculinization partially masculinized the XX female germline, leading to a 29% median extent of masculinization of transposon-targeting piRNAs (Figure 4D). For many transposon-targeting piRNAs (e.g., those targeting *copia* and *burdock*), the ectopic expression of Phf7 in XX germline shifted the piRNA profile from a female state towards a male state, but not as completely as losing Sxl did (Figure 4D right and S2). This partial reversal of the piRNA program from one sex to the other by Phf7 activation is also obvious when examining the expression of major piRNA loci in the genome. Each of the major piRNA loci in Phf7-expressing XX female germline resumed an activity somewhere in between the wildtype XX female and XO male, for both male- and female-biased loci (Figures 4E and 4F). For instance, Phf7 dampened the activity of *42AB*, a female-biased piRNA locus, and enhanced the activity of *38C*, a male-biased piRNA locus (Figures 4E and 4F). These observations indicate that Phf7 promotes a male piRNA program, and Sxl supports a female piRNA program in part through repressing Phf7. Thus, Phf7 mediates part of the sexual-identity effect on piRNA biogenesis, acting downstream of Sxl.

Recently, a female-specific piRNA biogenesis factor, the zinc finger protein Kipferl, was described to drive a subset of piRNA production in the female germline(*36*). In particular, piRNA production from *80F*, a sex-specific piRNA locus only active in the female germline (Figure 4E), depends on Kipferl(*36*) (Figure 4G), indicating that Kipferl is directly responsible for some of the germline piRNA sexual dimorphism. As Kipferl appears dedicated to the piRNA pathway and is absent in the male germline, we hypothesized that Kipferl is an effector protein that acts downstream of the sex determination pathway to elaborate a female piRNA program. Indeed, knocking-down *Sxl* in the female germline is sufficient to abrogate the expression of *Kipferl* in XX germline, suggesting that Kipferl expression depends on Sxl (Figure 4H). On the other hand, expressing Phf7 in the XX female germline did not perturb *Kipferl* expression (Figure 4H), suggesting that Phf7 and Kipferl act in parallel, both downstream of Sxl, to promote male and female piRNA programs, respectively (Figure 4I).

Taken together, we elucidated a genetic circuit that connects the sex determination pathway to germline piRNA sexual dimorphism (Figure 4I). XX germline activates Sxl, which positively regulates Kipferl to produce female-specific piRNAs and negatively regulates Phf7 to suppress a male piRNA program. On the contrary, XY germline lacks functional Sxl to activate Kipferl and instead expresses Phf7 to elaborate a male piRNA program. We conclude that male and female sexual identities enable divergent piRNA production programs, sculpting a sexually dimorphic molecular trait alongside the male-specific Y chromosome.

Discussion

In this work, we identified the germline as the source of piRNA sexual dimorphism in fly gonads. Building on this, we genetically separated the actions of sex chromosomes and the sex determination pathway to dissect their distinct roles towards the sex-specific piRNA program. We characterized the contribution of the Y chromosome to the male piRNA program, and we showed that the presence of the Y is sufficient to recapitulate some aspects of the male piRNA program in a female cellular environment. In fact, the effect of the Y is independent of its parental origin and mothers' sex chromosome contents. The ability of Y-linked piRNA loci to act in both male and female cellular environments independently of its inheritance path implies unique regulatory mechanisms(*37*) employed by the Y and distinctive evolutionary forces acting on the Y. What could be the forces that drove the evolution of Y-linked piRNAs? We envisage that Y-linked piRNAs might have evolved to facilitate male-specific transposon control and/or to suppress sex chromosome meiotic drive. Meanwhile, we showed that sexual identity is another major determinant of the piRNA program that regulates piRNA biogenesis outside the Y. Specifically, sexual identity shapes piRNA sexual dimorphism under the control of Sxl, which relays the sexual identity of a cell to piRNA biogenesis through the histone reader protein Phf7 and the zinc finger protein Kipferl. Even though we cannot completely exclude the effect of somatic perturbation in our germline sex reversal, our work unequivocally demonstrated an input from sex determination into piRNA biogenesis. We speculate that the sex determination pathway has been hijacked by transposons to facilitate their sex-biased germline invasion(*5*), so integrating the information of germline sexual identities into piRNA biogenesis provides a means to directly couple the sex-specific piRNA defense program with sex-specific transposon threats.

Together, our work revealed that sex chromosome and sexual identity control different facets of piRNA sexual dimorphism, and it is their collective action that sculpts the sex-specific piRNA program in fly germline. It is very likely that other sexually dimorphic traits are under the control of both sex chromosome and sexual identity, and disentangling the effects of the two promises to offer new insights into how a molecular pathway can be modified by each of the two sexes to execute essential functions.

References

1. A. Kopp, I. Duncan, D. Godt, S. B. Carroll, Genetic control and evolution of sexually dimorphic characters in Drosophila. *Nature* **408**, 553–559 (2000).

2. T. M. Williams, J. E. Selegue, T. Werner, N. Gompel, A. Kopp, S. B. Carroll, The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. *Cell* **134**, 610–623 (2008).

3. R. M. Cox, R. Calsbeek, Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *Am Nat* **173**, 176–187 (2009).

4. C. C. Galouzis, B. Prud'homme, Transvection regulates the sex-biased expression of a fly X-linked gene. *Science* **371**, 396–400 (2021).

5. P. Chen, A. A. Kotov, B. K. Godneeva, S. S. Bazylev, L. V. Olenina, A. A. Aravin, piRNA-mediated gene regulation and adaptation to sex-specific transposon expression in D. melanogaster male germline. *Genes Dev* **35**, 914–935 (2021).

6. V. V. Vagin, A. Sigova, C. Li, H. Seitz, V. Gvozdev, P. D. Zamore, A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).

7. J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G. J. Hannon, Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128**, 1089–1103 (2007).

8. K. Saito, K. M. Nishida, T. Mori, Y. Kawamura, K. Miyoshi, T. Nagami, H. Siomi, M. C. Siomi, Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev.* **20**, 2214–2222 (2006).

9. J. Brennecke, C. D. Malone, A. A. Aravin, R. Sachidanandam, A. Stark, G. J. Hannon, An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**, 1387–1392 (2008).

10. C. D. Malone, J. Brennecke, M. Dus, A. Stark, W. R. McCombie, R. Sachidanandam, G. J. Hannon, Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **137**, 522–535 (2009).

11. A. de Vanssay, A.-L. Bougé, A. Boivin, C. Hermant, L. Teysset, V. Delmarre, C. Antoniewski, S. Ronsseray, Paramutation in Drosophila linked to emergence of a piRNA-producing locus. *Nature* **490**, 112–115 (2012).

12. A. Le Thomas, E. Stuwe, S. Li, J. Du, G. Marinov, N. Rozhkov, Y.-C. A. Chen, Y. Luo, R. Sachidanandam, K. F. Toth, D. Patel, A. A. Aravin, Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev.* **28**, 1667–1680 (2014).

13. P. Chen, Y. Luo, A. A. Aravin, RDC complex executes a dynamic piRNA program during Drosophila spermatogenesis to safeguard male fertility. *PLoS Genet* **17**, e1009591 (2021).

14. C. B. Bridges, Non-Disjunction as Proof of the Chromosome Theory of Heredity. *Genetics* **1**, 1–52 (1916).

15. C. B. Bridges, Triploid Intersexes in Drosophila melanogaster. *Science* **54**, 252–254 (1921).

16. J. W. Erickson, J. J. Quintero, Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in Drosophila. *PLoS Biol* **5**, e332 (2007).

17. C. Li, V. V. Vagin, S. Lee, J. Xu, S. Ma, H. Xi, H. Seitz, M. D. Horwich, M. Syrzycka, B. M. Honda, E. L. W. Kittler, M. L. Zapp, C. Klattenhoff, N. Schulz, W. E. Theurkauf, Z. Weng, P. D. Zamore, Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. *Cell* **137**, 509–521 (2009).

18. N. Robine, N. C. Lau, S. Balla, Z. Jin, K. Okamura, S. Kuramochi-Miyagawa, M. D. Blower, E. C. Lai, A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol* **19**, 2066–2076 (2009).

19. A. Pélisson, S. U. Song, N. Prud'homme, P. A. Smith, A. Bucheton, V. G. Corces, Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. *EMBO J* **13**, 4401–4411 (1994).

20. F. Mohn, G. Sienski, D. Handler, J. Brennecke, The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in Drosophila. *Cell* **157**, 1364–1379 (2014).

21. K. A. Maggert, K. G. Golic, The Y chromosome of Drosophila melanogaster exhibits chromosome-wide imprinting. *Genetics* **162**, 1245–1258 (2002).

22. D. U. Menon, V. H. Meller, Imprinting of the Y chromosome influences dosage compensation in roX1 roX2 Drosophila melanogaster. *Genetics* **183**, 811–820 (2009).

23. B. Lemos, A. T. Branco, P.-P. Jiang, D. L. Hartl, C. D. Meiklejohn, Genome-wide gene expression effects of sex chromosome imprinting in Drosophila. *G3 (Bethesda)* **4**, 1–10 (2014).

24. T. Schüpbach, Normal female germ cell differentiation requires the female X chromosome to autosome ratio and expression of sex-lethal in Drosophila melanogaster. *Genetics* **109**, 529–548 (1985).

25. M. Steinmann-Zwicky, H. Schmid, R. Nöthiger, Cell-autonomous and inductive signals can determine the sex of the germ line of drosophila by regulating the gene Sxl. *Cell* **57**, 157–166 (1989).

26. R. Nöthiger, M. Jonglez, M. Leuthold, P. Meier-Gerschwiler, T. Weber, Sex determination in the germ line of Drosophila depends on genetic signals and inductive somatic factors. *Development* **107**, 505–518 (1989).

27. K. Hashiyama, Y. Hayashi, S. Kobayashi, Drosophila Sex lethal gene initiates female development in germline progenitors. *Science* **333**, 885–888 (2011).

28. L. Shapiro-Kulnane, A. E. Smolko, H. K. Salz, Maintenance of Drosophila germline stem cell sexual identity in oogenesis and tumorigenesis. *Development* **142**, 1073–1082 (2015).

29. S. Primus, C. Pozmanter, K. Baxter, M. Van Doren, Tudor-domain containing protein 5-like promotes male sexual identity in the Drosophila germline and is repressed in females by Sex lethal. *PLoS Genet* **15**, e1007617 (2019).

30. E. H. Brown, R. C. King, Studies on the Expression of the Transformer Gene of Drosophila Melanogaster. *Genetics* **46**, 143–156 (1961).

31. A. H. Sturtevant, A Gene in Drosophila Melanogaster That Transforms Females into Males. *Genetics* **30**, 297–299 (1945).

32. M. Steinmann-Zwicky, Sex determination in Drosophila: sis-b, a major numerator element of the X:A ratio in the soma, does not contribute to the X:A ratio in the germ line. *Development* **117**, 763–767 (1993).

33. A. Casper, M. Van Doren, The control of sexual identity in the Drosophila germline. *Development* **133**, 2783–2791 (2006).

34. L. Grmai, C. Pozmanter, M. Van Doren, The Regulation of Germline Sex Determination in Drosophila by Sex lethal. *Sex Dev*, 1–6 (2022).

35. S. Y. Yang, E. M. Baxter, M. Van Doren, Phf7 controls male sex determination in the Drosophila germline. *Dev Cell* **22**, 1041–1051 (2012).

36. L. Baumgartner, D. Handler, S. W. Platzer, C. Yu, P. Duchek, J. Brennecke, The Drosophila ZAD zinc finger protein Kipferl guides Rhino to piRNA clusters. *Elife* **11**, e80067 (2022).

37. Z. G. Venkei, I. Gainetdinov, M. R. Starostik, C. P. Choi, P. Chen, C. Balsara, T. W. Whitfield, G. W. Bell, S. Feng, S. E. Jacobsen, A. A. Aravin, J. K. Kim, P. D. Zamore, Y. M. Yamashita, "Drosophila Males Use 5′-to-3′ Phased Biogenesis to Make Stellate-silencing piRNAs that Lack Homology to Maternally Deposited piRNA Guides" (preprint, Developmental Biology, 2022); https://doi.org/10.1101/2022.09.12.507655.

38. H. M. T. Choi, M. Schwarzkopf, M. E. Fornace, A. Acharya, G. Artavanis, J. Stegmaier, A. Cunha, N. A. Pierce, Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145** (2018).

39. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein,

K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

40. T. Grentzinger, S. Oberlin, G. Schott, D. Handler, J. Svozil, V. Barragan-Borrero, A. Humbert, S. Duharcourt, J. Brennecke, O. Voinnet, A universal method for the rapid isolation of all known classes of functional silencing small RNAs. *Nucleic Acids Res* **48**, e79 (2020).

41. M. Ninova, Y.-C. A. Chen, B. Godneeva, A. K. Rogers, Y. Luo, K. Fejes Tóth, A. A. Aravin, Su(var)2-10 and the SUMO Pathway Link piRNA-Guided Target Recognition to Chromatin Silencing. *Mol Cell* **77**, 556–570.e6 (2020).

42. T. E. Oliphant, *Guide to NumPy* (Continuum Press, Austin, Tex., 2015).

43. W. McKinney, "Data Structures for Statistical Computing in Python" (Austin, Texas, 2010; https://conference.scipy.org/proceedings/scipy2010/mckinney.html), pp. 56–61.

44. J. VanderPlas, B. Granger, J. Heer, D. Moritz, K. Wongsuphasawat, A. Satyanarayan, E. Lees, I. Timofeev, B. Welsh, S. Sievert, Altair: Interactive Statistical Visualizations for Python. *JOSS* **3**, 1057 (2018).

45. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

46. J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, J. P. Mesirov, Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).

47. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics* **14**, 178–192 (2013).

48. P. R. Andersen, L. Tirian, M. Vunjak, J. Brennecke, A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* **549**, 54–59 (2017).

49. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

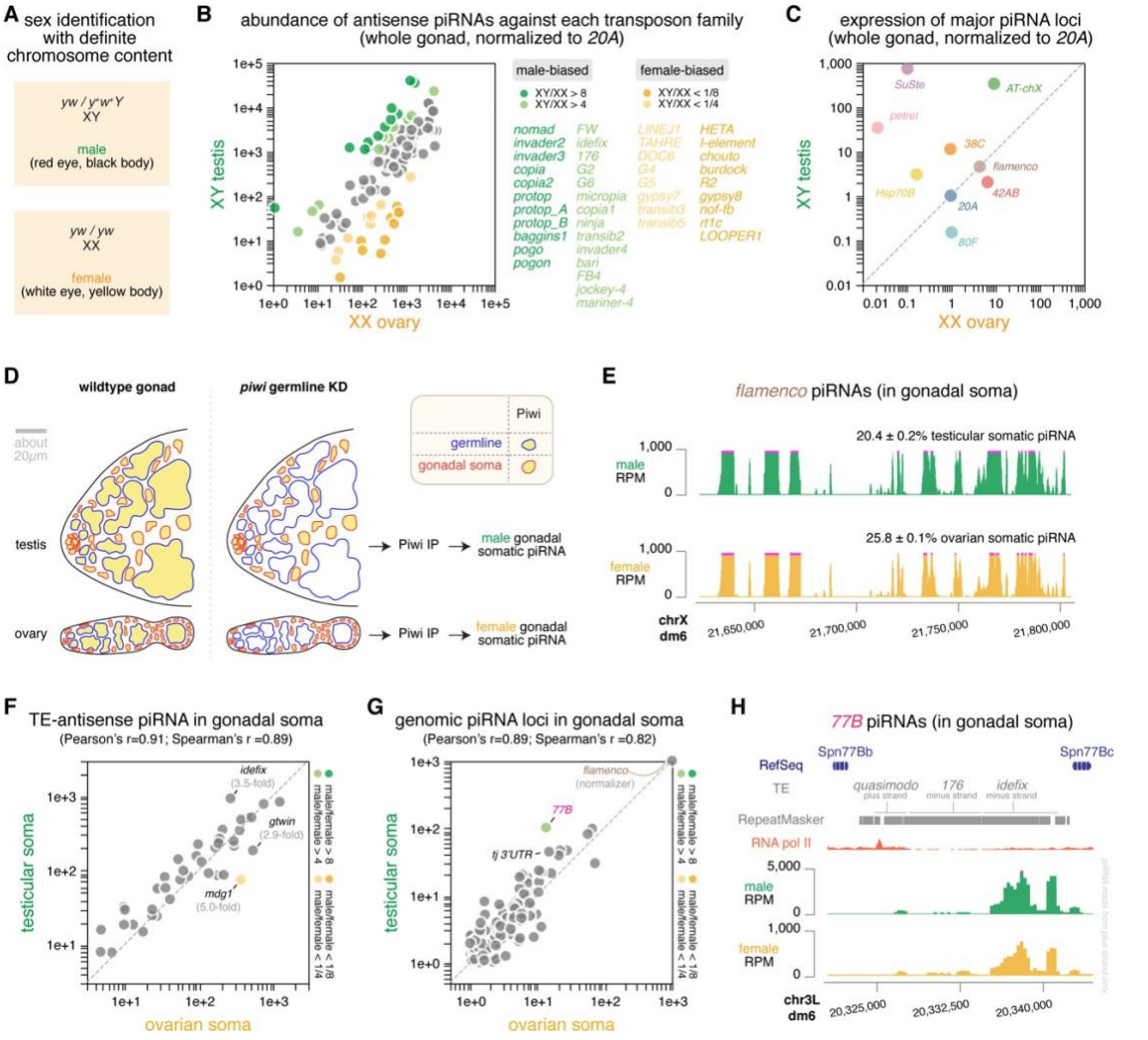


Figure 1 | Germline is the major cell type origin of piRNA sexual dimorphism in *D. mel* gonads.

(A) Genotype and phenotype of males and females that can be identified with definite chromosome content, employing an X chromosome lacking *y* and *w* genes as well as a Y chromosome that carries the wildtype y^+ and w^+ genes.

(B) Comparison of the abundance of piRNAs targeting individual transposon families in XY testis versus XX ovary, normalized to the expression of *20A* piRNAs. Sex-biased transposon-targeting piRNAs are color coded and listed on the right.

(C) Comparison of the expression of major piRNA loci in the genome(*5*) in XY testis versus XX ovary, normalized to the expression of *20A* piRNAs. Each locus is marked by a different color.

(D) Illustration of the experimental strategy to isolate somatic piRNAs in the gonad. Left: cartoon showing the cell type composition of testis and ovary, with germline having a blue outline and gonadal somatic cells having a red outline. Both germline and gonadal somatic cells express Piwi, which is marked by yellow. Right: cartoon showing Piwi expression in testis and ovary upon efficient, germline-specific knock-down of *piwi* that completely depletes Piwi in the germline, leaving the somatic cells as the only source of Piwi in the gonad. Gonadal somatic piRNAs are isolated by immunoprecipitating Piwi from these gonads that lose germline Piwi.

(E) UCSC genome browser view of the *flamenco* piRNA locus showing *flamenco* piRNAs take up similar fractions of gonadal somatic piRNAs in testis and ovary with similar coverage profiles.

(F) Comparison of the abundance of transposon-antisense piRNAs in testicular and ovarian soma, normalized to the expression of *flamenco* piRNAs. Sex-biased piRNAs are color coded in the same way as in (B) and the correlation coefficients are reported.

(G) Comparison of the expression of different piRNA loci in testicular and ovarian soma, normalized to the expression of *flamenco* piRNAs. Sex-biased piRNAs are color coded in the same way as in (B) and the correlation coefficients are reported.

(H) UCSC genome browser view of the *77B* piRNA locus, showing its flanking protein-coding genes, its transposon contents and piRNA coverage profiles in two sexes. Note the difference in y-axis scales that reflects a higher relative activity of *77B* in the testicular soma than the female counterpart. A putative promoter marked by an RNA pol II peak likely drives the expression of piRNAs from the plus strand that are antisense to two transposons, *176* and *idefix*.

See also Figure S1.

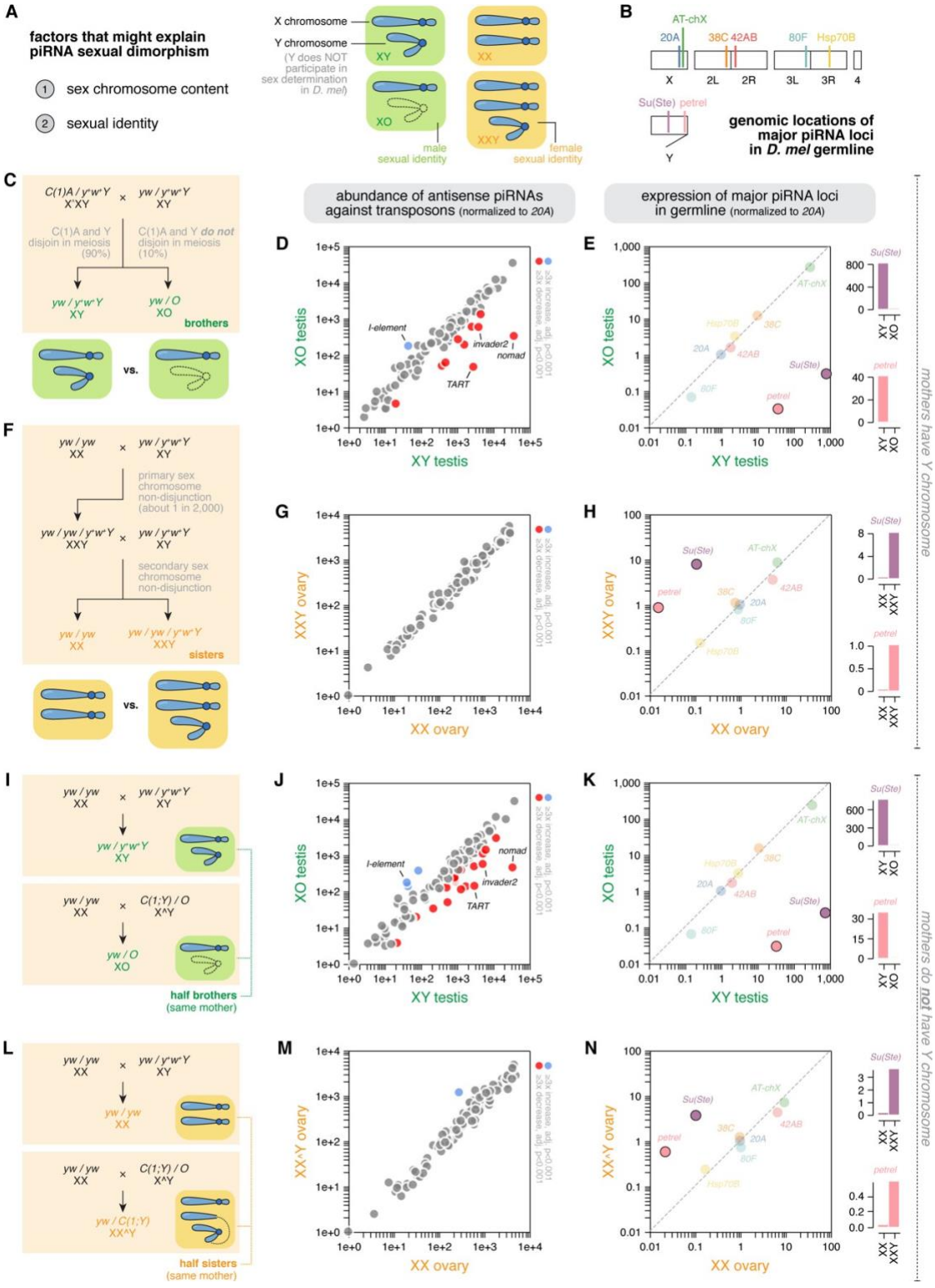

A
factors that might explain piRNA sexual dimorphism
1 sex chromosome content
2 sexual identity
X chromosome
Y chromosome
(Y does NOT participate in sex determination in D. mel)
XY
XX
XO
XXY
male sexual identity
female sexual identity
B
AT-chX
20A
38C 42AB
80F
Hsp70B
X
2L
2R
3L
3R
4
Su(Ste) petrel
Y
genomic locations of major piRNA loci in D. mel germline
C
C(1)A / y+w+ Y
X^XY
yw / y+w+ Y
XY
C(1)A and Y disjoin in meiosis (90%)
C(1)A and Y do not disjoin in meiosis (10%)
yw / y+w+ Y
XY
yw / O
XO
brothers
vs.
F
yw / yw
XX
yw / y+w+ Y
XY
primary sex chromosome non-disjunction (about 1 in 2,000)
yw / yw / y+w+ Y
XXY
yw / y+w+ Y
XY
secondary sex chromosome non-disjunction
yw / yw
XX
yw / yw / y+w+ Y
XXY
sisters
vs.
I
yw / yw
XX
yw / y+w+ Y
XY
yw / y+w+ Y
XY
yw / yw
XX
C(1;Y) / O
X^Y
yw / O
XO
half brothers
(same mother)
L
yw / yw
XX
yw / y+w+ Y
XY
yw / yw
XX
yw / yw
XX
C(1;Y) / O
X^Y
yw / C(1;Y)
XX^Y
half sisters
(same mother)
abundance of antisense piRNAs against transposons (normalized to 20A)
expression of major piRNA loci in germline (normalized to 20A)
D
XO testis
XY testis
I-element
invader2
nomad
TART
≥3x increase, adj. p<0.001
≥3x decrease, adj. p<0.001
E
XO testis
XY testis
AT-chX
Hsp70B
38C
20A
42AB
80F
Su(Ste)
petrel
XX
XO
G
XXY ovary
XX ovary
H
XXY ovary
XX ovary
I
XO testis
XY testis
K
XO testis
XY testis
M
XX^Y ovary
XX ovary
N
XX^Y ovary
XX ovary
mothers have Y chromosome
mothers do not have Y chromosome

Figure 2 | Y chromosome produces piRNAs in both males and females.

(A) Left: listed are factors that might explain piRNA sexual dimorphism. Right: cartoon showing different sex chromosome contents and respective sexual identities. X and Y chromosomes are depicted in different ways, and sexual identities are color coded with the male identity being green and the female one being orange. Note that the Y chromosome of *D. melanogaster* does not participate in sex determination, and the sex instead depends on the number of X. Hence, XY and XO are both males, and XX and XXY are both females.

(B) Illustration showing the karyotype of *D. melanogaster* with five chromosomes—X, Y, 2, 3, and 4, as well as the rough genomic locations of major piRNA loci in the germline.

(C) Cross scheme of the generation of XY and XO brothers.

(D) The abundance of transposon-targeting piRNAs in XO males compared to their XY brothers, showing the loss of piRNAs targeting several transposon families.

(E) The expression of major germline piRNA loci in XO males compared to their XY brothers, showing a specific loss of piRNAs from two Y-linked loci, *Su(Ste)* and *petrel*.

(F) Cross scheme of the generation of XX and XXY sisters.

(G) The abundance of transposon-targeting piRNAs in XXY females compared to their XX sisters, showing very limited differences.

(H) The expression of major germline piRNA loci in XXY females compared to their XX sisters, showing piRNA production from two Y-linked loci, *Su(Ste)* and *petrel*.

(I) Cross scheme of the generation of XY and XO half-brothers, with the same XX mothers.

(J) The abundance of transposon-targeting piRNAs in XO males compared to their XY half-brothers, both of which are sired by XX mothers, showing similar loss of piRNAs targeting several transposon families as seen in (D).

(K) The expression of major germline piRNA loci in XO males compared to their XY half-brothers, both of which are sired by XX mothers, showing a similar loss of piRNAs from two Y-linked loci, *Su(Ste)* and *petrel*, as seen in (E).

(L) Cross scheme of the generation of XX and XXY half-sisters, with the same XX mothers.

(M) The abundance of transposon-targeting piRNAs in XXY females compared to their XX half-sisters, both of which are sired by XX mothers, showing very few differences similar to (G) .

(N) The expression of major germline piRNA loci in XXY females compared to their XX half-sisters, both of which are sired by XX mothers, showing piRNA production from two Y-linked loci, *Su(Ste)* and *petrel*, as seen in (H).

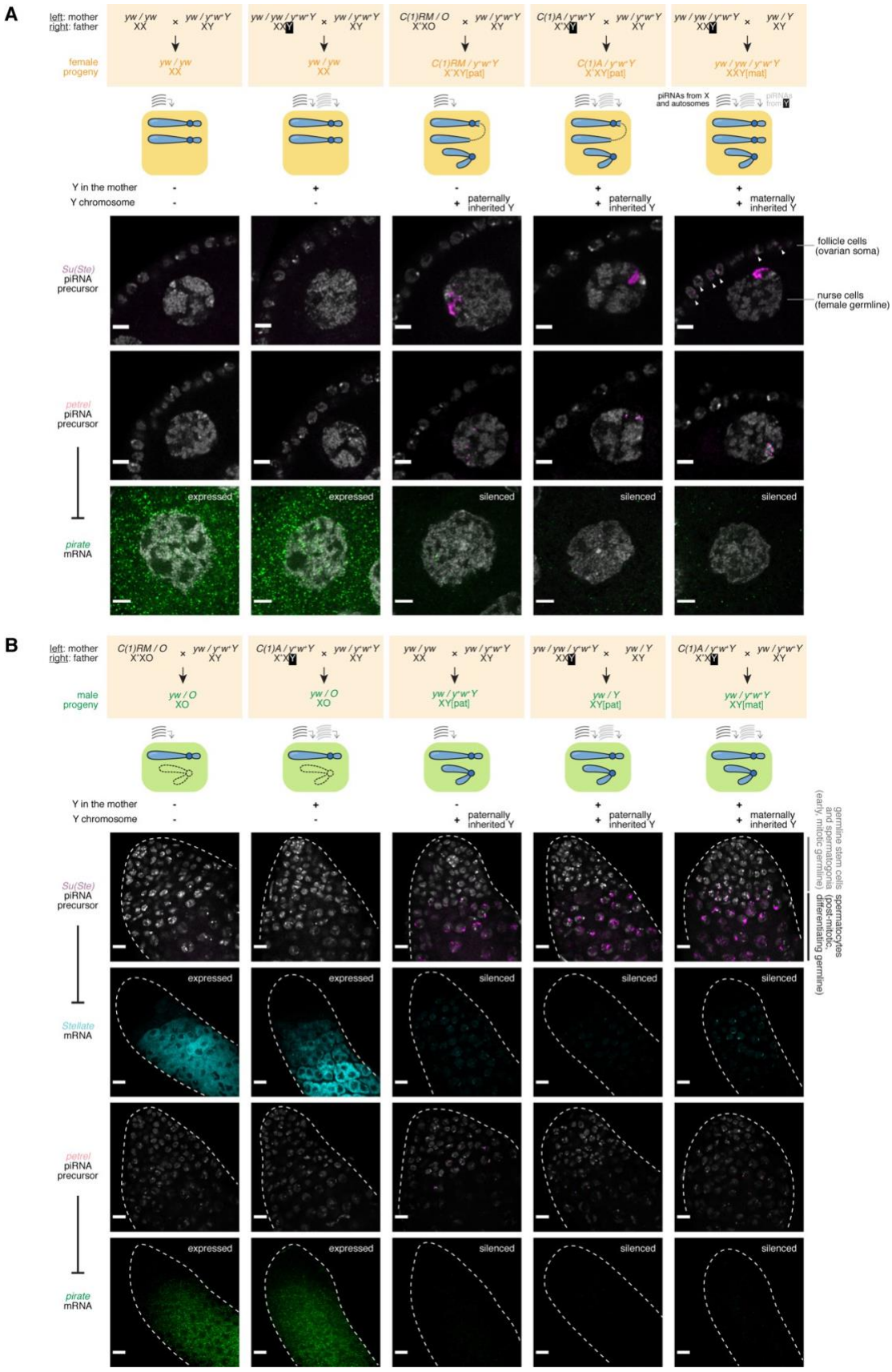

A
left: mother
right: father
female progeny
yw / yw XX × yw / y+w+Y XY → yw / yw XX
yw / yw / y+w+Y XXY × yw / y+w+Y XY → yw / yw XX
C(1)RM / O X+XO × yw / y+w+Y XY → C(1)RM / y+w+Y X+XY[pat]
C(1)A / y+w+Y X+XY × yw / y+w+Y XY → C(1)A / y+w+Y X+XY[pat]
yw / yw / y+w+Y XXY × yw / Y XY → yw / yw / y+w+Y XXY[mat]
piRNAs from X and autosomes
piRNAs from Y
Y in the mother
Y chromosome
paternally inherited Y
maternally inherited Y
Su(Ste) piRNA precursor
follicle cells (ovarian soma)
nurse cells (female germline)
petrel piRNA precursor
pirate mRNA
expressed
silenced
B
left: mother
right: father
male progeny
C(1)RM / O X+XO × yw / y+w+Y XY → yw / O XO
C(1)A / y+w+Y X+XY × yw / y+w+Y XY → yw / O XO
yw / yw XX × yw / y+w+Y XY → yw / y+w+Y XY[pat]
yw / yw / y+w+Y XXY × yw / Y XY → yw / Y XY[pat]
C(1)A / y+w+Y X+XY × yw / y+w+Y XY → yw / y+w+Y XY[mat]
Y in the mother
Y chromosome
paternally inherited Y
maternally inherited Y
Su(Ste) piRNA precursor
Stellate mRNA
petrel piRNA precursor
pirate mRNA
expressed
silenced
germline stem cells and spermatogonia (early, mitotic germline)
spermatocytes (post-mitotic, differentiating germline)

Figure 3 | Y-linked piRNA loci are active and functional in both sexes, when inherited from either parent, regardless of whether mothers carry a Y chromosome.

(A) Top: Cross schemes that generate: XX females without mothers bearing a Y (column 1), XX females with Y-bearing mothers (column 2), XXY females without mothers bearing a Y (column 3), XXY females with Y-bearing mothers but inhering the Y from the father (column 4), and XXY females with Y-bearing mothers and inheriting the Y from the mother (column 5). Middle: cartoon showing the genotype of each kind of females generated and whether their mothers carry a Y is reflected by whether they receive maternally deposited Y-derived piRNAs. Bottom: RNA *in situ* HCR detecting *Su(Ste)* piRNA precursors (row 1), *petrel* piRNA precursors (row 2), and *pirate* mRNAs (row 3) in stage 6–7 egg chambers. Scale bar: 5μm.

(B) Top: Cross schemes that generate: XO males without mothers bearing a Y (column 1), XO males with Y-bearing mothers (column 2), XY males without mothers bearing a Y (column 3), XY males with Y-bearing mothers but inhering the Y from the father (column 4), and XY males with Y-bearing mothers and inheriting the Y from the mother (column 5). Middle: cartoon showing the genotype of each kind of males generated and whether their mothers carry a Y is reflected by whether they receive maternally deposited Y-derived piRNAs. Bottom: RNA *in situ* HCR detecting *Su(Ste)* piRNA precursors (row 1), *Stellate* mRNAs (row 2), *petrel* piRNA precursors (row 3), and *pirate* mRNAs (row 4) and at the apical tips of the testes. Scale bar: 10μm.

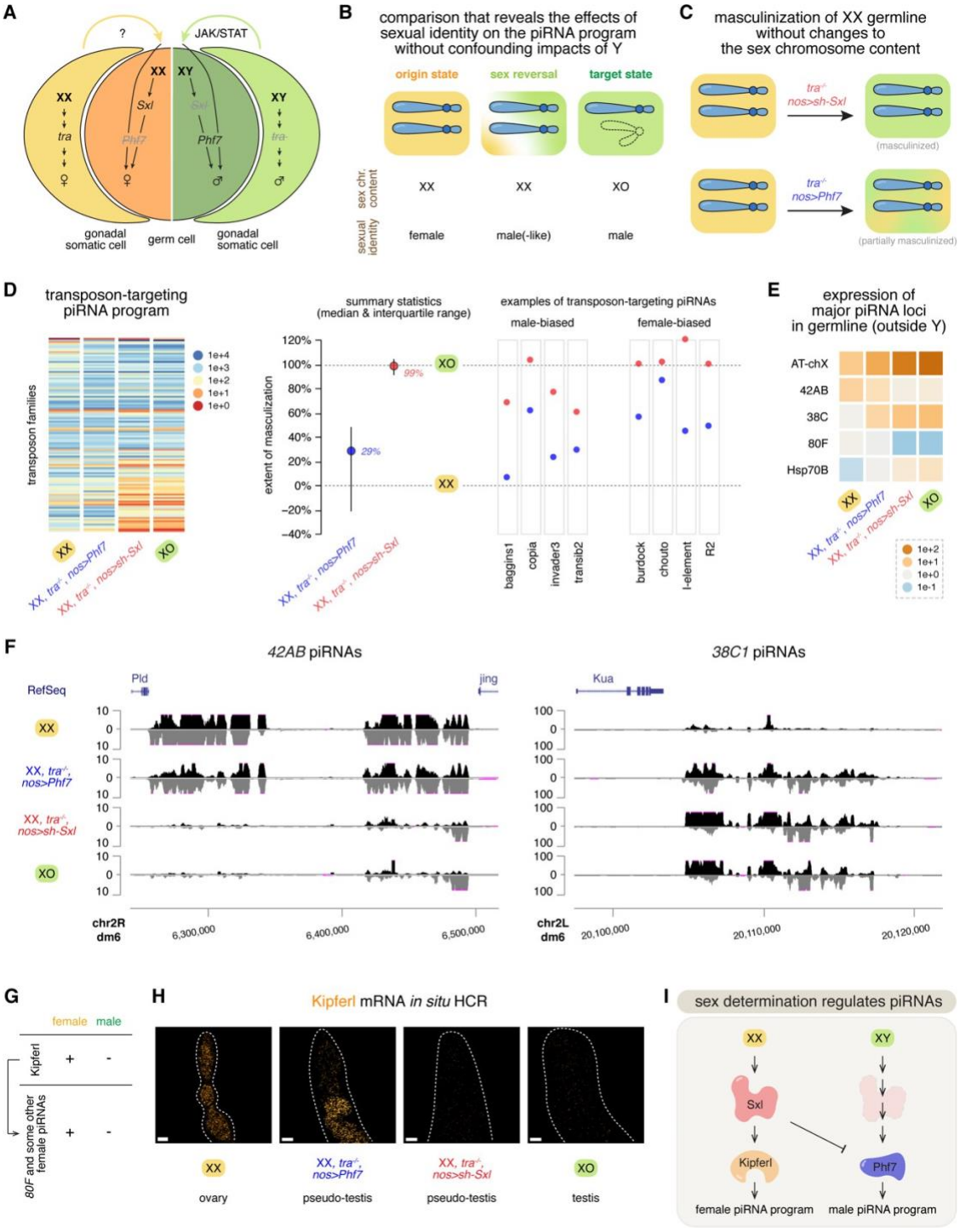

A
?
JAK/STAT
XX
XY
Sxl
tra
Phf7
♀
♂
gonadal somatic cell
germ cell
gonadal somatic cell
B
comparison that reveals the effects of sexual identity on the piRNA program without confounding impacts of Y
origin state
sex reversal
target state
sex chr. content
XX
XX
XO
sexual identity
female
male(-like)
male
C
masculinization of XX germline without changes to the sex chromosome content
tra-/- nos>sh-Sxl
(masculinized)
tra-/- nos>Phf7
(partially masculinized)
D
transposon-targeting piRNA program
transposon families
1e+4
1e+3
1e+2
1e+1
1e+0
XX
XX, tra-/-, nos>Phf7
XX, tra-/-, nos>sh-Sxl
XO
summary statistics (median & interquartile range)
extent of masculization
120%
100%
80%
60%
40%
20%
0%
−20%
−40%
99%
29%
examples of transposon-targeting piRNAs
male-biased
female-biased
baggins1
copia
invader3
transib2
burdock
chouto
I-element
R2
E
expression of major piRNA loci in germline (outside Y)
AT-chX
42AB
38C
80F
Hsp70B
1e+2
1e+1
1e+0
1e-1
F
42AB piRNAs
38C1 piRNAs
RefSeq
Pld
jing
Kua
XX
XX, tra-/-, nos>Phf7
XX, tra-/-, nos>sh-Sxl
XO
chr2R dm6
6,300,000
6,400,000
6,500,000
chr2L dm6
20,100,000
20,110,000
20,120,000
G
female
male
Kipferl
+
-
80F and some other female piRNAs
+
-
H
Kipferl mRNA in situ HCR
XX
ovary
XX, tra-/-, nos>Phf7
pseudo-testis
XX, tra-/-, nos>sh-Sxl
pseudo-testis
XO
testis
I
sex determination regulates piRNAs
XX
XY
Sxl
Kipferl
Phf7
female piRNA program
male piRNA program

Figure 4 | Sexual identity provides a key input into piRNA biogenesis and is a major determinant of the piRNA program.

(A) A simplified model of the sex determination pathway in germline and soma. On top of its own chromosome content, germline receives an additional input from the soma to determine its sex.

(B) A comparison scheme that uses sex reversal to examine the effects of sexual identities on the piRNA program. XX female germline is masculinized and compared to both wildtype XX females (the origin state) and XO males (the target state). Thus, any differences observed would reveal the effects of sexual identities on piRNAs, without confounding impacts of Y chromosome.

(C) Cartoon showing the masculinization of the XX female germline by genetic perturbations, without any changes to the sex chromosome content. To facilitate germline masculinization, the soma is masculinized by mutating *tra*. In addition, germline-specific knock-down of *Sxl* near completely masculinizes the female germline, while ectopic expression of *Phf7* in the germline led to partial masculinization.

(D) Left: a heatmap showing the abundance of piRNAs targeting different transposons in XX female, XO male, or XX masculinized by perturbing either *Phf7* or *Sxl* expression (in the *tra* mutant background). Each row represents piRNAs that target a different transposon, and their expression levels are color coded. Expression was normalized to the *20A* piRNAs. Middle: Quantification of the extent of masculinization for piRNAs targeting individual transposon families. For each transposon family, the abundance of corresponding antisense piRNAs in XX female is scaled to 0% and that in XO male is scaled to 100%, so the levels of transposon-targeting piRNAs in masculinized XX can be normalized to reflect the extent

of masculinization. Shown are the summary statistics (median and interquartile range) of the antisense piRNAs targeting different transposon families. Right: piRNA abundance upon XX masculinization for four examples of male-biased and female-biased transposon-targeting piRNAs, respectively. In these examples as well as the overall summary statistics, perturbing *Sxl* led to a stronger masculinization of the piRNA program than perturbing *Phf7*.

(E) A heatmap showing the expression of major germline piRNA loci (located outside Y) in XX female, XO male, or XX masculinized by perturbing either *Phf7* or *Sxl* expression (in the *tra* mutant background), normalized to the expression of *20A* piRNAs.

(F) UCSC genome browser view of the piRNA coverage profiles over the locus *42AB* (left) and the locus *38C1* (right) in XX female, XO male, or XX masculinized by perturbing either *Phf7* or *Sxl* expression (in the *tra* mutant background).

(G) Female-specific expression of Kipferl and Kipferl-dependent piRNAs.

(H) RNA *in situ* HCR detecting *Kipferl* mRNA in XX female, XO male, or XX masculinized by perturbing either *Phf7* or *Sxl* expression (in the *tra* mutant background).

(I) A genetic circuit that connects the sex determination pathway and piRNA biogenesis.

See also Figure S2 and S3.

Materials and Methods

Fly crosses

To minimize genetic background differences, *yw / y^+w^+*Y was backcrossed to the inbred *yw* line for six consecutive generations, via a single male at every generation. Similarly, after generating *C(1)A / y^+w^+*Y females, we backcrossed them to *yw / y^+w^+*Y males for six consecutive generations, via 2–3 females at every generation. To obtain an XXY exceptional female, we looked for a female carrying the marked Y chromosome (y^+w^+Y) in the *yw / y^+w^+*Y stock, which typically took no more than two months. To deplete germline Piwi, we expressed *sh-piwi* using both *nos-Gal4* and *bam-Gal4*, which led to efficient knock-down of Piwi in the germline as evidenced by the loss of germline GFP-Piwi expression in both testis and ovary. For sex reversal experiments, a Y chromosome marked by B^S (present in the *Df(tra)* stock) that alters the eye shape was employed, such that the sex chromosome content could be inferred independently of the morphological sex. Upon sex reversal, germline remains prevalent but shows phenotypes characteristic of either male or female germline development in Sxl and Phf7 perturbations, respectively (Figure S3).

RNA *in situ* hybridization and RNA *in situ* hybridization chain reaction (HCR)

For RNA *in situ* HCR, probes, amplifiers and buffers were purchased from Molecular Instruments (molecularinstruments.org) for *flam* (3893/E046), *petrel* (3872/E024), *pirate* (3916/E064), *Stellate* (4537/E832) and *Kipferl* (4708/E1062) transcripts. RNA *in situ* HCR v3.0(*38*) was done according to manufacturer's recommendations for generic samples in solution. To detect *Su(Ste)* transcripts, we did conventional RNA *in situ* hybridization using

DNA probes (75bp, position 994–1068 of *Su(Ste): CR42424*, sense direction) directly conjugated with fluorophore purchased from IDT.

Image acquisition and processing

Confocal images were acquired with Zeiss LSM 800 or LSM 980 using a 63x oil immersion objective (NA=1.4) and processed using Fiji(*39*). Single focal planes were shown in all images, where dotted outlines were manually drawn for illustration purposes.

Small RNA sequencing

Argonaute-associated small RNAs were isolated from ovaries (20 pairs per sample) or testes (30 pairs per sample) using TraPR columns(*40*). Purified small RNA was subject to library prep using NEBNext Multiplex Small RNA Sample Prep Set for Illumina (NEB E7330). Adaptor-ligated, reverse-transcribed, PCR-amplified samples were purified again by PAGE (6% polyacrylamide gel), from which we cut out the band within the desired size range. This additional size selection by PAGE eliminated other, longer RNAs (>30 nt) captured by TraPR columns. To isolate Piwi-associated small RNAs in gonadal soma, we first immunoprecipitated GFP-Piwi from gonads lacking germline Piwi (see above for fly crosses) using GFP-Trap (ChromoTek) magnetic agarose beads, as described before(*41*). Small RNAs associated with gonadal somatic Piwi are then purified by TraPR columns and library-prepared, as described above for all Argonaute-associated small RNAs. Libraries were sequenced on Illumina HiSeq 2500 or NextSeq 2000.

Analysis of transposon-targeting piRNAs

To computationally extract piRNAs from all Argonaute-associated small RNAs, adaptor-trimmed small RNAs were size-selected for 23–29nt (cutadapt 2.5) and those mapped to rRNA, tRNA, snRNA, snoRNA, miRNA, hpRNA and 7SL RNA were discarded (bowtie 1.2.2 with -v 3). piRNAs were then mapped to RepBase25.08 (manually curated) and those antisense to transposon consensus sequences with ≤3 mismatches are designated as transposon-targeting piRNAs. For LTR transposons, reads mapping to the LTR and internal sequences of a given transposon family were merged for quantification, given their well-correlated behaviors. All quantification was normalized to the expression of *20A* piRNAs (unless otherwise noted) and shown as the mean of two biological replicates.

Analysis of the expression of major piRNA loci

piRNAs were computationally extracted as described above. piRNAs were mapped to the dm6 genome using a previously described algorithm(*5*) that considers both piRNAs that map to single unique positions in the genome as well as those that map to "local repeats" (defined as repeats that are contained within a genomic window <2Mb in length in dm6 reference genome). Major piRNA loci, their coordinates and quantification method were described before(*5*). All quantification was normalized to *20A* piRNAs. The average of two biological replicates was shown in all figures.

Definition of piRNA-producing loci in gonadal soma

piRNA-producing loci in gonadal soma were defined as previously described for piRNA-producing loci (also known as "piRNA clusters") in whole gonads(*5*). Briefly, piRNAs isolated from gonadal soma were mapped to the genome and those that map uniquely or to

local repeats were kept and quantified over 1Kb windows that tile the entire genome. Neighboring 1Kb widows within 3Kb were merged. If merged windows were ≥5Kb, they were merged again within 15Kb, and this process was repeated twice. This *de novo* method recapitulated the *flamenco* locus and the 3'UTR of the *tj* gene—two loci that are known to make abundant piRNAs in ovarian soma—confirming its utility.

Inference of germline contribution to whole gonad piRNAs

Given that *flamenco* is only active in the gonadal soma but not in the germline, *flamenco* piRNAs found in whole gonad piRNAs must come from somatic cells in the gonad. Experimentally isolated gonadal somatic piRNAs revealed the contribution of *flamenco* piRNAs to total piRNAs in the gonadal soma (e.g., 25%), so if *flamenco* takes up 5% of whole gonad piRNAs, gonadal soma will contribute to 20% (5% / 25% = 20%) of whole gonad piRNAs. Then, the germline contribution to whole gonad piRNAs is 100% -20% = 80%. When calculating actual contributions of *flamenco* piRNAs to gonadal soma and whole gonads of both sexes, the mean of two replicates was used.

Data visualization

All data visualization were done in Python 3 via JupyterLab with the following software packages: numpy(*42*), pandas(*43*) and altair(*44*). The UCSC Genome Brower(*45*) and IGV(*46*, *47*) were used to explore sequencing data and to prepare genome browser track panels shown. UCSC genome browser tracks of piRNA coverages shown in the figures were normalized to CPM (counts per million 23–29nt reads that map uniquely to dm6 genome).

Two biological replicates showed similar coverage profiles on the genome browser, so one

of the two replicates was randomly selected to be shown in the figures. RNA pol II ChIP-seq data from wildtype fly ovary (w1118) was visualized using IGV using the publicly available bigwig file from GEO under the accession number GSM2576144(*48*).

Quantification and statistical analysis

Statistical analysis was done in Python 3 via JuputerLab. All sequencing experiments were done with two biological replicates. Differential expression analysis was done using DESeq2(*49*).

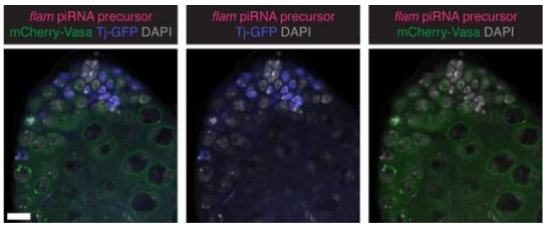


Figure S1 | The flamenco piRNA locus is only active in the gonadal soma in testis. RNA in situ HCR detecting flamenco piRNA precursors in testes expressing mCherry-Vasa and Tj-GFP under endogenous regulatory elements, stained for DAPI. flamenco transcripts are only detected in gonadal somatic cells (including both early cyst cells marked by Tj expression and hub cells marked by the lack of Tj and Vasa expression), but not in germline cells (marked by Vasa expression). Scale bar: 10µm.

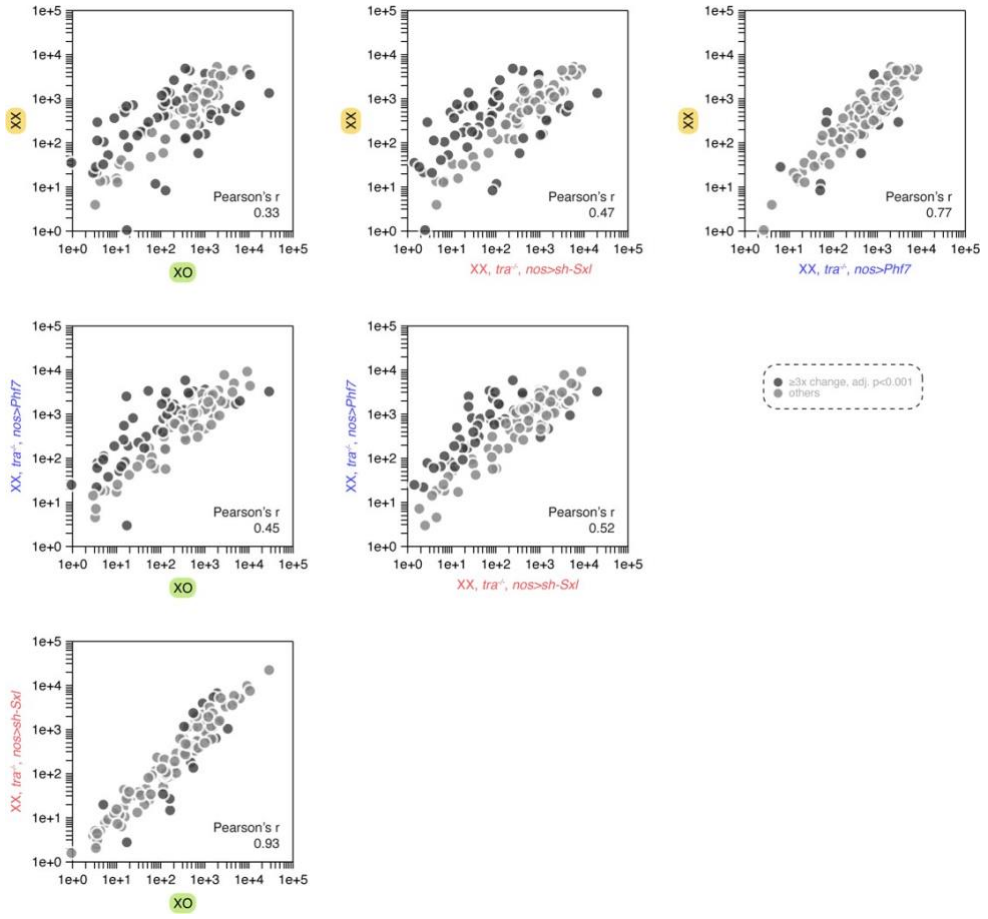


Figure S2 | Pairwise piRNA comparisons of XX female, XO male, and XX masculinized by perturbing either Phf7 or Sxl (in the tra mutant background).

Pairwise comparison of the abundance of piRNAs against different transposon families in four genotypes: XX female, XO male, XX tra-/- with Phf7 ectopic expression, and XX tra-/- with Sxl germline knock-down. Sxl perturbation led to a piRNA program that differed from XX female and resembled XO male, more than Phf7 perturbation did.

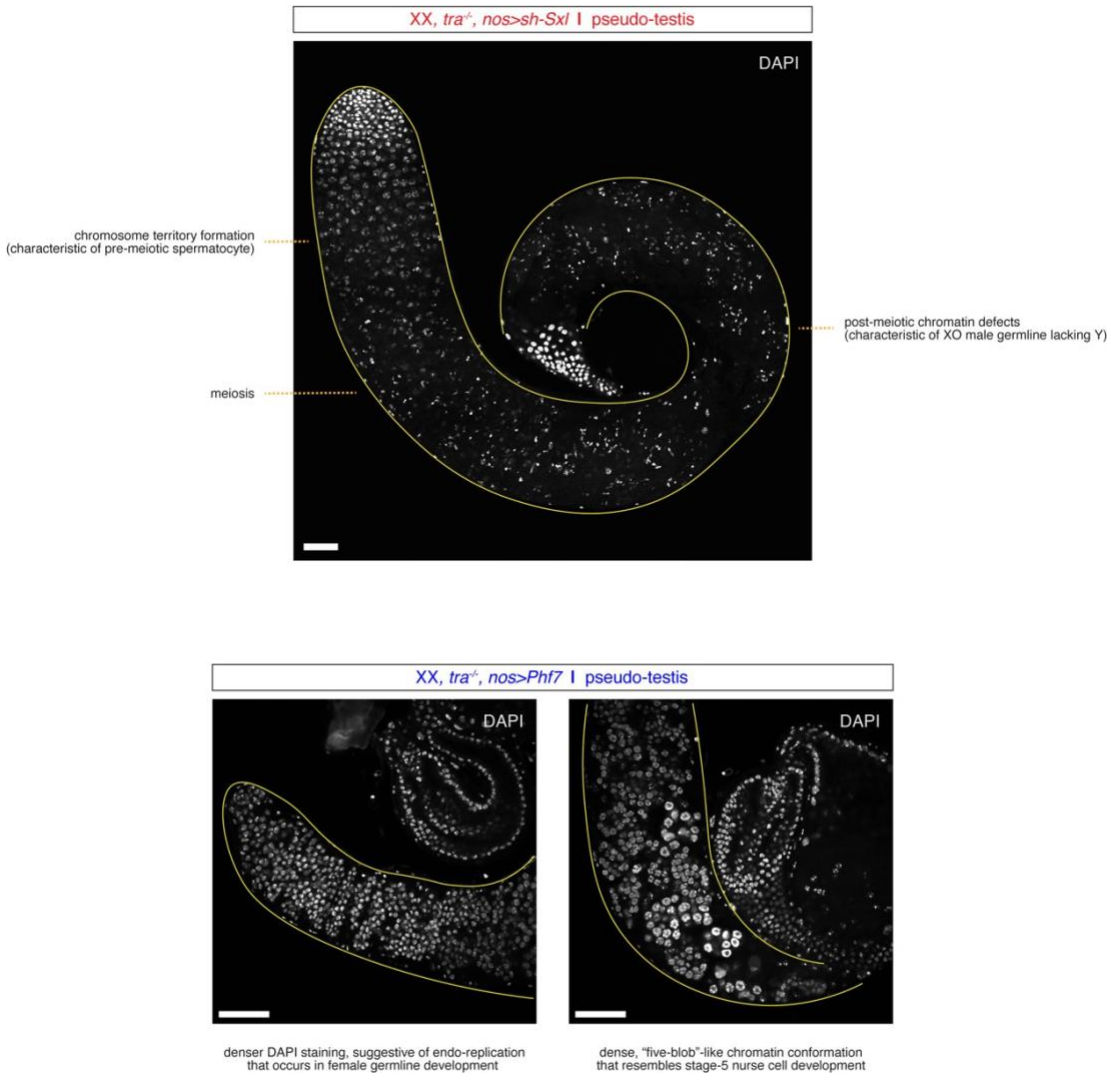


Figure S3 | Germline prevalence and phenotype in masculinized XX.

DAPI staining of masculinized XX tra-/- germline with either Sxl knock-down (top) or Phf7 ectopic expression (bottom). In both masculinization perturbations, germline is prevalent, but they show distinct phenotypes. Sxl perturbation mimics male germline development (with, e.g., chromosome territory formation characteristic of spermatocytes), while Phf7 perturbation resembles female germline development (with, e.g., nurse cell-like chromatin

conformation), though both surrounded by masculinized gonadal soma. Of note, these germline phenotypes correlate with the “sex” of the piRNA program. Scale bar: 50μm.

Chapter 6

ADAPTIVE PIRNA PATHWAY TUNING TAMES RECENTLY EXPANDED SELFISH GENES

P. Chen, K. C. Pan, E. H. Park, Y. Luo, Y. C. G. Lee, A. A. Aravin (2024).
submitted